Analysis & PDE

Volume 10 No. 5 2017

msp

Analysis & PDE

msp.org/apde

See inside back cover or msp.org/apde for submission instructions.

The subscription price for 2017 is US $265/year for the electronic version, and $470/year (+$55, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscriber address should be sent to MSP.

Analysis & PDE (ISSN 1948-206X electronic, 2157-5045 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

APDE peer review and production are managed by EditFlow® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

ANALYSIS AND PDE
Vol. 10, No. 5, 2017
dx.doi.org/10.2140/apde.2017.10.1017

msp

HARDY-SINGULAR BOUNDARY MASS AND SOBOLEV-CRITICAL VARIATIONAL PROBLEMS

NASSIF GHOUSSOUB AND FRÉDÉRIC ROBERT

We investigate the Hardy–Schrödinger operator $L_\gamma = -\Delta - \gamma/|x|^2$ on smooth domains $\Omega \subset \mathbb{R}^n$ whose boundaries contain the singularity 0. We prove a Hopf-type result and optimal regularity for variational solutions of corresponding linear and nonlinear Dirichlet boundary value problems, including the equation $L_\gamma u = u^{2^\star(s)-1}/|x|^s$, where $\gamma < \frac{1}{4}n^2$, $s \in [0, 2)$ and $2^\star(s) := 2(n-s)/(n-2)$ is the critical Hardy–Sobolev exponent. We also give a complete description of the profile of all positive solutions — variational or not — of the corresponding linear equation on the punctured domain. The value $\gamma = \frac{1}{4}(n^2-1)$ turns out to be a critical threshold for the operator L_γ. When $\frac{1}{4}(n^2-1) < \gamma < \frac{1}{4}n^2$, a notion of *Hardy singular boundary mass* $m_\gamma(\Omega)$ associated to the operator L_γ can be assigned to any conformally bounded domain Ω such that $0 \in \partial\Omega$. As a byproduct, we give a complete answer to problems of existence of extremals for Hardy–Sobolev inequalities, and consequently for those of Caffarelli, Kohn and Nirenberg. These results extend previous contributions by the authors in the case $\gamma = 0$, and by Chern and Lin for the case $\gamma < \frac{1}{4}(n-2)^2$. More specifically, we show that extremals exist when $0 \le \gamma \le \frac{1}{4}(n^2-1)$ if the mean curvature of $\partial\Omega$ at 0 is negative. On the other hand, if $\frac{1}{4}(n^2-1) < \gamma < \frac{1}{4}n^2$, extremals then exist whenever the Hardy singular boundary mass $m_\gamma(\Omega)$ of the domain is positive.

This work was carried out while F. Robert was visiting the Pacific Institute for the Mathematical Sciences (PIMS) at the University of British Columbia, as a member of the Unité Mixte Internationale of the French Centre National de la Recherche Scientifique (CNRS). He thanks the CNRS (INSMI) and UBC (PIMS) for this support. N. Ghoussoub was partially supported by a research grant from the Natural Science and Engineering Research Council of Canada (NSERC).
MSC2010: 35J35, 35J60, 58J05, 35B44.
Keywords: Hardy–Schrödinger operator, Hardy-singular boundary mass, Hardy–Sobolev inequalities, mean curvature.

1. Introduction

The borderline Dirichlet boundary value problem

$$\begin{cases} -\Delta u - \gamma \dfrac{u}{|x|^2} = u^{(n+2)/(n-2)} & \text{on } \Omega, \\ u > 0 & \text{on } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases} \tag{1-1}$$

on a smooth bounded domain Ω of $\mathbb{R}^n$ ($n \geq 3$) has no energy minimizing solutions if the singularity 0 belongs to the interior of the domain Ω; see the discussion after inequality (1-15). The situation changes dramatically, however, if 0 is situated on the boundary $\partial\Omega$. Indeed, Chern and Lin [2003; 2010] showed that solutions exist in this case provided the mean curvature of $\partial\Omega$ at 0 is negative, $n \geq 4$, and $0 < \gamma < \frac{1}{4}(n-2)^2$. The condition on γ ensures that the Hardy–Schrödinger operator $L_\gamma := -\Delta - \gamma/|x|^2$ is positive on $H_0^1(\Omega)$. This is the case as long as $\gamma < \gamma_H(\Omega)$, the latter being the best constant in the corresponding Hardy inequality, i.e.,

$$\gamma_H(\Omega) := \inf\left\{ \frac{\int_\Omega |\nabla u|^2\, dx}{\int_\Omega u^2/|x|^2\, dx} : u \in D^{1,2}(\Omega) \setminus \{0\} \right\}. \tag{1-2}$$

Here $D^{1,2}(\Omega)$ — or $H_0^1(\Omega)$ if the domain is bounded — is the completion of $C_c^\infty(\Omega)$ with respect to the norm given by $\|u\|^2 = \int_\Omega |\nabla u|^2\, dx$, and it is well known that for any domain Ω having 0 in its interior, we have

$$\gamma(\Omega) = \gamma_H(\mathbb{R}^n) = \tfrac{1}{4}(n-2)^2. \tag{1-3}$$

On the other hand, $\gamma_H(\mathbb{R}^n_+) = \frac{1}{4}n^2$ when $\mathbb{R}^n_+ := \{x \in \mathbb{R}^n : x_1 > 0\}$ is the half-space, and if Ω is any domain having 0 on its boundary, then necessarily

$$\tfrac{1}{4}(n-2)^2 < \gamma_H(\Omega) \leq \tfrac{1}{4}n^2. \tag{1-4}$$

The question of what happens when $\frac{1}{4}(n-2)^2 < \gamma < \gamma_H(\Omega)$ provided the initial motivation for this paper. To start with, we shall show that the negative mean curvature condition at 0 is still sufficient for the existence of solutions for (1-1) as long as γ remains below a new (higher) threshold, namely when $n \geq 4$ and

$$0 < \gamma \leq \tfrac{1}{4}(n^2-1). \tag{1-5}$$

However, the situation changes dramatically for the remaining interval, i.e., when

$$\tfrac{1}{4}(n^2-1) < \gamma < \gamma_H(\Omega). \tag{1-6}$$

In this case, we show that local geometric conditions at 0 become irrelevant for solving (1-1) and more global properties of the domain must come into play. This will be illustrated by the notion of *Hardy singular boundary mass* of the domain Ω that we introduce as follows.

We first consider the Hardy–Schrödinger operator $L_\gamma := -\Delta - \gamma/|x|^2$ on $\mathbb{R}^n_+$, and notice that the most basic solutions for $L_\gamma u = 0$ satisfying $u = 0$ on $\partial\mathbb{R}^n_+$ are of the form $u_\alpha(x) = x_1|x|^{-\alpha}$, and that $L_\gamma u_\alpha = 0$

on $\mathbb{R}^n_+$ if and only if α is either $\alpha_-(\gamma)$ or $\alpha_+(\gamma)$, where

$$\alpha_\pm(\gamma) := \tfrac{1}{2}n \pm \sqrt{\tfrac{1}{4}n^2 - \gamma}. \tag{1-7}$$

Actually, a byproduct of our analysis below gives that any nonnegative solution of $L_\gamma u = 0$ on $\mathbb{R}^n_+$ with $u = 0$ on $\partial\mathbb{R}^n_+$ is a linear combination of these two solutions. Note that $\alpha_-(\gamma) < \frac{1}{2}n < \alpha_+(\gamma)$, which points to the difference — in terms of behavior around 0 — between the "small" solution $x \mapsto x_1|x|^{-\alpha_-(\gamma)}$, and the "large" one $x \mapsto x_1|x|^{-\alpha_+(\gamma)}$. Indeed, the small solution is "variational", i.e., is locally in $D^{1,2}(\mathbb{R}^n_+)$, while the large one is not.

This turns out to hold in more general settings, as we show that any variational solution of $L_\gamma u = a(x)u$ behaves like $x \mapsto d(x, \partial\Omega)|x|^{-\alpha_-(\gamma)}$ around 0, while any positive nonvariational solution is necessarily like $x \mapsto d(x, \partial\Omega)|x|^{-\alpha_+(\gamma)}$ around 0. The profile can be made more explicit when $\gamma > \frac{1}{4}(n^2-1)$, as it is the only situation in which one can write a solution of $L_\gamma u = 0$ as the sum of the two above described profiles (plus lower-order terms), while if $\gamma \le \frac{1}{4}(n^2-1)$, there might be some intermediate terms between the two profiles. This led us to define the following notion of mass, which is reminiscent of the positive mass theorem of Schoen and Yau [1988] that was used to complete the solution of the Yamabe problem. This will allow us to settle the remaining cases left by Chern and Lin, since we establish that the positivity of such a boundary singular mass is sufficient to guarantee the existence of solutions for (1-1) in low dimensions.

Theorem 1.1. *Let Ω be a smooth bounded domain of $\mathbb{R}^n$ such that $0 \in \partial\Omega$. Assume $\frac{1}{4}(n^2-1) < \gamma < \gamma_H(\Omega)$. Then, up to multiplication by a positive constant, there exists a unique function $H \in C^2(\overline{\Omega} \setminus \{0\})$ such that*

$$\begin{cases} -\Delta H - \dfrac{\gamma}{|x|^2} H = 0 & \text{in } \Omega, \\ H > 0 & \text{in } \Omega, \\ H = 0 & \text{on } \partial\Omega \setminus \{0\}. \end{cases} \tag{1-8}$$

Moreover, there exists a constant $c \in \mathbb{R}$ and H satisfying (1-8) *such that*

$$H(x) = \frac{d(x, \partial\Omega)}{|x|^{\alpha_+(\gamma)}} + c\frac{d(x, \partial\Omega)}{|x|^{\alpha_-(\gamma)}} + o\left(\frac{d(x, \partial\Omega)}{|x|^{\alpha_-(\gamma)}}\right) \qquad \text{as } x \to 0.$$

Due to the uniqueness of solutions to (1-8) *up to multiplication by a constant, the coefficient c is uniquely defined. It will be denoted by $m_\gamma(\Omega) := c \in \mathbb{R}$, and will be referred to as the Hardy singular boundary mass of Ω.*

It will be shown in Section 7 that this notion of mass is conformally invariant in the following sense: if two sets are diffeomorphic via an inversion fixing 0 (see Definition 7.3 and (7-16)), then they have the same mass. As a consequence, we shall be able to define a notion of *Hardy singular boundary mass* for unbounded domains that are conformally bounded (that is, those that are smooth and bounded up to an inversion that fixes 0). We shall show that $\Omega \to m_\gamma(\Omega)$ is a monotone set-function and that $m_\gamma(\mathbb{R}^n_+) = 0$. These properties will allow us to construct in Section 9, examples of bounded domains Ω in $\mathbb{R}^n$ with $0 \in \partial\Omega$ with either positive or negative boundary mass, while satisfying any local behavior at 0 one wishes. In other words, the sign of the Hardy-singular boundary mass is totally independent of the local properties of $\partial\Omega$ around 0.

One motivation for considering equation (1-1) came from the problem of existence of extremals for the Caffarelli–Kohn–Nirenberg (CKN) inequalities [1984]. These state that in dimension $n \geq 3$, there is a constant $C := C(a, b, n) > 0$ such that, for all $u \in C_c^\infty(\mathbb{R}^n)$,

$$\left(\int_{\mathbb{R}^n} |x|^{-bq}|u|^q\right)^{2/q} \leq C \int_{\mathbb{R}^n} |x|^{-2a}|\nabla u|^2\, dx, \tag{1-9}$$

where

$$-\infty < a < \frac{n-2}{2}, \quad 0 \leq b-a \leq 1, \quad \text{and} \quad q = \frac{2n}{n-2+2(b-a)}. \tag{1-10}$$

If we let $D_a^{1,2}(\Omega)$ be the completion of $C_c^\infty(\Omega)$ with respect to the norm $\|u\|_a^2 = \int_\Omega |x|^{-2a}|\nabla u|^2\, dx$, then the best constant in (1-9) is given by

$$S(a, b, \Omega) = \inf\left\{ \frac{\int_\Omega |x|^{-2a}|\nabla u|^2\, dx}{\left(\int_\Omega |x|^{-bq}|u|^q\right)^{2/q} dx} : u \in D_a^{1,2}(\Omega)\backslash\{0\}\right\}. \tag{1-11}$$

The extremal functions for $S(a, b, \Omega)$ — whenever they exist — are then the least-energy solutions of the corresponding Euler–Lagrange equations

$$\begin{cases} -\operatorname{div}(|x|^{-2a}\nabla u) = |x|^{-bq}u^{q-1} & \text{on } \Omega, \\ u > 0 & \text{on } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases} \tag{1-12}$$

To make the connection with the Hardy–Schrödinger operator, note that the substitution $v(x) = |x|^{-a}u(x)$ with $a < \frac{1}{2}(n-2)$, gives — via the Hardy inequality — that $u \in D_a^{1,2}(\Omega)$ if and only if $v \in D^{1,2}(\Omega)$ and that u is a variational solution of (1-12) if and only if v is a solution of equation

$$\begin{cases} -\Delta v - \gamma \dfrac{v}{|x|^2} = \dfrac{v^{2^\star(s)-1}}{|x|^s} & \text{on } \Omega, \\ v > 0 & \text{on } \Omega, \\ v = 0 & \text{on } \partial\Omega, \end{cases} \tag{1-13}$$

where

$$\gamma = a(n-2-a), \quad s = (b-a)q \quad \text{and} \quad 2^\star = \frac{2n}{n-2+2(b-a)}. \tag{1-14}$$

The Caffarelli–Kohn–Nirenberg inequalities are then equivalent to the Hardy–Sobolev inequality

$$C\left(\int_\Omega \frac{u^{2^\star(s)}}{|x|^s}\, dx\right)^{2/2^\star(s)} \leq \int_\Omega |\nabla u|^2\, dx - \gamma \int_\Omega \frac{u^2}{|x|^2}\, dx \quad \text{for all } u \in D^{1,2}(\Omega), \tag{1-15}$$

at least in the case when $\gamma < \frac{1}{4}(n-2)^2$, which is optimal for domains Ω having 0 in their interior. If Ω is also bounded, then the best constant in (1-15) is never attained; that is, (1-13) has no energy minimizing solution.

However, when $0 \in \partial\Omega$, inequality (1-15) holds for γ all the way up to $\frac{1}{4}n^2$, and we shall work thereafter towards solving (1-13) by finding extremals for the variational problem

$$\mu_{\gamma,s}(\Omega) := \inf\big\{ J^{\Omega}_{\gamma,s}(u) : u \in D^{1,2}(\Omega) \setminus \{0\}\big\}, \tag{1-16}$$

where $J^{\Omega}_{\gamma,s}$ is the functional on $D^{1,2}(\Omega)$ defined by

$$J^{\Omega}_{\gamma,s}(u) := \frac{\int_\Omega |\nabla u|^2 - \gamma \int_\Omega u^2/|x|^2\, dx}{\big(\int_\Omega u^{2^\star(s)}/|x|^s\, dx\big)^{2/2^\star(s)}}. \tag{1-17}$$

We shall therefore consider the more general equation (1-13). The study of this type of nonlinear singular problems when $0 \in \partial\Omega$ was initiated by Ghoussoub and Kang [2004] and studied extensively by Ghoussoub and Robert [2006a; 2006b] in the case $\gamma = 0$. Chern and Lin [2003; 2010] and Lin and Wadade [2012] dealt with the case $\gamma < \frac{1}{4}(n-2)^2$. For more contributions, we refer to [Attar, Merchán and Peral 2015; Dávila and Peral 2011; Gmira and Véron 1991].

Theorem 1.2. *Let Ω be a smooth bounded domain in $\mathbb{R}^n$ ($n \geq 3$) such that $0 \in \partial\Omega$. Assume $\gamma \leq \frac{1}{4}(n^2-1)$ and $0 \leq s < 2$. If either $\{s > 0\}$ or $\{s = 0,\ n \geq 4$ and $\gamma > 0\}$, then there are extremals for $\mu_{\gamma,s}(\Omega)$ provided the mean curvature of $\partial\Omega$ at 0 is negative.*

As mentioned above, our main contribution here to this problem is to consider the cases when $\frac{1}{4}(n^2-1) \leq \gamma < \frac{1}{4}n^2$, as well as when $n = 3,\ s = 0$ and $\gamma > 0$, which were left open by Chern and Lin [2010]. We now introduce the new ingredients that we bring to the discussion.

We first note that standard compactness arguments [Ghoussoub and Kang 2004; Chern and Lin 2010] yield that for $\mu_{\gamma,s}(\Omega)$ to be attained it is sufficient to have that

$$\mu_{\gamma,s}(\Omega) < \mu_{\gamma,s}(\mathbb{R}^n_+), \tag{1-18}$$

and in order to prove the existence of such a gap, one tries to construct test functions for $\mu_{\gamma,s}(\Omega)$ that are based on the extremals of $\mu_{\gamma,s}(\mathbb{R}^n_+)$ provided the latter exist. The cases where this is known are given by the following standard proposition. See, for instance, [Bartsch, Peng and Zhang 2007; Chern and Lin 2010]. A complete proof is given in [Ghoussoub and Robert 2016].

Proposition 1.3. *Assume $\gamma < \frac{1}{4}n^2,\ n \geq 3$ and $0 \leq s < 2$. Then*:

(1) *$\mu_{\gamma,s}(\mathbb{R}^n_+)$ is attained provided either $\{s > 0\}$ or $\{s = 0,\ n \geq 4$ and $\gamma > 0\}$.*

(2) *On the other hand, there are no extremals for $\mu_{\gamma,s}(\mathbb{R}^n_+)$ for any $n \geq 3$ if $\{s = 0$ and $\gamma \leq 0\}$.*

(3) *Furthermore, whenever $\mu_{\gamma,0}(\mathbb{R}^n_+)$ has no extremals, then necessarily*

$$\mu_{\gamma,0}(\mathbb{R}^n_+) = \inf_{u \in D^{1,2}(\mathbb{R}^n)\setminus\{0\}} \frac{\int_{\mathbb{R}^n} |\nabla u|^2\, dx}{\big(\int_{\mathbb{R}^n} |u|^{2^\star}\, dx\big)^{2/2^\star}} = \frac{1}{K(n,2)^2}, \tag{1-19}$$

where $2^\star := 2n/(n-2)$ and $1/K(n,2)^2$ is the best constant in the Sobolev inequality.

The only unknown situation on $\mathbb{R}^n_+$ is again when $s = 0,\ n = 3$ and $\gamma > 0$, which we address in Section 10.

Assuming first that an extremal for $\mu_{\gamma,s}(\mathbb{R}^n_+)$ exists and that one knows its profile at infinity and at 0, this information can be used to construct test functions for $\mu_{\gamma,s}(\Omega)$. This classical method has been used by Kang and Ghoussoub [2004], by Ghoussoub and Robert [2006b; 2006a] when $\gamma = 0$, and by Chern and Lin [2010] for $0 < \gamma < \frac{1}{4}(n-2)^2$ in order to establish (1-18) under the assumption that $\partial\Omega$ has a negative mean curvature at 0. Actually, the estimates of Chern and Lin [2010] extend directly to establish Theorem 1.2 for all $\gamma < \frac{1}{4}(n^2-1)$ under the same negative mean curvature condition. However, the case where $\gamma = \frac{1}{4}(n^2-1)$ already requires estimates on the profile of variational solutions of (1-13) on $\mathbb{R}^n_+$ that are finer than those used by Chern and Lin [2010]. The following description of such a profile will allow us to construct sharper test functions and to prove existence of solutions for (1-13) when $\gamma = \frac{1}{4}(n^2-1)$.

Theorem 1.4. *Assume $\gamma < \frac{1}{4}n^2$ and $0 \le s < 2$, and let $u \in D^{1,2}(\mathbb{R}^n_+)$, $u \ge 0$, $u \not\equiv 0$ be a weak solution to*

$$-\Delta u - \frac{\gamma}{|x|^2}u = \frac{u^{2^\star(s)-1}}{|x|^s} \quad \text{in } \mathbb{R}^n_+. \tag{1-20}$$

Then, there exist $K_1, K_2 > 0$ such that

$$u(x) \sim_{x\to 0} K_1 \frac{x_1}{|x|^{\alpha_-(\gamma)}} \quad \text{and} \quad u(x) \sim_{|x|\to+\infty} K_2 \frac{x_1}{|x|^{\alpha_+(\gamma)}}.$$

The solution of the problem on $\mathbb{R}^n_+$ also enjoys the following natural symmetry that will be crucial for the sequel. This was carried out by Ghoussoub and Robert [2006a] when $\gamma = 0$, and their proof extends immediately to the case $0 \le \gamma < \frac{1}{4}n^2$. Chern and Lin [2010] gave another proof which also includes the case where $\gamma < 0$.

Theorem 1.5 [Chern and Lin 2010]. *If u is a nonnegative solution to* (1-20) *in $D^{1,2}(\mathbb{R}^n_+)$, then $u \circ \sigma = u$ for all isometries of $\mathbb{R}^n$ such that $\sigma(\mathbb{R}^n_+) = \mathbb{R}^n_+$. In particular, there exists $v \in C^\infty((0,+\infty)\times\mathbb{R})$ such that for all $x_1 > 0$ and all $x' \in \mathbb{R}^{n-1}$, we have that $u(x_1, x') = v(x_1, |x'|)$.*

The following theorem summarizes the situation for low dimensions.

Theorem 1.6. *Let Ω be a bounded smooth domain of $\mathbb{R}^n$ ($n \ge 3$) such that $0 \in \partial\Omega$, hence $\frac{1}{4}(n-2)^2 < \gamma_H(\Omega) \le \frac{1}{4}n^2$. Let $0 \le s < 2$.*

(1) *If $\gamma_H(\Omega) \le \gamma < \frac{1}{4}n^2$, then there are extremals for $\mu_{\gamma,s}(\Omega)$ for all $n \ge 3$.*

(2) *If $\frac{1}{4}(n^2-1) < \gamma < \gamma_H(\Omega)$ and either $\{s > 0\}$ or $\{s = 0,\ n \ge 4$ and $\gamma > 0\}$, then there are extremals for $\mu_{\gamma,s}(\Omega)$ provided the Hardy singular boundary mass $m_\gamma(\Omega)$ is positive.*

(3) *If $\{s = 0$ and $\gamma \le 0\}$, then there are no extremals for $\mu_{\gamma,0}(\Omega)$ for any $n \ge 3$.*

Finally, we address in Section 10 the only remaining case, i.e., $n = 3$, $s = 0$ and $\gamma \in \left(0, \frac{9}{4}\right)$. In this situation, there may or may not be extremals for $\mu_{\gamma,0}(\mathbb{R}^3_+)$. If they do exist, we can then argue as before — using the same test functions — to conclude existence of extremals under the same conditions, that is, either $\gamma \le 2$ and the mean curvature of $\partial\Omega$ at 0 is negative, or $\gamma > 2$ and the mass $m_\gamma(\Omega)$ is positive. However, if no extremals exist for $\mu_{\gamma,0}(\mathbb{R}^3_+)$, then as noted in (1-19), we have that

$$\mu_{\gamma,0}(\mathbb{R}^3_+) = \inf_{u\in D^{1,2}(\mathbb{R}^3)\setminus\{0\}} \frac{\int_{\mathbb{R}^3} |\nabla u|^2\,dx}{\left(\int_{\mathbb{R}^3} |u|^{2^\star}\,dx\right)^{2/2^\star}} = \frac{1}{K(3,2)^2},$$

Hardy term	dimension	geometric condition	extremal
$-\infty < \gamma \le \frac{1}{4}(n^2-1)$	$n \ge 3$	negative mean curvature at 0	yes
$\frac{1}{4}(n^2-1) < \gamma < \frac{1}{4}n^2$	$n \ge 3$	positive boundary-mass	yes

Table 1. Singular Sobolev-critical term: $s > 0$.

Hardy term	dimension	geometric condition	extremal
$0 < \gamma \le \frac{1}{4}(n^2-1)$	$n = 3$ $n \ge 4$	negative mean curvature at 0 and positive internal mass negative mean curvature at 0	yes yes
$\frac{1}{4}(n^2-1) < \gamma < \frac{1}{4}n^2$	$n = 3$ $n \ge 4$	positive boundary-mass and positive internal mass positive boundary mass	yes yes
$\gamma \le 0$	$n \ge 3$	—	no

Table 2. Nonsingular Sobolev-critical term: $s = 0$.

and we are back to the case of the Yamabe problem with no boundary singularity. This means that one needs to resort to a more standard notion of mass $R_\gamma(\Omega, x_0)$ associated to L_γ and an interior point $x_0 \in \Omega$ in order to construct suitable test functions in the spirit of [Schoen 1984]. Such an *interior mass* will be introduced in Section 10. We get the following (note that the boundary mass $m_\gamma(\Omega)$ was defined in Theorem 1.1).

Theorem 1.7. *Let Ω be a bounded smooth domain of $\mathbb{R}^3$ such that $0 \in \partial\Omega$. In particular $\frac{1}{4} < \gamma_H(\Omega) \le \frac{9}{4}$.*

(1) *If $\gamma_H(\Omega) \le \gamma < \frac{9}{4}$, then there are extremals for $\mu_{\gamma,0}(\Omega)$.*

(2) *If $0 < \gamma < \gamma_H(\Omega)$ and if there exists $x_0 \in \Omega$ such that $R_\gamma(\Omega, x_0) > 0$, then there are extremals for $\mu_{\gamma,0}(\Omega)$ under either one of the following conditions*:

  (a) *$\gamma \le 2$ and the mean curvature of $\partial\Omega$ at 0 is negative.*

  (b) *$\gamma > 2$ and the boundary mass $m_\gamma(\Omega)$ is positive.*

More precisely, if there are extremals for $\mu_{\gamma,0}(\mathbb{R}^3)$, then conditions (a) and (b) are sufficient to get extremals for $\mu_{\gamma,0}(\Omega)$. If there are no extremals for $\mu_{\gamma,0}(\mathbb{R}^3)$, then the positivity of the internal mass $R_\gamma(\Omega, x_0)$ is sufficient to get extremals for $\mu_{\gamma,0}(\Omega)$. Tables 1 and 2 summarize our findings.

Notation. In the sequel, $C_i(a, b, \dots)$ $(i = 1, 2, \dots)$ will denote constants depending on $a, b, \dots$. The same notation can be used for different constants, even in the same line. We will always refer to the monograph [Gilbarg and Trudinger 1998] for the standard results on elliptic PDEs.

2. Old and new inequalities involving singular weights

The following general form of the Hardy inequality is well known. See, for example, [Cowan 2010] or the book [Ghoussoub and Moradifam 2013].

Theorem 2.1. *Let Ω be a connected open subset of $\mathbb{R}^n$ and consider $\rho \in C^\infty(\Omega)$ such that $\rho > 0$ and $-\Delta\rho > 0$. Then, for any $u \in D^{1,2}(\Omega)$ we have*

$$\int_\Omega \frac{-\Delta\rho}{\rho} u^2\, dx \le \int_\Omega |\nabla u|^2\, dx. \tag{2-1}$$

Moreover, the case of equality is achieved exactly on $\mathbb{R}\rho \cap D^{1,2}(\Omega)$. In particular, if $\rho \notin D^{1,2}(\Omega)$, there are no nontrivial extremals for (2-1).

The above theorem applies to various weight functions ρ. See, for example, [Cowan 2010; Ghoussoub and Moradifam 2013]. For this paper, we use it to derive the following inequality.

Corollary 2.2. *Fix $1 \le k \le n$. We then have the following inequality.*

$$\left(\frac{n+2k-2}{2}\right)^2 = \inf_u \frac{\int_{\mathbb{R}^k_+ \times \mathbb{R}^{n-k}} |\nabla u|^2\, dx}{\int_{\mathbb{R}^k_+ \times \mathbb{R}^{n-k}} u^2/|x|^2\, dx},$$

where the infimum is taken over all u in $D^{1,2}(\mathbb{R}^k_+ \times \mathbb{R}^{n-k}) \setminus \{0\}$. Moreover, the infimum is never achieved.

Proof. Take $\rho(x) := x_1 \cdots x_k |x|^{-\alpha}$ for all $x \in \Omega := \mathbb{R}^k_+ \times \mathbb{R}^{n-k} \setminus \{0\}$. Then

$$\frac{-\Delta\rho}{\rho} = \frac{\alpha(n+2k-2-\alpha)}{|x|^2}.$$

We then maximize the constant by taking $\alpha := \frac{1}{2}(n+2k-2)$. Since $\rho \notin D^{1,2}(\mathbb{R}^k_+ \times \mathbb{R}^{n-k})$, Theorem 2.1 applies and we obtain that

$$\left(\frac{n+2k-2}{2}\right)^2 \int_{\mathbb{R}^k_+ \times \mathbb{R}^{n-k}} \frac{u^2}{|x|^2}\, dx \le \int_{\mathbb{R}^k_+ \times \mathbb{R}^{n-k}} |\nabla u|^2\, dx \tag{2-2}$$

for all $u \in D^{1,2}(\mathbb{R}^k_+ \times \mathbb{R}^{n-k})$, and that the extremals are trivial.

It remains to prove that the constant in (2-2) is optimal. This will be achieved via the following test function estimates. Construct a sequence $(\rho_\epsilon)_{\epsilon>0} \in D^{1,2}(\mathbb{R}^k_+ \times \mathbb{R}^{n-k})$ as follows. Starting with $\rho(x) = x_1 \cdots x_k |x|^{-\alpha}$, we fix $\beta > 0$ and define

$$\rho_\epsilon(x) := \begin{cases} |x/\epsilon|^\beta \rho(x) & \text{if } |x| < \epsilon, \\ \rho(x) & \text{if } \epsilon \le |x| \le 1/\epsilon, \\ |\epsilon \cdot x|^{-\beta} \rho(x) & \text{if } |x| > 1/\epsilon, \end{cases} \tag{2-3}$$

with $\alpha := \frac{1}{2}(n+2k-2)$. As one checks, $\rho_\epsilon \in D^{1,2}(\mathbb{R}^k_+ \times \mathbb{R}^{n-k})$ for all $\epsilon > 0$. The changes of variables $x = \epsilon y$ and $x = \epsilon^{-1} z$ yield

$$\begin{aligned} \int_{B_\epsilon(0)} \frac{\rho_\epsilon^2}{|x|^2}\, dx = O(1), \qquad \int_{\mathbb{R}^n \setminus \bar{B}_{\epsilon^{-1}}(0)} \frac{\rho_\epsilon^2}{|x|^2}\, dx = O(1), \\ \int_{B_\epsilon(0)} |\nabla \rho_\epsilon|^2\, dx = O(1), \quad \int_{\mathbb{R}^n \setminus \bar{B}_{\epsilon^{-1}}(0)} |\nabla \rho_\epsilon|^2\, dx = O(1), \end{aligned} \tag{2-4}$$

when $\epsilon \to 0$. By integrating by parts, we get

$$\int_{B_{\epsilon^{-1}}(0)\setminus\bar{B}_\epsilon(0)} |\nabla\rho_\epsilon|^2\,dx = \int_{B_{\epsilon^{-1}}(0)\setminus\bar{B}_\epsilon(0)} \frac{-\Delta\rho}{\rho}\rho^2\,dx + O(1)$$
$$= \left(\frac{n+2k-2}{2}\right)^2 \int_{B_{\epsilon^{-1}}(0)\setminus\bar{B}_\epsilon(0)} \frac{\rho^2}{|x|^2}\,dx + O(1), \tag{2-5}$$

when $\epsilon \to 0$. Using polar coordinates, we obtain

$$\int_{B_{\epsilon^{-1}}(0)\setminus\bar{B}_\epsilon(0)} \frac{\rho^2}{|x|^2}\,dx = C(2)\ln\frac{1}{\epsilon}, \quad \text{where } C(2) := 2\int_{\mathbb{S}^{n-1}} \left|\prod_{i=1}^k x_i\right|^2 d\sigma. \tag{2-6}$$

Therefore, by using (2-4), (2-5) and (2-6),

$$\frac{\int_{\mathbb{R}^k_+\times\mathbb{R}^{n-k}} |\nabla\rho_\epsilon|^2\,dx}{\int_{\mathbb{R}^k_+\times\mathbb{R}^{n-k}} \rho_\epsilon^2/|x|^2\,dx} = \left(\frac{n+2k-2}{2}\right)^2 + o(1)$$

as $\epsilon \to 0$, and we are done. Note that the infimum is never achieved since $\rho \notin D^{1,2}(\mathbb{R}^k_+ \times \mathbb{R}^{n-k})$. □

Another approach to prove Corollary 2.2 is to see $\mathbb{R}^k_+ \times \mathbb{R}^{n-k}$ as a cone generated by a domain of the unit sphere. Then the Hardy constant is given by the Hardy constant of $\mathbb{R}^n$ plus the first eigenvalue of the Laplacian of the Dirichlet of the above domain of the unit sphere endowed with its canonical metric. This point of view is developed in [Pinchover and Tintarev 2005] (see also [Fall and Musina 2012; Ghoussoub and Moradifam 2013] for an exposition in book form).

We also have the following generalized Caffarelli–Kohn–Nirenberg inequality.

Proposition 2.3. *Let Ω be an open subset of $\mathbb{R}^n$. Let $\rho, \rho' \in C^\infty(\Omega)$ be such that $\rho, \rho' > 0$ and $-\Delta\rho, -\Delta\rho' > 0$. Fix $s \in [0, 2]$ and assume that there exists $\epsilon \in (0, 1)$ and $\rho_\epsilon \in C^\infty(\Omega)$ such that*

$$\frac{-\Delta\rho}{\rho} \le (1-\epsilon)\frac{-\Delta\rho_\epsilon}{\rho_\epsilon} \quad \text{in } \Omega \text{ with } \rho_\epsilon, -\Delta\rho_\epsilon > 0.$$

Then, for all $u \in C_c^\infty(\Omega)$,

$$\left(\int_\Omega \left(\frac{-\Delta\rho'}{\rho'}\right)^{s/2} \rho^{2^\star(s)}|u|^{2^\star(s)}\,dx\right)^{2/2^\star(s)} \le C\int_\Omega \rho^2|\nabla u|^2\,dx. \tag{2-7}$$

Proof. The Sobolev inequality yields the existence of $C(n) > 0$ such that

$$\left(\int_\Omega |u|^{2^\star}\,dx\right)^{2/2^\star} \le C(n)\int_\Omega |\nabla u|^2\,dx$$

for all $u \in C_c^\infty(\Omega)$, where $2^\star = 2^\star(0) = 2n/(n-2)$. A Hölder inequality interpolating between this Sobolev inequality and the Hardy inequality (2-1) for ρ' yields the existence of $C > 0$ such that for all $u \in C_c^\infty(\Omega)$,

$$\left(\int_\Omega \left(\frac{-\Delta\rho'}{\rho'}\right)^{s/2} |u|^{2^\star(s)}\,dx\right)^{2/2^\star(s)} \le C\int_\Omega |\nabla u|^2\,dx. \tag{2-8}$$

By applying (2-1) to ρ_ϵ, we get for $v \in C_c^\infty(\Omega)$,

$$\int_\Omega \rho^2 |\nabla v|^2 \, dx = \int_\Omega |\nabla(\rho v)|^2 \, dx - \int_\Omega \frac{-\Delta\rho}{\rho}(\rho v)^2 \, dx$$
$$\geq \int_\Omega |\nabla(\rho v)|^2 \, dx - (1-\epsilon)\int_\Omega \frac{-\Delta\rho_\epsilon}{\rho_\epsilon}(\rho v)^2 \, dx \geq \epsilon \int_\Omega |\nabla(\rho v)|^2.$$

Taking $u := \rho v$ in (2-8) and using this latest inequality yield (2-7). □

Corollary 2.4. *Fix $k \in \{1, \dots, n-1\}$. There exists then a constant $C := C(a, b, n) > 0$ such that for all $u \in C_c^\infty(\mathbb{R}_+^k \times \mathbb{R}^{n-k})$,*

$$\left(\int_{\mathbb{R}_+^k \times \mathbb{R}^{n-k}} |x|^{-bq} \left(\prod_{i=1}^k x_i\right)^q |u|^q\right)^{2/q} \leq C \int_{\mathbb{R}_+^k \times \mathbb{R}^{n-k}} \left(\prod_{i=1}^k x_i\right)^2 |x|^{-2a} |\nabla u|^2 \, dx, \tag{2-9}$$

where

$$-\infty < a < \frac{n-2+2k}{2}, \quad 0 \leq b-a \leq 1, \quad q = \frac{2n}{n-2+2(b-a)}. \tag{2-10}$$

Proof. Apply Proposition 2.3 with $\rho(x) = \rho'(x) = \left(\prod_{i=1}^k x_i\right)|x|^{-a}$ and $\rho_\epsilon(x) = \left(\prod_{i=1}^k x_i\right)|x|^{-(n-2+2k)/2}$ for all $x \in \mathbb{R}_+^k \times \mathbb{R}^{n-k}$. Corollary 2.4 then follows for suitable a, b, q. □

Remark. Observe that by taking $k = 0$, we recover the classical Caffarelli–Kohn–Nirenberg inequalities (1-9). However, one does not see any improvement in the integrability of the weight functions since $\left(\prod_{i=1}^k x_i\right)|x|^{-a}$ is of order $k - a > -\frac{1}{2}(n-2)$, hence as close as we wish to $(n-2)/2$ with the right choice of a. The relevance here appears when one considers the Hardy inequality of Corollary 2.2.

3. On the best constants in the Hardy and Hardy–Sobolev inequalities

As mentioned in the Introduction, the best constant in the Hardy inequality $\gamma_H(\Omega)$ does not depend on the domain $\Omega \subset \mathbb{R}^n$ if the singularity 0 belongs to the interior of Ω, and it is always equal to $\frac{1}{4}(n-2)^2$. We have seen, however, in the last section that the situation changes whenever $0 \in \partial\Omega$, since $\gamma_H(\mathbb{R}_+^n) = \frac{1}{4}n^2$. Some properties of the best Hardy constants were studied in [Fall and Musina 2012; Fall 2012]. In this section, we shall collect whatever information we shall need later on about γ_H.

Proposition 3.1. *The best Hardy constant γ_H satisfies the following properties*:

(1) $\gamma_H(\Omega) = \frac{1}{4}(n-2)^2$ *for any smooth domain Ω such that $0 \in \Omega$.*

(2) *If $0 \in \partial\Omega$, then $\frac{1}{4}(n-2)^2 < \gamma_H(\Omega) \leq \frac{1}{4}n^2$.*

(3) $\gamma_H(\Omega) = \frac{1}{4}n^2$ *for every Ω such that $0 \in \partial\Omega$ and $\Omega \subset \mathbb{R}_+^n$.*

(4) *If $\gamma_H(\Omega) < \frac{1}{4}n^2$, then it is attained in $D^{1,2}(\Omega)$.*

(5) *We have* $\inf\{\gamma_H(\Omega) : 0 \in \partial\Omega\} = \frac{1}{4}(n-2)^2$.

(6) *For every $\epsilon > 0$, there exists a smooth domain $\mathbb{R}_+^n \subsetneq \Omega_\epsilon \subsetneq \mathbb{R}^n$ such that $0 \in \partial\Omega_\epsilon$ and $\frac{1}{4}n^2 - \epsilon \leq \gamma_H(\Omega_\epsilon) < \frac{1}{4}n^2$.*

Proof. Properties (1)–(4) are well known (see [Fall and Musina 2012; Fall 2012]). We sketch proofs since we will make frequent use of the test functions involved. Note first that Corollary 2.2 already yields that $\gamma_H(\mathbb{R}^n_+) = \frac{1}{4}n^2$.

(2) Since $\Omega \subset \mathbb{R}^n$, we have that $\gamma_H(\Omega) \geq \gamma_H(\mathbb{R}^n) = \frac{1}{4}(n-2)^2$. Assume by contradiction that $\gamma_H(\Omega) = \frac{1}{4}(n-2)^2$. It then follows from Theorem 3.6 below (applied with $s = 2$) that $\gamma_H(\Omega)$ is achieved by a function in $u_0 \in D^{1,2}(\Omega) \setminus \{0\}$ (note that $\mu_{0,\gamma}(\Omega) = \gamma_H(\Omega) - \gamma$). Therefore, $\gamma_H(\mathbb{R}^n)$ is achieved in $D^{1,2}(\mathbb{R}^n)$. Up to taking $|u_0|$, we can assume that $u_0 \geq 0$. Therefore, the Euler–Lagrange equation and the maximum principle yield $u_0 > 0$ in $\mathbb{R}^n$: this is impossible since $u_0 \in D^{1,2}(\Omega)$. Therefore $\gamma_H(\Omega) > \frac{1}{4}(n-2)^2$.

For the other inequality, the standard proof normally uses the fact that the domain contains an interior sphere that is tangent to the boundary at 0. We choose here to perform another proof based on test functions, which will be used again to prove Proposition 3.3. It goes as follows: since Ω is a smooth bounded domain of $\mathbb{R}^n$ such that $0 \in \partial\Omega$, there exist U, V open subsets of $\mathbb{R}^n$ such that $0 \in U$ and $0 \in V$ and there exists $\varphi \in C^\infty(U, V)$ a diffeomorphism such that $\varphi(0) = 0$ and

$$\varphi(U \cap \{x_1 > 0\}) = \varphi(U) \cap \Omega \quad \text{and} \quad \varphi(U \cap \{x_1 = 0\}) = \varphi(U) \cap \partial\Omega.$$

Moreover, we can and shall assume that $d\varphi_0$ is an isometry. Let $\eta \in C^\infty_c(U)$ such that $\eta(x) = 1$ for $x \in B_\delta(0)$ for some $\delta > 0$ small enough, and consider $(\alpha_\epsilon)_{\epsilon>0} \in (0, +\infty)$ such that $\alpha_\epsilon = o(\epsilon)$ as $\epsilon \to 0$. For $\epsilon > 0$, define

$$u_\epsilon(x) := \begin{cases} \eta(y)\alpha_\epsilon^{-(n-2)/2}\rho_\epsilon(y/\alpha_\epsilon) & \text{for all } x \in \varphi(U) \cap \Omega,\ x = \varphi(y), \\ 0 & \text{elsewhere.} \end{cases} \tag{3-1}$$

Here ρ_ϵ is constructed as in (2-3) with $k = 1$. Now fix $\sigma \in [0, 2]$, and note that only the case $\sigma = 2$ is needed for the above proposition. Immediate computations yield

$$\int_\Omega \frac{|u_\epsilon(y)|^{2^\star(\sigma)}}{|y|^\sigma}\, dy = C(\sigma) \ln \frac{1}{\epsilon} + O(1) \quad \text{as } \epsilon \to 0, \tag{3-2}$$

where $C(\sigma) := 2\int_{\mathbb{S}^{n-1}} \left|\prod_{i=1}^k x_i\right|^{2^\star(\sigma)} d\sigma$. Similar arguments yield

$$\int_\Omega |\nabla u_\epsilon|^2\, dy = \frac{n^2}{4} C(2) \ln \frac{1}{\epsilon} + O(1) \quad \text{as } \epsilon \to 0. \tag{3-3}$$

As a consequence, we get that

$$\frac{\int_\Omega |\nabla u_\epsilon|^2\, dx}{\int_\Omega u_\epsilon^2/|x|^2\, dx} = \frac{n^2}{4} + o(1) \quad \text{as } \epsilon \to 0.$$

In particular, we get that $\gamma_H(\Omega) \leq \frac{1}{4}n^2$, which proves the upper bound in item (2) of the proposition.

(3) Assume that $\Omega \subset \mathbb{R}^n_+$. Then $D^{1,2}(\Omega) \subset D^{1,2}(\mathbb{R}^n_+)$, and therefore $\gamma_H(\Omega) \geq \gamma_H(\mathbb{R}^n_+) = \frac{1}{4}n^2$. With the reverse inequality already given by item (2), we get that $\gamma_H(\Omega) = \frac{1}{4}n^2$ for all $\Omega \subset \mathbb{R}^n_+$ such that $0 \in \partial\Omega$.

(4) This will be a particular case of Theorem 3.6 when $s = 2$.

(5) Let Ω_0 be a bounded domain of $\mathbb{R}^n$ such that $0 \in \Omega_0$ (i.e., it is not on the boundary). Given $\delta > 0$, we chop out a ball of radius $\frac{1}{4}\delta$ with 0 on its boundary to define $\Omega_\delta := \Omega_0 \setminus \bar{B}_{\delta/4}\big(\big(-\frac{1}{4}\delta, 0, \dots, 0\big)\big)$. Note that for $\delta > 0$ small enough, Ω is smooth and $0 \in \partial\Omega$. We now prove that

$$\lim_{\delta\to 0} \gamma_H(\Omega_\delta) = \tfrac{1}{4}(n-2)^2. \tag{3-4}$$

Define $\eta_1 \in C^\infty(\mathbb{R}^n)$ such that $\eta_1(x) = 0$ if $|x| < 1$ and $\eta_1(x) = 1$ if $|x| > 2$. Let $\eta_\delta(x) := \eta_1(\delta^{-1}x)$ for all $\delta > 0$ and $x \in \mathbb{R}^n$. Fix $U \in C_c^\infty(\mathbb{R}^n)$ and consider, for any $\delta > 0$, an $\epsilon_\delta > 0$ such that $\lim_{\delta\to 0} \delta/\epsilon_\delta = \lim_{\delta\to 0} \epsilon_\delta = 0$. For $\delta > 0$, we define

$$u_\delta(x) := \eta_\delta(x)\epsilon_\delta^{-(n-2)/2} U(\epsilon_\delta^{-1}x) \quad \text{for all } x \in \Omega_\delta.$$

For $\delta > 0$ small enough, we have that $u_\delta \in C_c^\infty(\Omega_\delta)$. Since $\delta = o(\epsilon_\delta)$ as $\delta \to 0$, a change of variable yields

$$\lim_{\delta\to 0} \int_{\Omega_\delta} \frac{u_\delta^2}{|x|^2}\,dx = \int_{\mathbb{R}^n} \frac{U^2}{|x|^2}\,dx.$$

We also have for δ small,

$$\begin{aligned}\int_{\Omega_\delta} |\nabla u_\delta|^2\,dx &= \int_{\mathbb{R}^n} |\nabla u_\delta|^2\,dx = \int_{\mathbb{R}^n} |\nabla(U\cdot\eta_{\delta/\epsilon_\delta})|^2\,dx \\ &= \int_{\mathbb{R}^n} |\nabla U|^2 \eta_{\delta/\epsilon_\delta}^2\,dx + \int_{\mathbb{R}^n} \eta_{\delta/\epsilon_\delta}(-\Delta\eta_{\delta/\epsilon_\delta})U^2\,dx. \end{aligned}\tag{3-5}$$

Let $R > 0$ be such that U has support in $B_R(0)$. Since $n \geq 3$, we have

$$\int_{\mathbb{R}^n} \eta_{\delta/\epsilon_\delta}(-\Delta\eta_{\delta/\epsilon_\delta})U^2\,dx = O\left(\left(\frac{\epsilon_\delta}{\delta}\right)^2 \mathrm{Vol}\big(B_R(0)\cap \mathrm{Supp}(-\Delta\eta_{\delta/\epsilon_\delta})\big)\right) = O\left(\left(\frac{\delta}{\epsilon_\delta}\right)^{n-2}\right) = o(1)$$

as $\delta \to 0$. This latest identity, (3-5) and the dominated convergence theorem yield

$$\lim_{\delta\to 0} \int_{\Omega_\delta} |\nabla u_\delta|^2\,dx = \int_{\mathbb{R}^n} |\nabla U|^2\,dx.$$

Therefore, for $U \in C_c^\infty(\mathbb{R}^n)$, we have

$$\limsup_{\delta\to 0} \gamma_H(\Omega_\delta) \leq \lim_{\delta\to 0} \frac{\int_{\Omega_\delta} |\nabla u_\delta|^2\,dx}{\int_{\Omega_\delta} u_\delta^2/|x|^2\,dx} = \frac{\int_{\mathbb{R}^n} |\nabla U|^2\,dx}{\int_{\mathbb{R}^n} U^2/|x|^2\,dx}.$$

Taking the infimum over all $U \in C_c^\infty(\mathbb{R}^n)$, we get that

$$\limsup_{\delta\to 0} \gamma_H(\Omega_\delta) \leq \inf_{U\in D^{1,2}(\mathbb{R}^n)\setminus\{0\}} \frac{\int_{\mathbb{R}^n} |\nabla U|^2\,dx}{\int_{\mathbb{R}^n} U^2/|x|^2\,dx} = \gamma_H(\mathbb{R}^n) = \tfrac{1}{4}(n-2)^2.$$

Since $\gamma_H(\Omega_\delta) \geq \frac{1}{4}(n-2)^2$ for all $\delta > 0$, this completes the proof of (3-4), yielding (5).

For (6) we use the following observation.

Lemma 3.2. *Let $(\Phi_k)_{k\in\mathbb{N}} \in C^1(\mathbb{R}^n, \mathbb{R}^n)$ be such that*

$$\lim_{k\to+\infty} \big(\|\Phi_k - \mathrm{Id}_{\mathbb{R}^n}\|_\infty + \|\nabla(\Phi_k - \mathrm{Id}_{\mathbb{R}^n})\|_\infty\big) = 0 \quad \textit{and} \quad \Phi_k(0) = 0. \tag{3-6}$$

Let $D \subset \mathbb{R}^n$ be an open domain such that $0 \in \partial D$ (not necessarily bounded or regular), and set $D_k := \Phi_k(D)$ for all $k \in \mathbb{N}$. Then $0 \in \partial D_k$ for all $k \in \mathbb{N}$ and

$$\lim_{k\to+\infty} \gamma_H(D_k) = \gamma_H(D). \tag{3-7}$$

Proof. If $u \in C_c^\infty(D_k)$, then $u \circ \Phi_k \in C_c^\infty(D)$ and

$$\int_{D_k} |\nabla u|^2 \, dx = \int_{\mathbb{R}^n_+} |\nabla(u \circ \Phi_k)|^2_{\Phi_k^\star \mathrm{Eucl}} \, |\mathrm{Jac}\, \Phi_k| \, dx, \tag{3-8}$$

$$\int_{D_k} \frac{u^2}{|x|^2} \, dx = \int_{\mathbb{R}^n_+} \frac{(u \circ \Phi_k(x))^2}{|\Phi_k(x)|^2} \, |\mathrm{Jac}\, \Phi_k| \, dx, \tag{3-9}$$

where here and in the sequel $\Phi_k^\star \mathrm{Eucl}$ is the pull-back of the Euclidean metric via the diffeomorphism Φ_k. Assumption (3-6) yields

$$\lim_{k\to+\infty} \sup_{x\in D} \left(\left| \frac{|\Phi_k(x)|}{|x|} - 1 \right| + \sup_{i,j} \left| \big(\partial_i \Phi_k(x), \partial_j \Phi_k(x)\big) - \delta_{ij} \right| + |\mathrm{Jac}\, \Phi_k - 1| \right) = 0,$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. This limit, (3-8), (3-9) and a density argument yield (3-7). □

We now prove (6) of Proposition 3.1. Let $\varphi \in C^\infty(\mathbb{R}^{n-1})$ such that $0 \le \varphi \le 1$, $\varphi(0) = 0$, and $\varphi(x') = 1$ for all $x' \in \mathbb{R}^{n-1}$ be such that $|x'| \ge 1$. For $t \ge 0$, define $\Phi_t(x_1, x') := (x_1 - t\varphi(x'), x')$ for all $(x_1, x') \in \mathbb{R}^n$. Set $\widetilde{\Omega}_t := \Phi_t(\mathbb{R}^n_+)$ and apply Lemma 3.2 to note that $\lim_{\epsilon\to 0} \gamma_H(\widetilde{\Omega}_t) = \gamma_H(\mathbb{R}^n_+) = \frac{1}{4}n^2$. Since $\varphi \ge 0$ and $\varphi \not\equiv 0$, we have $\mathbb{R}^n_+ \subsetneq \widetilde{\Omega}_t$ for all $t > 0$. To get (6) it suffices to take $\Omega_\epsilon := \widetilde{\Omega}_t$ for $t > 0$ small enough. □

As in the case of $\gamma_H(\Omega)$, the best Hardy–Sobolev constant

$$\mu_{\gamma,s}(\Omega) := \inf \left\{ \frac{\int_\Omega |\nabla u|^2 \, dx - \gamma \int_\Omega u^2/|x|^2 \, dx}{\left(\int_\Omega u^{2^\star(s)}/|x|^s \, dx\right)^{2/2^\star(s)}} : u \in D^{1,2}(\Omega) \setminus \{0\} \right\}$$

will depend on the geometry of Ω whenever $0 \in \partial\Omega$.

Proposition 3.3. *Let Ω be a bounded smooth domain such that $0 \in \partial\Omega$.*

(1) *If $\gamma < \frac{1}{4}n^2$, then $\mu_{\gamma,s}(\Omega) > -\infty$.*

(2) *If $\gamma > \frac{1}{4}n^2$, then $\mu_{\gamma,s}(\Omega) = -\infty$.*

Moreover,

(3) *If $\gamma < \gamma_H(\Omega)$, then $\mu_{\gamma,s}(\Omega) > 0$.*

(4) *If $\gamma_H(\Omega) < \gamma < \frac{1}{4}n^2$, then $0 > \mu_{\gamma,s}(\Omega) > -\infty$.*

(5) *If $\gamma = \gamma_H(\Omega) < \frac{1}{4}n^2$, then $\mu_{\gamma,s}(\Omega) = 0$.*

Proof. Assume that $\gamma < \frac{1}{4}n^2$ and let $\epsilon > 0$ be such that $(1+\epsilon)\gamma \le \frac{1}{4}n^2$. It follows from Proposition 3.5 that there exists $C_\epsilon > 0$ such that for $u \in D^{1,2}(\Omega)$,

$$\frac{n^2}{4} \int_\Omega \frac{u^2}{|x|^2} \, dx \le (1+\epsilon) \int_\Omega |\nabla u|^2 \, dx + C_\epsilon \int_\Omega u^2 \, dx.$$

For any $u \in D^{1,2}(\Omega) \setminus \{0\}$, we have

$$J^{\Omega}_{\gamma,s}(u) \geq \frac{\big(1-(4\gamma/n^2)(1+\epsilon)\big)\int_\Omega |\nabla u|^2\,dx - (4\gamma/n^2)C_\epsilon \int_\Omega u^2\,dx}{\Big(\int_\Omega |u|^{2^\star(s)}/|x|^s\,dx\Big)^{2/2^\star(s)}} \geq -\frac{4\gamma}{n^2}C_\epsilon \frac{\int_\Omega u^2\,dx}{\Big(\int_\Omega |u|^{2^\star(s)}/|x|^s\,dx\Big)^{2/2^\star(s)}}.$$

It follows from Hölder's inequality that there exists $C > 0$ independent of u such that

$$\int_\Omega u^2\,dx \leq C\bigg(\int_\Omega \frac{|u|^{2^\star(s)}}{|x|^s}\,dx\bigg)^{2/2^\star(s)}.$$

It then follows that $J^{\Omega}_{\gamma,s}(u) \geq -(4\gamma/n^2)C_\epsilon C$ for all $u \in D^{1,2}(\Omega)\setminus\{0\}$. Therefore $\mu_{\gamma,s}(\Omega) > -\infty$ whenever $\gamma < \frac{1}{4}n^2$.

Assume now that $\gamma > \frac{1}{4}n^2$ and define for every $\epsilon > 0$ a function $u_\epsilon \in D^{1,2}(\Omega)$ as in (3-1). It then follows from (3-2) and (3-3) that as $\epsilon \to 0$,

$$J^{\Omega}_{\gamma,s}(u_\epsilon) = \frac{\big(\frac{1}{4}n^2-\gamma\big)C(2)\ln(1/\epsilon)+O(1)}{\big(C(s)\ln(1/\epsilon)+O(1)\big)^{2/2^\star(s)}} = \bigg(\Big(\tfrac{1}{4}n^2-\gamma\Big)\frac{C(2)}{C(s)^{2/2^\star(s)}}+o(1)\bigg)\bigg(\ln\frac{1}{\epsilon}\bigg)^{(2-s)/(n-s)}.$$

Since $s < 2$ and $\gamma > \frac{1}{4}n^2$, we have $\lim_{\epsilon\to 0} J^{\Omega}_{\gamma,s}(u_\epsilon) = -\infty$; therefore $\mu_{\gamma,s}(\Omega) = -\infty$.

If $\gamma < \gamma_H(\Omega)$, Sobolev's embedding theorem yields $\mu_{0,s}(\Omega) > 0$; hence the result is clear for all $\gamma \leq 0$ since then $\mu_{\gamma,s}(\Omega) \geq \mu_{0,s}(\Omega)$. If now $0 \leq \gamma < \gamma_H(\Omega)$, it follows from the definition of $\gamma_H(\Omega)$ that for all $u \in D^{1,2}(\Omega)\setminus\{0\}$,

$$J^{\Omega}_{\gamma,s}(u) = \frac{\int_\Omega |\nabla u|^2 - \gamma\int_\Omega u^2/|x|^2\,dx}{\Big(\int_\Omega u^{2^\star(s)}/|x|^s\,dx\Big)^{2/2^\star(s)}} \geq \bigg(1-\frac{\gamma}{\gamma_H(\Omega)}\bigg)\frac{\int_\Omega |\nabla u|^2\,dx}{\Big(\int_\Omega |u|^{2^\star(s)}/|x|^s\,dx\Big)^{2/2^\star(s)}} \geq \bigg(1-\frac{\gamma}{\gamma_H(\Omega)}\bigg)\mu_{0,s}(\Omega).$$

Therefore $\mu_{\gamma,s}(\Omega) \geq (1-\gamma/\gamma_H(\Omega))\mu_{0,s}(\Omega) > 0$ when $\gamma < \gamma_H(\Omega)$.

If $\gamma_H(\Omega) < \gamma < \frac{1}{4}n^2$, then Proposition 3.1(4) yields that $\gamma_H(\Omega)$ is attained. We let u_0 be such an extremal. In particular $J^{\Omega}_{\gamma_H(\Omega),s}(u) \geq 0 = J^{\Omega}_{\gamma_H(\Omega),s}(u_0)$, and therefore $\mu_{\gamma_H(\Omega),s}(\Omega) = 0$. Since $\gamma_H(\Omega) < \gamma < \frac{1}{4}n^2$, we have that $J^{\Omega}_{\gamma,s}(u_0) < 0$, and therefore $\mu_{\gamma,s}(\Omega) < 0$ when $\gamma_H(\Omega) < \gamma < \frac{1}{4}n^2$. □

Remark 3.4. The case $\gamma = \frac{1}{4}n^2$ is unclear and anything can happen at that value of γ. For example, if $\gamma_H(\Omega) < \frac{1}{4}n^2$ then $\mu_{n^2/4,s}(\Omega) < 0$, while if $\gamma_H(\Omega) = \frac{1}{4}n^2$ then $\mu_{n^2/4,s}(\Omega) \geq 0$. It is our guess that many examples reflecting different regimes can be constructed.

We shall need the following standard result.

Proposition 3.5. *Assume $\gamma < \frac{1}{4}n^2$ and $s \in [0,2]$. Then, for any $\epsilon > 0$, there exists $C_\epsilon > 0$ such that, for all $u \in D^{1,2}(\Omega)$,*

$$\bigg(\int_\Omega \frac{|u|^{2^\star(s)}}{|x|^s}\,dx\bigg)^{2/2^\star(s)} \leq \bigg(\frac{1}{\mu_{\gamma,s}(\mathbb{R}^n_+)}+\epsilon\bigg)\int_\Omega\bigg(|\nabla u|^2-\gamma\frac{u^2}{|x|^2}\bigg)\,dx + C_\epsilon\int_\Omega u^2\,dx. \tag{3-10}$$

This result says that, up to adding an L^2-term (indeed, any subcritical term fits), the best constant in the Hardy–Sobolev embedding can be chosen to be as close as one wishes to the best constant in the model space $\mathbb{R}^n_+$. One can see this by noting that for functions that are supported in a small neighborhood

of 0, the domain Ω looks like $\mathbb{R}^n_+$, and the distortion is determined by the radius of the neighborhood. The case of general functions in $D^{1,2}(\Omega)$ is dealt with by using a cut-off, which induces the L^2-norm. A detailed proof is given in [Ghoussoub and Robert 2016].

The following result is central for the sequel. The proof is standard, ever since T. Aubin's proof of the Yamabe conjecture in high dimensions, where he noted that the compactness of minimizing sequences is restored if the infimum is strictly below the energy of a "bubble". In our case below, this translates to $\mu_{\gamma,s}(\Omega) < \mu_{\gamma,s}(\mathbb{R}^n_+)$. We omit the proof, which can be found in [Ghoussoub and Robert 2016].

Theorem 3.6. *Assume that $\gamma < \frac{1}{4}n^2$, $0 \le s \le 2$ and $\mu_{\gamma,s}(\Omega) < \mu_{\gamma,s}(\mathbb{R}^n_+)$. Then there are extremals for $\mu_{\gamma,s}(\Omega)$. In particular, there exists a minimizer u in $D^{1,2}(\Omega) \setminus \{0\}$ that is a positive solution to the equation*

$$\begin{cases} -\Delta u - \gamma \dfrac{u}{|x|^2} = \mu_{\gamma,s}(\Omega) \dfrac{u^{2^\star(s)-1}}{|x|^s} & \text{in } \Omega, \\ u > 0 & \text{in } \partial\Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases} \tag{3-11}$$

4. Profile at 0 of the variational solutions of $L_\gamma u = a(x)u$

Here and in the sequel, we shall assume that $0 \in \partial\Omega$, where Ω is a smooth domain. Recall from the Introduction that two solutions for $L_\gamma u = 0$, with $u = 0$ on $\partial\mathbb{R}^n_+$, are of the form $u_\alpha(x) = x_1|x|^{-\alpha}$, where $\alpha \in \{\alpha_-(\gamma), \alpha_+(\gamma)\}$ with

$$\alpha_-(\gamma) := \tfrac{1}{2}n - \sqrt{\tfrac{1}{4}n^2 - \gamma} \quad \text{and} \quad \alpha_+(\gamma) := \tfrac{1}{2}n + \sqrt{\tfrac{1}{4}n^2 - \gamma}. \tag{4-1}$$

These solutions will be the building blocks for sub- and supersolutions of more general linear equations involving L_γ on other domains. This section is devoted to the proof of the following result. To state the theorem, we use the following terminology:

We say that $u \in D^{1,2}(\Omega)_{\mathrm{loc},0}$ if there exists $\eta \in C^\infty_c(\mathbb{R}^n)$ such that $\eta \equiv 1$ around 0 and $\eta u \in D^{1,2}(\Omega)$. We say that $u \in D^{1,2}(\Omega)_{\mathrm{loc},0}$ is *a weak solution* to the equation

$$-\Delta u = F \in (D^{1,2}(\Omega)_{\mathrm{loc},0})'$$

if for any $\varphi \in D^{1,2}(\Omega)$ and $\eta \in C^\infty_c(\mathbb{R}^n)$ with sufficiently small support around 0, we have

$$\int_\Omega (\nabla u, \nabla(\eta\varphi))\, dx = \langle F, \eta\varphi\rangle.$$

Theorem 4.1. *Fix $\gamma < \frac{1}{4}n^2$ and $\tau > 0$, and let $u \in D^{1,2}(\Omega)_{\mathrm{loc},0}$ be a weak solution of*

$$-\Delta u - \frac{\gamma + O(|x|^\tau)}{|x|^2} u = 0 \quad \text{in } D^{1,2}(\Omega)_{\mathrm{loc},0}. \tag{4-2}$$

Then, there exists $K \in \mathbb{R}$ such that

$$\lim_{x\to 0} \frac{u(x)}{d(x, \partial\Omega)|x|^{-\alpha_-(\gamma)}} = K.$$

Moreover, if $u \ge 0$ and $u \not\equiv 0$, we have that $K > 0$.

By a slight abuse of notation,

$$u \mapsto -\Delta u - \frac{\gamma + O(|x|^\tau)}{|x|^2}u$$

will denote an operator

$$u \mapsto -\Delta u - \frac{\gamma + a(x)}{|x|^2}u,$$

where $a \in C^0(\overline{\Omega})$ such that $a(x) = O(|x|^\tau)$ as $\tau \to 0$. In Section 6, we will give a full description of solutions to (4-2) that are not necessarily variational (we also refer to [Pinchover 1994] for related problems).

We need the following lemmas, which will be used frequently throughout the paper. The first is only the initial step towards proving rigidity for the solutions of $L_\gamma u = 0$ on $\mathbb{R}^n_+$. Indeed, the pointwise assumption $u(x) \le C|x|^{1-\alpha}$ will not be necessary as it will be eventually removed in Proposition 6.4, which will be a consequence of the classification Theorem 6.1. We omit the proof as it can be inferred from the work of Pinchover and Tintarev [2005].

Lemma 4.2 (rigidity). *Let $u \in C^2(\overline{\mathbb{R}^n_+} \setminus \{0\})$ be a nonnegative solution of*

$$\begin{cases} -\Delta u - \dfrac{\gamma}{|x|^2}u = 0 & \text{in } \mathbb{R}^n_+, \\ u = 0 & \text{on } \partial\mathbb{R}^n_+. \end{cases} \tag{4-3}$$

Suppose $u(x) \le C|x|^{1-\alpha}$ on $\mathbb{R}^n_+$ for $\alpha \in \{\alpha_-(\gamma), \alpha_+(\gamma)\}$, then there exists $\lambda \ge 0$ such that $u(x) = \lambda x_1 |x|^{-\alpha}$ for all $x \in \mathbb{R}^n_+$.

We now construct basic sub- and supersolutions for the equation $L_\gamma u = a(x)u$, where $a(x) = O(|x|^{\tau-2})$ for some $\tau > 0$.

Proposition 4.3. *Let $\gamma < \frac{1}{4}n^2$ and $\alpha \in \{\alpha_-(\gamma), \alpha_+(\gamma)\}$. Let $0 < \tau \le 1$ and $\beta \in \mathbb{R}$ be such that $\alpha - \tau < \beta < \alpha$ and $\beta \notin \{\alpha_-(\gamma), \alpha_+(\gamma)\}$. Then, there exist $r > 0$, and $u_{\alpha,+}, u_{\alpha,-} \in C^\infty(\overline{\Omega} \setminus \{0\})$ such that*

$$\begin{cases} u_{\alpha,+},\ u_{\alpha,-} > 0 & \text{in } \Omega \cap B_r(0), \\ u_{\alpha,+},\ u_{\alpha,-} = 0 & \text{on } \partial\Omega \cap B_r(0), \\ -\Delta u_{\alpha,+} - \dfrac{\gamma + O(|x|^\tau)}{|x|^2}u_{\alpha,+} > 0 & \text{in } \Omega \cap B_r(0), \\ -\Delta u_{\alpha,-} - \dfrac{\gamma + O(|x|^\tau)}{|x|^2}u_{\alpha,-} < 0 & \text{in } \Omega \cap B_r(0). \end{cases} \tag{4-4}$$

Moreover, we have as $x \to 0$, $x \in \Omega$, that

$$u_{\alpha,+}(x) = \frac{d(x, \partial\Omega)}{|x|^\alpha}(1 + O(|x|^{\alpha-\beta})) \quad \text{and} \quad u_{\alpha,-}(x) = \frac{d(x, \partial\Omega)}{|x|^\alpha}(1 + O(|x|^{\alpha-\beta})). \tag{4-5}$$

Proof. We first choose an adapted chart to lift the basic solutions from $\mathbb{R}^n_+$. Since $0 \in \partial\Omega$ and Ω is smooth, there exist $\widetilde{U}$, $\widetilde{V}$ two bounded domains of $\mathbb{R}^n$ such that $0 \in \widetilde{U}$ and $0 \in \widetilde{V}$, and there exists a C^∞-diffeomorphism $c \in C^\infty(\widetilde{U}, \widetilde{V})$ such that $c(0) = 0$,

$$c(\widetilde{U} \cap \{x_1 > 0\}) = c(\widetilde{U}) \cap \Omega \quad \text{and} \quad c(\widetilde{U} \cap \{x_1 = 0\}) = c(\widetilde{U}) \cap \partial\Omega.$$

The orientation of $\partial\Omega$ is chosen in such a way that for any $x' \in \widetilde{U} \cap \{x_1 = 0\}$,

$$\big\{\partial_1 c(0,x'), \partial_2 c(0,x'), \dots, \partial_n c(0,x')\big\}$$

is a direct basis of $\mathbb{R}^n$ (canonically oriented). For $x' \in \widetilde{U} \cap \{x_1 = 0\}$, we define $\nu(x')$ as the unique orthonormal inner vector at the tangent space $T_{c(0,x')}\partial\Omega$ (it is chosen such that $\{\nu(x'), \partial_2 c(0,x'), \dots, \partial_n c(0,x')\}$ is a direct basis of $\mathbb{R}^n$). In particular, on $\mathbb{R}^n_+ := \{x_1 > 0\}$, we have $\nu(x') := (1,0,\dots,0)$.

Here and in the sequel, we write for any $r > 0$

$$\widetilde{B}_r := (-r,r) \times B_r^{(n-1)}(0), \tag{4-6}$$

where $B_r^{(n-1)}(0)$ denotes the ball of center 0 and radius r in $\mathbb{R}^{n-1}$. It is standard that there exists $\delta > 0$ such that

$$\begin{aligned} \varphi : \widetilde{B}_{2\delta} &\to \mathbb{R}^n, \\ (x_1,x') \in \mathbb{R} \times \mathbb{R}^{n-1} &\mapsto c(0,x') + x_1 \nu(x'), \end{aligned} \tag{4-7}$$

is a C^∞-diffeomorphism onto its open image $\varphi(\widetilde{B}_{2\delta})$, and

$$\varphi(\widetilde{B}_{2\delta} \cap \{x_1 > 0\}) = \varphi(\widetilde{B}_{2\delta}) \cap \Omega \quad \text{and} \quad \varphi(\widetilde{B}_{2\delta} \cap \{x_1 = 0\}) = \varphi(\widetilde{B}_{2\delta}) \cap \partial\Omega. \tag{4-8}$$

We also have, for all $x' \in B_\delta(0)^{(n-1)}$,

$$\nu(x') \text{ is the inner orthonormal unit vector at the tangent space } T_{\varphi(0,x')}\partial\Omega. \tag{4-9}$$

An important remark is that

$$d\big(\varphi(x_1,x'), \partial\Omega\big) = |x_1| \quad \text{for all } (x_1,x') \in \widetilde{B}_{2\delta} \text{ close to 0}. \tag{4-10}$$

Consider the metric $g := \varphi^\star \mathrm{Eucl}$ on $\widetilde{B}_{2\delta}$, that is, the pull-back of the Euclidean metric Eucl via the diffeomorphism φ. Following classical notations, we define

$$g_{ij}(x) := \big(\partial_i \varphi(x), \partial_j \varphi(x)\big)_{\mathrm{Eucl}} \quad \text{for all } x \in \widetilde{B}_{2\delta} \text{ and } i,j = 1,\dots,n. \tag{4-11}$$

Up to a change of coordinates, we can assume that $(\partial_2\varphi(0), \dots, \partial_n\varphi(0))$ is an orthogonal basis of $T_0\partial\Omega$. In other words, we then have that

$$g_{ij}(0) = \delta_{ij} \quad \text{for all } i,j = 1,\dots,n. \tag{4-12}$$

As one checks,

$$g_{i1}(x) = \delta_{i1} \quad \text{for all } x \in \widetilde{B}_{2\delta} \text{ and } i = 1,\dots,n. \tag{4-13}$$

Fix now $\alpha \in \mathbb{R}$ and consider $\Theta \in C^\infty(\widetilde{B}_{2\delta})$ such that $\Theta(0) = 0$ and which will be constructed later (independently of α) with additional needed properties. Fix $\eta \in C_c^\infty(\widetilde{B}_{2\delta})$ such that $\eta(x) = 1$ for all $x \in \widetilde{B}_\delta$. Define $u_\alpha \in C^\infty(\overline{\Omega} \setminus \{0\})$ as

$$u_\alpha \circ \varphi(x_1,x') := \eta(x) x_1 |x|^{-\alpha} (1 + \Theta(x)) \quad \text{for all } (x_1,x') \in \widetilde{B}_{2\delta} \setminus \{0\}. \tag{4-14}$$

In particular, $u_\alpha(x) > 0$ for all $x \in \varphi(\widetilde{B}_{2\delta}) \cap \Omega$ and $u_\alpha(x) = 0$ on $\Omega \setminus \varphi(\widetilde{B}_{2\delta})$.

We claim that with a good choice of Θ, we have that

$$-\Delta u_\alpha = \frac{\alpha(n-\alpha)}{|x|^2} u_\alpha + O\left(\frac{u_\alpha(x)}{|x|}\right) \quad \text{as } x \to 0. \tag{4-15}$$

Indeed, using the chart φ, we have that

$$(-\Delta u_\alpha) \circ \varphi(x_1, x') = -\Delta_g (u_\alpha \circ \varphi)(x_1, x')$$

for all $(x_1, x') \in \widetilde{B}_\delta \setminus \{0\}$. Here, $-\Delta_g$ is the Laplace operator associated to the metric g; that is,

$$-\Delta_g := -g^{ij}(\partial_{ij} - \Gamma_{ij}^k \partial_k),$$

where

$$\Gamma_{ij}^k := \tfrac{1}{2} g^{km}(\partial_i g_{jm} + \partial_j g_{im} - \partial_m g_{ij}),$$

and (g^{ij}) is the inverse of the matrix (g_{ij}). Here and in the sequel, we have adopted Einstein's convention of summation. It follows from (4-13) that

$$(-\Delta u_\alpha)\circ\varphi = -\Delta_{\mathrm{Eucl}}(u_\alpha\circ\varphi) - \sum_{i,j\geq 2} (g^{ij} - \delta^{ij})\, \partial_{ij}(u_\alpha\circ\varphi) + g^{ij}\Gamma_{ij}^1\, \partial_1(u_\alpha\circ\varphi) + \sum_{k\geq 2} g^{ij}\Gamma_{ij}^k\, \partial_k(u_\alpha\circ\varphi). \tag{4-16}$$

It follows from the definition (4-14) that there exists $C > 0$ such that for any $i, j, k \geq 2$, we have that

$$\left|\partial_{ij}(u_\alpha \circ \varphi)(x_1, x')\right| \leq C|x_1| \cdot |x|^{-\alpha-2} \quad \text{and} \quad \left|\partial_k(u_\alpha \circ \varphi)(x_1, x')\right| \leq C|x_1| \cdot |x|^{-\alpha-1}$$

for all $(x_1, x') \in \widetilde{B}_\delta \setminus \{0\}$. It follows from (4-12) that $g^{ij} - \delta^{ij} = O(|x|)$ as $x \to 0$. Therefore, (4-16) yields that as $x \to 0$,

$$(-\Delta u_\alpha) \circ \varphi = -\Delta_{\mathrm{Eucl}}(u_\alpha \circ \varphi) + g^{ij}\Gamma_{ij}^1\, \partial_1(u_\alpha \circ \varphi) + O(x_1|x|^{-\alpha-1}). \tag{4-17}$$

The definition of g_{ij} and the expression of $\varphi(x_1, x')$ then yield that as $x \to 0$,

$$\begin{aligned} g^{ij}\Gamma_{ij}^1 &= -\frac{1}{2}\sum_{i,j\geq 2} g^{ij}\, \partial_1 g_{ij} \\ &= -\sum_{i,j\geq 2} g^{ij}(x_1, x')\big(\partial_i\varphi(0, x'), \partial_j \nu(x')\big) + x_1\big(\partial_i(x'), \partial_j \nu(x')\big)\big) \\ &= -\sum_{i,j\geq 2} g^{ij}(0, x')\big(\partial_i\varphi(0, x'), \partial_j \nu(x')\big) + O(|x_1|) = H(x') + O(|x_1|), \end{aligned}$$

where $H(x')$ is the mean curvature of the $(n-1)$-manifold $\partial\Omega$ at $\varphi(0, x')$ oriented by the outer normal vector $-\nu(x')$. Using the expression (4-14) and using the smoothness of Θ, (4-17) yields

$$(-\Delta u_\alpha) \circ \varphi = (-\Delta_{\mathrm{Eucl}}(x_1|x|^{-\alpha})) \cdot (1+\Theta) + |x|^{-\alpha}\big(H(x')(1+\Theta) - 2\partial_1\Theta\big) + O(x_1|x|^{-\alpha-1}) \quad \text{as } x \to 0.$$

We now define

$$\Theta(x_1, x') := e^{-x_1 H(x')/2} - 1 \quad \text{for all } x = (x_1, x') \in \widetilde{B}_{2\delta}.$$

Clearly $\Theta(0) = 0$ and $\Theta \in C^\infty(\widetilde{B}_{2\delta})$. We then get that as $x \to 0$,

$$(-\Delta u_\alpha) \circ \varphi = \frac{\alpha(n-\alpha)}{|x|^2} x_1|x|^{-\alpha} \cdot (1+\Theta) + O(x_1|x|^{-\alpha-1}). \tag{4-18}$$

With the choice that $g_{ij}(0)=\delta_{ij}$, we have that $(\partial_i\varphi(0))_{i=1,\dots,n}$ is an orthonormal basis of $\mathbb{R}^n$, and therefore $|\varphi(x)|=|x|(1+O(|x|))$ as $x\to 0$. It then follows from (4-18) and (4-14) that

$$-\Delta u_\alpha=\frac{\alpha(n-\alpha)}{|x|^2}u_\alpha+O(|x|^{-1}u_\alpha)\quad\text{as } x\to 0. \tag{4-19}$$

This proves (4-15). We now proceed with the construction of the sub- and supersolutions. Let $\alpha\in\{\alpha_-(\gamma),\alpha_+(\gamma)\}$ in such a way that $\alpha(n-\alpha)=\gamma$ and consider $\beta,\lambda\in\mathbb{R}$ to be chosen later. It follows from (4-15) that

$$\begin{aligned}\left(-\Delta-\frac{\gamma+O(|x|^\tau)}{|x|^2}\right)(u_\alpha+\lambda u_\beta)&=\frac{\lambda(\beta(n-\beta)-\gamma)}{|x|^2}u_\beta+\frac{O(|x|^\tau)}{|x|^2}u_\alpha+O(|x|^{-1}u_\alpha)+O(|x|^{\tau-2}u_\beta)\\&=\frac{u_\beta}{|x|^2}\big(\lambda(\beta(n-\beta)-\gamma)+O(|x|^\tau)+O(|x|^{\tau+\beta-\alpha})+O(|x|^{1+\beta-\alpha})\big)\end{aligned}$$

as $x\to 0$. Choose β such that $\alpha-\tau<\beta<\alpha$ in such a way that $\beta\neq\alpha_-(\gamma)$ and $\beta\neq\alpha_+(\gamma)$. In particular, $\beta>\alpha-1$ and $\beta(n-\beta)-\gamma\neq 0$. We then have

$$\left(-\Delta-\frac{\gamma+O(|x|^\tau)}{|x|^2}\right)(u_\alpha+\lambda u_\beta)=\frac{u_\beta}{|x|^2}\big(\lambda(\beta(n-\beta)-\gamma)+O(|x|^{\tau+\beta-\alpha})\big) \tag{4-20}$$

as $x\to 0$. Choose $\lambda\in\mathbb{R}$ such that $\lambda(\beta(n-\beta)-\gamma)>0$. Finally, let $u_{\alpha,+}:=u_\alpha+\lambda u_\beta$ and $u_{\alpha,-}:=u_\alpha-\lambda u_\beta$. They clearly satisfy (4-4) and (4-5), which completes the proof of Proposition 4.3. □

Lemma 4.4. *Assume that $u\in D^{1,2}(\Omega)_{\mathrm{loc},0}$ is a weak solution of*

$$\begin{cases}-\Delta u-\dfrac{\gamma+O(|x|^\tau)}{|x|^2}u=0 & \text{in } D^{1,2}(\Omega)_{\mathrm{loc},0},\\ u=0 & \text{on } B_{2\delta}(0)\cap\partial\Omega\end{cases} \tag{4-21}$$

for some $\tau>0$ and $\delta>0$. Then, there exists $C_1>0$ such that

$$|u(x)|\le C_1 d(x,\partial\Omega)|x|^{-\alpha_-(\gamma)}\quad\text{for } x\in\Omega\cap B_\delta(0). \tag{4-22}$$

Moreover, if $u>0$ in Ω, then there exists $C_2>0$ such that

$$u(x)\ge C_2 d(x,\partial\Omega)|x|^{-\alpha_-(\gamma)}\quad\text{for } x\in\Omega\cap B_\delta(0). \tag{4-23}$$

Proof. Assume first that $u\in D^{1,2}(\Omega)_{\mathrm{loc},0}$ and $u>0$ on $B_\delta(0)\cap\Omega$. We claim that there exists $C_0>0$ such that

$$\frac{1}{C_0}\frac{d(x,\partial\Omega)}{|x|^{\alpha_-(\gamma)}}\le u(x)\le C_0\frac{d(x,\partial\Omega)}{|x|^{\alpha_-(\gamma)}}\quad\text{for all } x\in\Omega\cap B_\delta(0). \tag{4-24}$$

Indeed, since u is smooth outside 0, it follows from Hopf's maximum principle that there exists $C_1,C_2>0$ such that

$$C_1 d(x,\partial\Omega)\le u(x)\le C_2 d(x,\partial\Omega)\quad\text{for all } x\in\Omega\cap\partial B_\delta(0). \tag{4-25}$$

Let $u_{\alpha_-(\gamma),+}$ be the supersolution constructed in Proposition 4.3. It follows from (4-25) and the asymptotics (4-5) of $u_{\alpha_-(\gamma),+}$ that there exists $C_3>0$ such that

$$u(x)\le C_3 u_{\alpha_-(\gamma),+}(x)\quad\text{for all } x\in\partial(B_\delta(0)\cap\Omega).$$

Since u is a solution and $u_{\alpha_-(\gamma),+}$ is a supersolution, both being in $D^{1,2}(\Omega)_{\mathrm{loc},0}$, it follows from the maximum principle (by choosing $\delta > 0$ small enough so that $-\Delta - (\gamma + O(|x|^\tau))|x|^{-2}$ is coercive on $B_\delta(0) \cap \Omega$) that $u(x) \le C_3 u_{\alpha_-(\gamma),+}(x)$ for all $x \in B_\delta(0) \cap \Omega$. In particular, it follows from the asymptotics (4-5) of $u_{\alpha_-(\gamma),+}$ that there exists $C_4 > 0$ such that $u(x) \le C_4 d(x, \partial\Omega)|x|^{-\alpha_-(\gamma)}$ for all $x \in \Omega \cap B_\delta(0)$. Arguing similarly with the lower-bound in (4-25) and the subsolution $u_{\alpha_-(\gamma),-}$, we get the existence of $C_0 > 0$ such that (4-24) holds. This yields Lemma 4.4 for $u > 0$.

Now we deal with the case when u is a sign-changing solution for (4-21). We then define $u_1, u_2 : B_\delta(0) \cap \Omega \to \mathbb{R}$ such that

$$\begin{cases} -\Delta u_1 - \dfrac{\gamma + O(|x|^\tau)}{|x|^2} u_1 = 0 & \text{in } B_\delta(0) \cap \Omega, \\ u_1(x) = \max\{u(x), 0\} & \text{on } \partial(B_\delta(0) \cap \Omega), \end{cases} \qquad \begin{cases} -\Delta u_2 - \dfrac{\gamma + O(|x|^\tau)}{|x|^2} u_2 = 0 & \text{in } B_\delta(0) \cap \Omega, \\ u_2(x) = \max\{-u(x), 0\} & \text{on } \partial(B_\delta(0) \cap \Omega). \end{cases}$$

The existence of such solutions is ensured by choosing $\delta > 0$ small enough so that the operator $-\Delta - (\gamma + O(|x|^\tau))|x|^{-2}$ is coercive on $B_\delta(0) \cap \Omega$. In particular, $u_1, u_2 \in D^{1,2}(\Omega)_{\mathrm{loc},0}$, $u_1, u_2 \ge 0$ and $u = u_1 - u_2$. It follows from the maximum principle that for all i, either $u_i \equiv 0$ or $u_i > 0$. The first part of the proof yields the upper bound for u_1, u_2. Since $u = u_1 - u_2$, we then get (4-22). □

The following lemma allows to construct sub- and supersolutions with Dirichlet boundary conditions on any small smooth domain.

Proposition 4.5. *Let Ω be a smooth bounded domain of $\mathbb{R}^n$, and let W be a smooth domain of $\mathbb{R}^n$ such that for some $r > 0$ small enough, we have*

$$B_r(0) \cap \Omega \subset W \subset B_{2r}(0) \cap \Omega \quad \text{and} \quad B_r(0) \cap \partial W = B_r(0) \cap \partial\Omega. \tag{4-26}$$

Fix $\gamma < \frac{1}{4}n^2$, $0 < \tau \le 1$ and $\beta \in \mathbb{R}$ such that $\alpha_+(\gamma) - \tau < \beta < \alpha_+(\gamma)$ and $\beta \ne \alpha_-(\gamma)$. Then, for r small enough, there exists $u^{(d)}_{\alpha_+(\gamma),+}, u^{(d)}_{\alpha_+(\gamma),-} \in C^\infty(\overline{W} \setminus \{0\})$ such that

$$\begin{cases} u^{(d)}_{\alpha_+(\gamma),+},\ u^{(d)}_{\alpha_+(\gamma),+} = 0 & \text{in } \partial W \setminus \{0\}, \\ -\Delta u^{(d)}_{\alpha_+(\gamma),+} - \dfrac{\gamma + O(|x|^\tau)}{|x|^2} u^{(d)}_{\alpha_+(\gamma),+} > 0 & \text{in } W, \\ -\Delta u^{(d)}_{\alpha_+(\gamma),-} - \dfrac{\gamma + O(|x|^\tau)}{|x|^2} u^{(d)}_{\alpha_+(\gamma),-} < 0 & \text{in } W. \end{cases} \tag{4-27}$$

Moreover, we have as $x \to 0$, $x \in \Omega$ that

$$u^{(d)}_{\alpha_+(\gamma),+}(x) = \frac{d(x, \partial\Omega)}{|x|^{\alpha_+(\gamma)}}(1 + O(|x|^{\alpha-\beta})), \tag{4-28}$$

$$u^{(d)}_{\alpha_+(\gamma),-}(x) = \frac{d(x, \partial\Omega)}{|x|^{\alpha_+(\gamma)}}(1 + O(|x|^{\alpha-\beta})). \tag{4-29}$$

Proof. Take $\eta \in C^\infty(\mathbb{R}^n)$ such that $\eta(x) = 0$ for $x \in B_{\delta/4}(0)$ and $\eta(x) = 1$ for $x \in \mathbb{R}^n \setminus B_{\delta/3}(0)$. Define on W the function

$$f(x) := \left(-\Delta - \frac{\gamma + O(|x|^\tau)}{|x|^2}\right)(\eta u_{\alpha_+(\gamma),+}),$$

where $u_{\alpha_+(\gamma),+}$ is given by Proposition 4.3. Note that f vanishes around 0 and that it is in $C^\infty(\overline{W})$. Let $v \in D^{1,2}(W)$ be such that

$$\begin{cases} -\Delta v - \dfrac{\gamma + O(|x|^\tau)}{|x|^2} v = f & \text{in } W, \\ v = 0 & \text{on } \partial W. \end{cases}$$

Note that for $r > 0$ small enough, $-\Delta - (\gamma + O(|x|^\tau))|x|^{-2}$ is coercive on W, and therefore, the existence of v is ensured for small r. Define

$$u^{(d)}_{\alpha_+(\gamma),+} := u_{\alpha_+(\gamma),+} - \eta u_{\alpha_+(\gamma),+} + v.$$

The properties of W and the definitions of η and v yield

$$\begin{cases} u^{(d)}_{\alpha_+(\gamma),+} = 0 & \text{in } \partial W \setminus \{0\}, \\ -\Delta u^{(d)}_{\alpha_+(\gamma),+} - \dfrac{\gamma + O(|x|^\tau)}{|x|^2} u^{(d)}_{\alpha_+(\gamma),+} > 0 & \text{in } W. \end{cases}$$

Since $-\Delta v - (\gamma + O(|x|^\tau))|x|^{-2} v = 0$ around 0 and $v \in D^{1,2}(W)$, it follows from Lemma 4.4 that there exists $C > 0$ such that $|v(x)| \le C d(x, W)|x|^{-\alpha_-(\gamma)}$ for all $x \in W$. Then (4-28) follows from the asymptotics (4-5) of $u_{\alpha_+(\gamma),+}$ and the fact that $\alpha_-(\gamma) < \alpha_+(\gamma)$. We argue similarly for $u^{(d)}_{\alpha_+(\gamma),-}$. □

Lemma 4.6. *Let $u \in D^{1,2}(\Omega)_{\mathrm{loc},0}$ such that* (4-2) *holds. Assume there exists $C > 0$ and $\alpha \in \{\alpha_+(\gamma), \alpha_-(\gamma)\}$ such that*

$$|u(x)| \le C|x|^{1-\alpha} \quad \text{for } x \to 0,\ x \in \Omega. \tag{4-30}$$

(1) *Then, there exists $C_1 > 0$ such that*

$$|\nabla u(x)| \le C_1 |x|^{-\alpha} \quad \text{as } x \to 0,\ x \in \Omega. \tag{4-31}$$

(2) *If $\lim_{x\to 0} |x|^{\alpha-1} u(x) = 0$, then $\lim_{x\to 0} |x|^\alpha |\nabla u(x)| = 0$. Moreover, if $u > 0$, then there exists $l \ge 0$ such that*

$$\lim_{x\to 0} \frac{|x|^\alpha u(x)}{d(x, \partial\Omega)} = l \quad \text{and} \quad \lim_{x\to 0,\, x\in\partial\Omega} |x|^\alpha |\nabla u(x)| = l. \tag{4-32}$$

Proof. Assume that (4-30) holds. Set $\omega(x) := |x|^\alpha u(x)/d(x, \partial\Omega)$ for $x \in \Omega$. Let $(x_i)_i \in \Omega$ be such that

$$\lim_{i\to+\infty} x_i = 0 \quad \text{and} \quad \lim_{i\to+\infty} \omega(x_i) = l. \tag{4-33}$$

Choose a chart φ as in (4-7) such that $d\varphi_0 = \mathrm{Id}_{\mathbb{R}^n}$. For any i, define $X_i \in \mathbb{R}^n_+$ such that $x_i = \varphi(X_i)$, $r_i := |X_i|$ and $\theta_i := X_i/|X_i|$. In particular, $\lim_{i\to+\infty} r_i = 0$ and $|\theta_i| = 1$ for all i. Set

$$\tilde{u}_i(x) := r_i^{\alpha-1} u(\varphi(r_i x)) \quad \text{for all } i \text{ and } x \in B_R(0) \cap \mathbb{R}^n_+,\ x \ne 0.$$

Equation (4-2) can then be rewritten as

$$\begin{cases} -\Delta_{g_i} \tilde{u}_i - \dfrac{\gamma + o(1)}{|x|^2} \tilde{u}_i = 0 & \text{in } B_R(0) \cap \mathbb{R}^n_+, \\ \tilde{u}_i = 0 & \text{in } B_R(0) \cap \partial\mathbb{R}^n_+, \end{cases} \tag{4-34}$$

where $g_i(x) := (\varphi^\star \text{Eucl})(r_i x)$ is a metric that goes to Eucl on every compact subset of $\mathbb{R}^n$ as $i \to \infty$. Here, $o(1) \to 0$ in $C^0_{\text{loc}}(\overline{\mathbb{R}^n_+} \setminus \{0\})$. It follows from (4-30) and (4-33) that

$$|\tilde{u}_i(x)| \le C|x|^{1-\alpha} \quad \text{for all } i \text{ and all } x \in B_R(0) \cap \mathbb{R}^n_+, \tag{4-35}$$

It follows from elliptic theory that there exists $\tilde{u} \in C^2(\overline{\mathbb{R}^n_+} \setminus \{0\})$ such that $\tilde{u}_i \to \tilde{u}$ in $C^1_{\text{loc}}(\overline{\mathbb{R}^n_+} \setminus \{0\})$. By letting $\theta := \lim_{i\to+\infty} \theta_i$ ($|\theta| = 1$), we then have that $\partial_j \tilde{u}_i(\theta_i) \to \partial_j \tilde{u}(\theta)$ as $i \to +\infty$ for any $j = 1, \dots, n$, which can be rewritten as

$$\lim_{i\to+\infty} |x_i|^\alpha \, \partial_j u(x_i) = \partial_j \tilde{u}(\theta) \quad \text{for all } j = 1, \dots, n. \tag{4-36}$$

We now prove (4-31). For that, we argue by contradiction and assume that there exists a sequence $(x_i)_i \in \Omega$ that goes to 0 as $i \to +\infty$ and such that $|x_i|^\alpha |\nabla u(x_i)| \to +\infty$ as $i \to +\infty$. It then follows from (4-36) that $|x_i|^\alpha |\nabla u(x_i)| = O(1)$ as $i \to +\infty$. This is a contradiction to our assumption, which proves (4-31). The case when $|x|^\alpha u(x) \to 0$ as $x \to 0$ goes similarly.

Now we consider the case when $u > 0$, which implies that $\tilde{u}_i \ge 0$ and $\tilde{u} \ge 0$. We let $l \in [0, +\infty]$ and $(x_i)_i \in \Omega$ be such that

$$\lim_{i\to+\infty} x_i = 0 \quad \text{and} \quad \lim_{i\to+\infty} \omega(x_i) = l. \tag{4-37}$$

We claim that

$$0 \le l < +\infty \quad \text{and} \quad \lim_{x\to 0} \omega(x) = l \in [0, +\infty). \tag{4-38}$$

Indeed, using the notations above, we get that

$$\lim_{i\to+\infty} \frac{\tilde{u}_i(\theta_i)}{(\theta_i)_1} = l.$$

The convergence of $\tilde{u}_i$ in $C^1_{\text{loc}}(\overline{\mathbb{R}^n_+} \setminus \{0\})$ then yields $l < +\infty$. Passing to the limit as $i \to +\infty$ in (4-34), we get

$$\begin{cases} -\Delta_{\text{Eucl}} \tilde{u} - \dfrac{\gamma}{|x|^2} \tilde{u} = 0 & \text{in } \mathbb{R}^n_+, \\ \tilde{u} \ge 0 & \text{in } \mathbb{R}^n_+, \\ \tilde{u} = 0 & \text{in } \partial\mathbb{R}^n_+. \end{cases}$$

The limit (4-37) can be rewritten as $\tilde{u}(\theta) = l\theta_1$ if $\theta \in \mathbb{R}^n_+$ and $\partial_1 \tilde{u}(\theta) = l$ if $\theta \in \partial\mathbb{R}^n_+$. The rigidity lemma, Lemma 4.2, then yields

$$\tilde{u}(x) = l x_1 |x|^{-\alpha} \quad \text{for all } x \in \mathbb{R}^n_+.$$

In particular, since the differential of φ at 0 is the identity map, it follows from the convergence of $\tilde{u}_i$ to $\tilde{u}$ locally in C^1 that

$$\lim_{i\to+\infty} \sup_{x\in\Omega\cap\partial B_{r_i}(0)} \frac{u(x)}{d(x, \partial\Omega)|x|^{-\alpha}} = \sup_{x\in\mathbb{R}^n_+\cap\partial B_1(0)} \frac{\tilde{u}(x)}{x_1|x|^{-\alpha}} = l \tag{4-39}$$

and

$$\lim_{i\to+\infty} \inf_{x\in\Omega\cap\partial B_{r_i}(0)} \frac{u(x)}{d(x, \partial\Omega)|x|^{-\alpha}} = \inf_{x\in\mathbb{R}^n_+\cap\partial B_1(0)} \frac{\tilde{u}(x)}{x_1|x|^{-\alpha}} = l. \tag{4-40}$$

We distinguish two cases:

Case 1: $\alpha = \alpha_+(\gamma)$. Let W and $u^{(d)}_{\alpha_+(\gamma),-}$ be as in Proposition 4.5, and fix $\epsilon > 0$. Note that the existence and properties of $u^{(d)}_{\alpha_+(\gamma),-}$ do not use the lemma that is currently being proved. It follows from (4-40) that there exists i_0 such that for $i \geq i_0$, we have

$$u(x) \geq (l-\epsilon)u^{(d)}_{\alpha_+(\gamma),-}(x) \quad \text{for all } x \in W \cap \partial B_{r_i}(0).$$

Since $\big(-\Delta - (\gamma + O(|x|^\tau))|x|^{-2}\big)\big(u - (l-\epsilon)u^{(d)}_{\alpha_+(\gamma),-}\big) \geq 0$ in $W \setminus B_{r_i}(0)$ and since $u_{\alpha_+(\gamma),-}$ vanishes on $\partial W \setminus \{0\}$, it follows from the comparison principle that

$$u(x) \geq (l-\epsilon)u^{(d)}_{\alpha_+(\gamma),-}(x) \quad \text{for all } x \in W \setminus \partial B_{r_i}(0).$$

Letting $i \to +\infty$ yields

$$u(x) \geq (l-\epsilon)u^{(d)}_{\alpha_+(\gamma),-}(x) \quad \text{for all } x \in W \setminus \{0\}.$$

It follows from this inequality and the asymptotics for $u^{(d)}_{\alpha_+(\gamma),-}$ that

$$\liminf_{x\to 0} \omega(x) \geq l.$$

Note that this is valid for any $l \in \mathbb{R}$ satisfying (4-37). By taking $l := \limsup_{x\to 0} \omega(x)$, we then get that $\lim_{x\to 0} \omega(x) = l$.

Case 2: $\alpha = \alpha_-(\gamma)$. Consider the super- and subsolutions $u_{\alpha_-(\gamma),+}$, $u_{\alpha_-(\gamma),-}$ constructed in Proposition 4.3. It follows from (4-39) and (4-40) that for $\epsilon > 0$, there exists i_0 such that, for $i \geq i_0$, we have

$$(l-\epsilon)u_{\alpha_-(\gamma),-}(x) \leq u(x) \leq (l+\epsilon)u_{\alpha_-(\gamma),+}(x) \quad \text{for all } x \in \Omega \cap \partial B_{r_i}(0).$$

Since the operator $-\Delta - (\gamma + O(|x|^\tau))|x|^{-2}$ is coercive on $\Omega \cap B_{r_i}(0)$ and the functions we consider are in $D^{1,2}_{\mathrm{loc},0}(\Omega \cap B_{r_i}(0))$ (i.e., they are variational), it follows from the maximum principle that

$$(l-\epsilon)u_{\alpha_-(\gamma),-}(x) \leq u(x) \leq (l+\epsilon)u_{\alpha_-(\gamma),+}(x) \quad \text{for all } x \in \Omega \cap B_{r_i}(0).$$

Using the asymptotics (4-5) of the sub- and supersolutions, we get that

$$(l-\epsilon) \leq \liminf_{x\to 0} \frac{u(x)}{d(x,\partial\Omega)|x|^{-\alpha_-(\gamma)}} \leq \limsup_{x\to 0} \frac{u(x)}{d(x,\partial\Omega)|x|^{-\alpha_-(\gamma)}} \leq (l+\epsilon).$$

Letting $\epsilon \to 0$ yields $\lim_{x\to 0} \omega(x) = l \geq 0$. This ends Case 2 and completes the proof of (4-38).

The case $u > 0$ is a consequence of (4-38) and (4-36) (note that for the second limit, $x_i \in \partial\Omega$ can be rewritten as $\theta_i \in \partial\mathbb{R}^n_+$ and therefore $(\theta_i)_1 = 0$). This ends the proof of Lemma 4.6. □

Proof of Theorem 4.1. First, assume that $u \in D^{1,2}(\Omega)_{\mathrm{loc},0}$ satisfies (4-2) and $u > 0$ on $B_\delta(0) \cap \Omega$. It then follows from Lemma 4.4 that there exists $C_0 > 0$ such that

$$\frac{1}{C_0}\frac{d(x,\partial\Omega)}{|x|^{\alpha_-(\gamma)}} \leq u(x) \leq C_0\frac{d(x,\partial\Omega)}{|x|^{\alpha_-(\gamma)}} \quad \text{for all } x \in \Omega \cap B_\delta(0).$$

Since $u > 0$, this estimate coupled with Lemma 4.6 yields the theorem for $u > 0$.

If now u is a sign-changing solution for (4-2), we define $u_1, u_2 : B_\delta(0) \cap \Omega \to \mathbb{R}_{\geq 0}$ as in the proof of Lemma 4.4. The first part of the proof yields that there exist $l_1, l_2 \geq 0$ such that

$$\lim_{x\to 0} \frac{u_1(x)}{d(x,\partial\Omega)|x|^{-\alpha_-(\gamma)}} = l_1 \quad \text{and} \quad \lim_{x\to 0} \frac{u_2(x)}{d(x,\partial\Omega)|x|^{-\alpha_-(\gamma)}} = l_2.$$

Since $u = u_1 - u_2$, we get Theorem 4.1 by taking $l := l_1 - l_2$. □

Here is an immediate consequence.

Corollary 4.7. *Suppose $\gamma < \gamma_H(\Omega)$ and consider the first eigenvalue of L_γ, i.e.,*

$$\lambda_1(\Omega,\gamma) := \inf_{u\in D^{1,2}(\Omega)\setminus\{0\}} \frac{\int_\Omega \big(|\nabla u|^2 - u^2\gamma/|x|^2\big)\,dx}{\int_\Omega u^2\,dx} > 0.$$

If $u_0 \in D^{1,2}(\Omega)\setminus\{0\}$ is a minimizer, then there exists $A \neq 0$ such that

$$u_0(x) \sim_{x\to 0} A\frac{d(x,\partial\Omega)}{|x|^{\alpha_-(\gamma)}}.$$

Proof. The existence of a minimizer u_0 that doesn't change sign is standard. The Euler–Lagrange equation is $-\Delta u - u\gamma/|x|^2 = ku$ for some $k \in \mathbb{R}$. We then apply Theorem 4.1. □

5. Regularity of solutions for related nonlinear variational problems

This section is devoted to the proof of the following key result.

Theorem 5.1 (optimal regularity and generalized Hopf's lemma). *Fix $\gamma < \frac{1}{4}n^2$ and let $f : \Omega\times\mathbb{R}\to\mathbb{R}$ be a Carathéodory function such that*

$$|f(x,v)| \leq C|v|\left(1 + \frac{|v|^{2^\star(s)-2}}{|x|^s}\right) \quad \text{for all } x\in\Omega \text{ and } v\in\mathbb{R}.$$

Let $u \in D^{1,2}(\Omega)_{\mathrm{loc},0}$ be a weak solution of

$$-\Delta u - \frac{\gamma + O(|x|^\tau)}{|x|^2}u = f(x,u) \quad \text{in } D^{1,2}(\Omega)_{\mathrm{loc},0} \tag{5-1}$$

for some $\tau > 0$. Then, there exists $K \in \mathbb{R}$ such that

$$\lim_{x\to 0} \frac{u(x)}{d(x,\partial\Omega)|x|^{-\alpha_-(\gamma)}} = K. \tag{5-2}$$

Moreover, if $u \geq 0$ and $u \not\equiv 0$, we have that $K > 0$.

Note that when $f \equiv 0$, this is nothing but Theorem 4.1. The result can be viewed as a generalization of Hopf's lemma in the following sense: when $\gamma = 0$ (and then $\alpha_-(\gamma) = 0$), the classical Nash–Moser regularity scheme yields $u \in C^1_{\mathrm{loc}}$, and when $u \geq 0$, $u \not\equiv 0$, Hopf's comparison principle yields $\partial_\nu u(0) < 0$, which is a reformulation of (5-2) when $\alpha_-(\gamma) = 0$.

The following lemma will be of frequent use in the sequel.

Lemma 5.2. *Let $f : \Omega \times \mathbb{R} \to \mathbb{R}$ be as in the statement of Theorem 5.1, and consider $u \in D^{1,2}(\Omega)_{\mathrm{loc},0}$ such that* (5-1) *holds. Assume that for some $C > 0$,*

$$|u(x)| \le C|x|^{1-\alpha_-(\gamma)} \quad \text{for } x \to 0,\ x \in \Omega. \tag{5-3}$$

Then, u satisfies the conclusion of Lemma 4.6.

Proof. Assume that (5-3) holds. We claim that we can assume that for some $\tau > 0$,

$$-\Delta u - \frac{\gamma + O(|x|^\tau)}{|x|^2} u = 0 \quad \text{in } D^{1,2}(\Omega)_{\mathrm{loc},0}. \tag{5-4}$$

Indeed, we have as $x \to 0$,

$$|f(x,u)| \le C|u|\big(1 + |x|^{-s}|x|^{-(2^\star(s)-2)(\alpha_-(\gamma)-1)}\big) \le C\frac{|u|}{|x|^2}\big(|x|^2 + |x|^{(2^\star(s)-2)(n/2-\alpha_-(\gamma))}\big) = O\left(|x|^{\tau'}\frac{u}{|x|^2}\right)$$

for some $\tau' > 0$. Plugging this inequality into (5-1) and replacing τ by $\min\{\tau, \tau'\}$ yields (5-4). The lemma now follows from Lemma 4.6. □

Proof of Theorem 5.1. We let here $u \in D^{1,2}(\Omega)_{\mathrm{loc},0}$ be a solution to (5-1); that is,

$$-\Delta u - \frac{\gamma + O(|x|^\tau)}{|x|^2} u = f(x,u) \quad \text{weakly in } D^{1,2}(\Omega)_{\mathrm{loc},0} \tag{5-5}$$

for some $\tau > 0$. We shall first use the classical De Giorgi–Nash–Moser iterative scheme (see [Gilbarg and Trudinger 1998; Hebey 1997] for expositions in book form). We skip most of the computations and refer to [Ghoussoub and Robert 2006a, Proposition A.1] for the details. We fix $\delta_0 > 0$ such that

(i) there exists $\tilde\eta \in C^\infty(B_{4\delta_0}(0))$ such that $\tilde\eta(x) = 1$ for $x \in B_{2\delta_0}(0)$,

(ii) $\tilde\eta u \in D^{1,2}(\Omega)$, and

(iii) u is a weak solution to (5-5) when tested on $\tilde\eta\varphi$ with $\varphi \in D^{1,2}(\Omega)$ (see the definition of weak solution given in the preceding section).

The proof goes through four steps.

Step 1: Let $\beta \ge 1$ be such that $4\beta/(\beta+1)^2 > 4\gamma/n^2$. Assume that $u \in L^{\beta+1}(\Omega \cap B_{\delta_0}(0))$. We claim that

$$u \in L^{n/(n-2)(\beta+1)}(\Omega \cap B_{\delta_0}(0)). \tag{5-6}$$

Indeed, fix $\beta \ge 1,\ L > 0$, and define $G_L, H_L : \mathbb{R} \to \mathbb{R}$ as

$$G_L(t) := \begin{cases} |t|^{\beta-1}t & \text{if } |t| \le L, \\ \beta L^{\beta-1}(t-L) + L^\beta & \text{if } t \ge L, \\ \beta L^{\beta-1}(t+L) - L^\beta & \text{if } t \le -L \end{cases} \tag{5-7}$$

and

$$H_L(t) := \begin{cases} |t|^{(\beta-1)/2}t & \text{if } |t| \le L, \\ \frac{1}{2}(\beta+1)L^{(\beta-1)/2}(t-L) + L^{(\beta+1)/2} & \text{if } t \ge L, \\ \frac{1}{2}(\beta+1)L^{(\beta-1)/2}(t+L) - L^{(\beta+1)/2} & \text{if } t \le -L. \end{cases} \tag{5-8}$$

As is easily checked,

$$0 \le tG_L(t) \le H_L(t)^2 \quad \text{and} \quad G_L'(t) = \frac{4\beta}{(\beta+1)^2}(H_L'(t))^2 \tag{5-9}$$

for all $t \in \mathbb{R}$ and all $L > 0$. We fix $\delta > 0$ small, which will be chosen later. We let $\eta \in C_c^\infty(\mathbb{R}^n)$ be such that $\eta(x) = 1$ for $x \in B_{\delta/2}(0)$ and $\eta(x) = 0$ for $x \in \mathbb{R}^n \setminus B_\delta(0)$. Multiplying equation (5-5) with $\eta^2 G_L(u) \in D^{1,2}(\Omega)$, we get that

$$\int_\Omega (\nabla u, \nabla(\eta^2 G_L(u)))\,dx - \int_\Omega \frac{\gamma + O(|x|^\tau)}{|x|^2}\eta^2 u G_L(u)\,dx = \int_\Omega f(x,u)\eta^2 G_L(u)\,dx. \tag{5-10}$$

Integrating by parts, and using formulae (5-7)–(5-9) (see [Ghoussoub and Robert 2006a] for details) yields

$$\int_\Omega \big(\nabla u, \nabla(\eta^2 G_L(u))\big)\,dx = \frac{4\beta}{(\beta+1)^2}\int_\Omega \big(|\nabla(\eta H_L(u))|^2 - \eta(-\Delta)\eta H_L(u)^2\big)\,dx + \int_\Omega -\Delta(\eta^2) J_L(u)\,dx, \tag{5-11}$$

where $J_L(t) := \int_0^t G_L(\tau)\,d\tau$. This identity and (5-10) yield

$$\begin{aligned}\frac{4\beta}{(\beta+1)^2}\int_\Omega |\nabla(\eta H_L(u))|^2\,dx - \int_\Omega \frac{\gamma + O(|x|^\tau)}{|x|^2}\eta^2 u G_L(u)\,dx \\ \le \int_\Omega |-\Delta(\eta^2)|\cdot|J_L(u)|\,dx + C(\beta,\delta)\int_{\Omega\cap B_\delta(0)} |H_L(u)|^2\,dx + C\int_\Omega \frac{|u|^{2^\star(s)-2}}{|x|^s}(\eta H_L(u))^2\,dx.\end{aligned} \tag{5-12}$$

Hölder's inequality and the Sobolev constant given in (1-16) yield

$$\begin{aligned}\int_\Omega \frac{|u|^{2^\star(s)-2}}{|x|^s}(\eta H_L(u))^2\,dx &\le \left(\int_{\Omega\cap B_\delta(0)} \frac{|u|^{2^\star(s)}}{|x|^s}\,dx\right)^{(2^\star(s)-2)/2^\star(s)} \left(\int_\Omega \frac{|\eta H_L(u)|^{2^\star(s)}}{|x|^s}\,dx\right)^{2/2^\star(s)} \\ &\le \left(\int_{\Omega\cap B_\delta(0)} \frac{|u|^{2^\star(s)}}{|x|^s}\,dx\right)^{(2^\star(s)-2)/2^\star(s)} \cdot \frac{1}{\mu_{0,s}(\Omega)}\int_\Omega |\nabla(\eta H_L(u))|^2\,dx.\end{aligned}$$

Plugging this estimate into (5-12) and defining $\gamma_+ := \max\{\gamma, 0\}$ yields

$$\begin{aligned}\frac{4\beta}{(\beta+1)^2}\int_\Omega \|\nabla(\eta H_L(u))\|^2\,dx - (\gamma_+ + C\delta^\tau)\int_\Omega \frac{(\eta H_L(u))^2}{|x|^2}\,dx \\ \le C(\beta,\delta)\int_{\Omega\cap B_\delta(0)} \big(|H_L(u)|^2 + |J_L(u)|\big)\,dx + \alpha(\delta)\int_\Omega |\nabla(\eta H_L(u))|^2\,dx,\end{aligned}$$

where

$$\alpha(\delta) := C\left(\int_{\Omega\cap B_\delta(0)} \frac{|u|^{2^\star(s)}}{|x|^s}\,dx\right)^{(2^\star(s)-2)/2^\star(s)} \cdot \frac{1}{\mu_{0,s}(\Omega)},$$

so that

$$\lim_{\delta\to 0}\alpha(\delta) = 0.$$

It follows from Hardy's inequality that

$$\frac{n^2}{4}\int_\Omega \frac{(\eta H_L(u))^2}{|x|^2}\,dx \le (1+\epsilon(\delta))\int_\Omega |\nabla(\eta H_L(u))|^2\,dx,$$

where $\lim_{\delta\to 0}\epsilon(\delta)=0$. Therefore, we get that

$$\left(\frac{4\beta}{(\beta+1)^2}-\alpha(\delta)-(\gamma_+ + C\delta^\tau)\frac{4}{n^2}(1+\epsilon(\delta))\right)\int_\Omega |\nabla(\eta H_L(u))|^2\,dx$$
$$\le C(\beta,\delta)\int_{\Omega\cap B_\delta(0)}\left(|H_L(u)|^2+|J_L(u)|\right)dx \le C(\beta,\delta)\int_{B_\delta(0)\cap\Omega}|u|^{\beta+1}\,dx.$$

Let $\delta\in(0,\delta_0)$ be such that

$$\frac{4\beta}{(\beta+1)^2}-\alpha(\delta)-(\gamma_+ + C\delta^\tau)\frac{4}{n^2}(1+\epsilon(\delta))>0.$$

This is possible since $4\beta/(\beta+1)^2>4\gamma/n^2$. Using Sobolev's embedding, we then get that

$$\left(\int_{B_{\delta/2}(0)\cap\Omega}|H_L(u)|^{2^\star}\,dx\right)^{2/2^\star}\le\left(\int_{\mathbb{R}^n}|\eta H_L(u)|^{2^\star}\,dx\right)^{2/2^\star}$$
$$\le\mu_{0,0}(\Omega)^{-1}\int_\Omega|\nabla(\eta H_L(u))|^2\,dx\le C(\beta,\delta,\gamma)\int_{B_\delta(0)\cap\Omega}|u|^{\beta+1}\,dx.$$

Since $u\in L^{\beta+1}(B_{\delta_0}(0)\cap\Omega)$, let $L\to+\infty$ and use Fatou's lemma to obtain that $u\in L^{(2^\star/2)(\beta+1)}(B_{\delta/2}(0)\cap\Omega)$. The standard iterative scheme then yields $u\in C^1(\overline{\Omega}\cap B_{\delta_0}(0)\setminus\{0\})$. Therefore $u\in L^{(2^\star/2)(\beta+1)}(B_{\delta_0}(0)\cap\Omega)$.

Step 2: We now show that

$$\text{if } \gamma\le 0,\ \text{then } u\in L^p(\Omega\cap B_\delta(0)) \text{ for all } p\ge 1, \tag{5-13}$$

$$\text{if } \gamma>0,\ \text{then } u\in L^p(\Omega\cap B_\delta(0)) \text{ for all } p\in\left(1,\frac{n}{n-2}\frac{n}{\alpha_-(\gamma)}\right). \tag{5-14}$$

The case $\gamma\le 0$ is standard, so we only consider the case where $\gamma>0$. Fix $p\ge 2$ and set $\beta:=p-1$. We have

$$\frac{4\beta}{(\beta+1)^2}>\frac{4}{n^2}\gamma \iff \frac{n}{\alpha_+(\gamma)}<p<\frac{n}{\alpha_-(\gamma)}.$$

Since $\alpha_+(\gamma)>\frac{1}{2}n$ and $p\ge 2$,

$$\frac{4\beta}{(\beta+1)^2}>\frac{4}{n^2}\gamma \iff p<\frac{n}{\alpha_-(\gamma)}.$$

Therefore, it follows from Step 1 that if $u\in L^p(\Omega\cap B_{\delta_0})$, with $p<n/\alpha_-(\gamma)$, then $u\in L^{pn/(n-2)}(\Omega\cap B_{\delta_0})$. Since $u\in L^2(\Omega\cap B_{\delta_0})$, (5-14) follows.

Step 3: We claim that for any $\lambda>0$,

$$|x|^{(n-2)/2}|u(x)|=O(|x|^{(n-2)/n\left(n/2-\max\{\alpha_-(\gamma),0\}-\lambda\right)}) \quad \text{as } x\to 0. \tag{5-15}$$

Indeed, take $p\in\left(2^\star, n^2/((n-2)\alpha_-(\gamma))\right)$ if $\gamma>0$, and $p>2^\star$ if $\gamma\le 0$. This is possible since $2^\star=2n/(n-2)$ and $\alpha_-(\gamma)<\frac{1}{2}n$. We fix a sequence $(\epsilon_i)_i\in(0,+\infty)$ such that $\lim_{i\to+\infty}\epsilon_i=0$ and we fix a chart φ as in (4-7) to (4-12). For any $i\in\mathbb{N}$, we define

$$u_i(x):=\epsilon_i^{n/p}u(\varphi(\epsilon_i x)) \quad \text{for all } x\in\widetilde{B}_{\delta/\epsilon_i}.$$

Equation (5-5) then can be written as

$$-\Delta_{g_i} u_i - \frac{\epsilon_i^2(\gamma + O(\epsilon_i^\tau |x|^\tau))}{|\varphi(\epsilon_i x)|^2} u_i = f_i(x, u_i), \quad u_i = 0 \text{ on } \partial\mathbb{R}^n_+ \cap \widetilde{B}_{\delta/\epsilon_i}, \tag{5-16}$$

where $g_i(x) := \varphi^\star \mathrm{Eucl}(\epsilon_i x)$ and

$$|f_i(x, u_i)| \le C\epsilon_i^2 |u_i| + C\epsilon_i^{(2^\star(s)-2)((n-2)/2-n/p)} |x|^{-s} |u_i|^{2^\star(s)-1} \quad \text{in } \widetilde{B}_{\delta/\epsilon_i}.$$

We fix $R > 0$ and define $\omega_R := (\widetilde{B}_R \setminus \widetilde{B}_{R^{-1}}) \cap \mathbb{R}^n_+$. With our choice of p above and using (5-14), we get that

$$\|u_i\|_{L^p(\omega_R)} \le C, \tag{5-17}$$

and

$$|f_i(x, u_i)| \le C_R |u_i| + C_R |u_i|^{2^\star(s)-1} \quad \text{for all } x \in \omega_R. \tag{5-18}$$

Fix $q \ge p > 2^\star$. It follows from elliptic regularity that

$$\|u_i\|_{L^q(\omega_R)} \le C \implies \begin{cases} \|u_i\|_{L^{q'}(\omega_{R/2})} \le C' & \text{if } q < \frac{1}{2}n(2^\star(s)-1), \\ \|u_i\|_{L^r(\omega_{R/2})} \le C' & \text{for all } r \ge 1 \text{ if } q = \frac{1}{2}n(2^\star(s)-1), \\ \|u_i\|_{L^\infty(\omega_{R/2})} \le C' & \text{if } q > \frac{1}{2}n(2^\star(s)-1), \end{cases}$$

where

$$\frac{1}{q'} = \frac{2^\star(s)-1}{q} - \frac{2}{n}$$

and the constants C, C' are uniform with respect to i. It then follows from the standard bootstrap iterative argument and the initial bound (5-17) that $\|u_i\|_{L^\infty(\omega_{R/4})} \le C'$. Taking $R > 0$ large enough and going back to the definition of u_i, we get that for all $i \in \mathbb{N}$,

$$|x|^{n/p} |u(x)| \le C \quad \text{for all } x \in \Omega \cap B_{2\epsilon_i}(0) \setminus B_{\epsilon_i/2}(0).$$

Since this holds for any sequence $(\epsilon_i)_i$, we get that $|x|^{n/p} |u(x)| \le C$ around 0 for any

$$2^\star < p < \frac{n^2}{(n-2)\alpha_-(\gamma)}$$

when $\gamma > 0$. Letting p go to $n^2/((n-2)\alpha_-(\gamma))$ yields (5-15) when $\gamma > 0$. For $\gamma \le 0$, we let $p \to +\infty$.

To finish the proof of Theorem 5.1, we rewrite equation (5-5) as

$$-\Delta u - \frac{a(x)}{|x|^2} u = 0,$$

where for $x \in \Omega$,

$$\begin{aligned} a(x) &= \gamma + O(|x|^\tau) + O(|x|^2) + O(|x|^{2-s} |u|^{2^\star(s)-2}) \\ &= \gamma + O(|x|^\tau) + O(|x|^2) + O(|x|^{(n-2)/2} |u(x)|)^{2^\star(s)-2}. \end{aligned}$$

Since $\alpha_-(\gamma) < \frac{1}{2}n$, it then follows from (5-15) that there exists $\tau' > 0$ such that $a(x) = \gamma + O(|x|^{\tau'})$ as $x \to 0$. We are therefore back to the linear case; hence we can apply Theorem 4.1 and deduce Theorem 5.1. □

As a consequence we get the following result, which will be crucial for the sequel.

Corollary 5.3. *Suppose $u \in D^{1,2}(\mathbb{R}^n_+)$, $u \geq 0$, $u \not\equiv 0$ is a weak solution of*

$$-\Delta u - \frac{\gamma}{|x|^2}u = \frac{u^{2^\star-1}}{|x|^s} \quad \text{in } \mathbb{R}^n_+.$$

Then, there exist $K_1, K_2 > 0$ such that

$$u(x) \sim_{x\to 0} K_1 \frac{x_1}{|x|^{\alpha_-(\gamma)}} \quad \text{and} \quad u(x) \sim_{|x|\to+\infty} K_2 \frac{x_1}{|x|^{\alpha_+(\gamma)}}. \tag{5-19}$$

Proof. Theorem 5.1 yields the behavior when $x \to 0$. The Kelvin transform $\hat{u}(x) := |x|^{2-n}u(x/|x|^2)$ is a solution to the same equation in $D^{1,2}(\mathbb{R}^n_+)$, and its behavior at 0 is given by Theorem 5.1. Going back to u yields the behavior at ∞. □

6. Profile around 0 of positive singular solutions of $L_\gamma u = a(x)u$

In this section we describe the profile of any positive solution — variational or not — of linear equations involving L_γ. Here is the main result of this section.

Theorem 6.1. *Let $u \in C^2(B_\delta(0) \cap (\overline{\Omega} \setminus \{0\}))$ be such that*

$$\begin{cases} -\Delta u - \dfrac{\gamma + O(|x|^\tau)}{|x|^2}u = 0 & \text{in } \Omega \cap B_\delta(0), \\ u > 0 & \text{in } \Omega \cap B_\delta(0), \\ u = 0 & \text{on } (\partial\Omega \cap B_\delta(0)) \setminus \{0\}. \end{cases} \tag{6-1}$$

Then, there exists $K > 0$ such that either

$$u(x) \sim_{x\to 0} K \frac{d(x,\partial\Omega)}{|x|^{\alpha_-(\gamma)}} \quad \text{or} \quad u(x) \sim_{x\to 0} K \frac{d(x,\partial\Omega)}{|x|^{\alpha_+(\gamma)}}.$$

In the first case, the solution $u \in D^{1,2}(\Omega)_{\mathrm{loc},0}$ is a variational solution to (6-1).

It is worth noting that Pinchover [1994] tackled similar issues. The proof of Theorem 6.1 will require two additional results. The first is a Harnack-type result.

Proposition 6.2. *Let Ω be a smooth bounded domain of $\mathbb{R}^n$, and let $a \in L^\infty(\Omega)$ be such that $\|a\|_\infty \leq M$ for some $M > 0$. Assume U is an open subset of $\mathbb{R}^n$ and consider $u \in C^2(U \cap \overline{\Omega})$ to be a solution of*

$$\begin{cases} -\Delta_g u + au = 0 & \text{in } U \cap \Omega, \\ u \geq 0 & \text{in } U \cap \Omega, \\ u = 0 & \text{on } U \cap \partial\Omega. \end{cases}$$

Here g is a smooth metric on U. If $U' \Subset U$ is such that $U' \cap \Omega$ is connected, then there exists $C > 0$ depending only on Ω, U', M and g such that

$$\frac{u(x)}{d(x,\partial\Omega)} \leq C \frac{u(y)}{d(y,\partial\Omega)} \quad \text{for all } x, y \in U' \cap \Omega. \tag{6-2}$$

Proof. We first prove a local result. The global result will be the consequence of a covering of U'. Fix $x_0 \in \partial\Omega$. For $\delta > 0$ small enough, there exists a smooth open domain W such that

$$B_\delta(x_0) \cap \Omega \subset W \subset B_{2\delta}(x_0) \cap \Omega \quad \text{and} \quad B_\delta(x_0) \cap \partial W = B_\delta(x_0) \cap \partial\Omega. \tag{6-3}$$

Let G be the Green's function of $-\Delta_g + a$ with Dirichlet boundary condition on W, then its representation formula reads as

$$u(x) = \int_{\partial W} u(\sigma)(-\partial_{\nu,\sigma} G(x,\sigma))\, d\sigma = \int_{\partial W \setminus \partial\Omega} u(\sigma)(-\partial_{\nu,\sigma} G(x,\sigma))\, d\sigma \tag{6-4}$$

for all $x \in W$, where $\partial_{\nu,\sigma} G(x,\sigma)$ is the normal derivative of $y \mapsto G(x,y)$ at $\sigma \in \partial W$. Estimates of the Green's function (see [Robert 2010; Ghoussoub and Robert 2006a]) yield the existence of $C > 0$ such that for all $x \in W$ and $\sigma \in \partial W$,

$$\frac{1}{C}\frac{d(x,\partial W)}{|x-\sigma|^n} \le -\partial_{\nu,\sigma} G(x,\sigma) \le C\frac{d(x,\partial W)}{|x-\sigma|^n}.$$

It follows from (6-3) that there exists $C(\delta) > 0$ such that for all $x \in B_{\delta/2}(x_0) \cap \Omega \subset W$ and $\sigma \in \partial W \setminus \partial\Omega$,

$$\frac{1}{C(\delta)} d(x,\partial W) \le -\partial_{\nu,\sigma} G(x,\sigma) \le C(\delta)\, d(x,\partial W).$$

Since u vanishes on $\partial\Omega$, it then follows from (6-4) that for all $x \in B_{\delta/2}(x_0) \cap \Omega$,

$$\frac{1}{C(\delta)} d(x,\partial W) \int_{\partial W} u(\sigma)\, d\sigma \le u(x) \le C(\delta)\, d(x,\partial W) \int_{\partial W} u(\sigma)\, d\sigma.$$

It is easy to check that under the assumption (6-3), we have that $d(x,\partial\Omega) = d(x,\partial W)$. Therefore, we have for all $x \in B_{\delta/2}(x_0) \cap \Omega$,

$$\frac{1}{C(\delta)} \int_{\partial W} u(\sigma)\, d\sigma \le \frac{u(x)}{d(x,\partial\Omega)} \le C(\delta) \int_{\partial W} u(\sigma)\, d\sigma.$$

Since these lower and upper bounds are independent of x, we get inequality (6-2) for any $x, y \in B_{\delta/2}(x_0) \cap \Omega$.

The general case is a consequence of a covering of $U' \cap \Omega$ by finitely many balls. Note that for balls intersecting $\partial\Omega$, we apply the preceding result, while for balls not intersecting $\partial\Omega$, we apply the classical Harnack inequality. This completes the proof of Proposition 6.2. □

Proof of Theorem 6.1. Let u be a solution of (6-1) as in the statement of Theorem 6.1. We claim that

$$u(x) = O(d(x,\partial\Omega)|x|^{-\alpha_+(\gamma)}) \quad \text{for } x \to 0,\ x \in \Omega. \tag{6-5}$$

Indeed, otherwise we can assume that

$$\limsup_{x\to 0} \frac{u(x)}{d(x,\partial\Omega)|x|^{-\alpha_+(\gamma)}} = +\infty. \tag{6-6}$$

In particular, there exists $(x_k)_k \in \Omega$ such that for all $k \in \mathbb{N}$,

$$\lim_{k\to+\infty} x_k = 0 \quad \text{and} \quad \frac{u(x_k)}{d(x_k,\partial\Omega)|x_k|^{-\alpha_+(\gamma)}} \ge k. \tag{6-7}$$

We claim that there exists $C > 0$ such that

$$\frac{u(x)}{d(x, \partial\Omega)|x|^{-\alpha_+(\gamma)}} \geq Ck \quad \text{for all } x \in \Omega \cap \partial B_{r_k}(0), \text{ with } r_k := |x_k| \to 0. \tag{6-8}$$

We prove the claim by using the Harnack inequality (6-2): first take the chart φ at 0 as in (4-7), and define

$$u_k(x) := u \circ \varphi(r_k x) \quad \text{for } x \in \mathbb{R}^n_+ \cap B_3(0) \setminus \{0\}.$$

Equation (6-1) can be written as

$$-\Delta_{g_k} u_k + a_k u_k = 0 \quad \text{in } \mathbb{R}^n_+ \cap B_3(0) \setminus \{0\}, \tag{6-9}$$

with

$$a_k(x) := -r_k^2 \frac{\gamma + O(r_k^\tau |x|^\tau)}{|\varphi(r_k x)|^2}.$$

In particular, there exists $M > 0$ such that $|a_k(x)| \leq M$ for all $x \in \mathbb{R}^n_+ \cap B_3(0) \setminus \bar{B}_{1/3}(0)$. Since $u_k \geq 0$, the Harnack inequality (6-2) yields the existence of $C > 0$ such that

$$\frac{u_k(y)}{y_1} \geq C \frac{u_k(x)}{x_1} \quad \text{for all } x, y \in \mathbb{R}^n_+ \cap B_2(0) \setminus \bar{B}_{1/2}(0). \tag{6-10}$$

Let $\tilde{x}_k \in \mathbb{R}^n_+$ be such that $x_k = \varphi(r_k \tilde{x}_k)$. In particular, $|\tilde{x}_k| = 1 + o(1)$ as $k \to +\infty$. It then follows from (6-7), (6-9) and (6-10) that

$$\frac{u \circ \varphi(r_k y)}{d(\varphi(r_k y), \partial\Omega)} \geq C \cdot k \quad \text{for all } y \in \mathbb{R}^n_+ \cap B_2(0) \setminus \bar{B}_{1/2}(0).$$

In particular, (6-8) holds.

We let now W be a smooth domain such that (4-26) holds for $r > 0$ small enough. Take the supersolution $u^{(d)}_{\alpha_+(\gamma),-}$ defined in Proposition 4.5. We have that

$$u(x) \geq \frac{C \cdot k}{2} u^{(d)}_{\alpha_+(\gamma),-}(x) \quad \text{for all } x \in W \cap \partial B_{r_k}(0).$$

Since $u^{(d)}_{\alpha_+(\gamma),-}$ vanishes on ∂W, we have $u(x) \geq \frac{1}{2}(C \cdot k) u^{(d)}_{\alpha_+(\gamma),-}(x)$ for all $x \in \partial(W \cap B_{r_k}(0))$. Moreover, we have that

$$-\Delta u^{(d)}_{\alpha_+(\gamma),-} - \frac{\gamma + O(|x|^\tau)}{|x|^2} u^{(d)}_{\alpha_+(\gamma),-} < 0 = -\Delta u - \frac{\gamma + O(|x|^\tau)}{|x|^2} u \quad \text{on } W.$$

Up to taking r even smaller, it follows from the coercivity of the operator and the maximum principle that

$$u(x) \geq \frac{C \cdot k}{2} u^{(d)}_{\alpha_+(\gamma),-}(x) \quad \text{for all } x \in W \cap B_{r_k}(0). \tag{6-11}$$

For any $x \in W$, we let $k_0 \in \mathbb{N}$ such that $r_k < |x|$ for all $k \geq k_0$. It then follows from (6-11) that $u(x) \geq \frac{1}{2}(C \cdot k) u^{(d)}_{\alpha_+(\gamma),-}(x)$ for all $k \geq k_0$. Letting $k \to +\infty$ yields that $u^{(d)}_{\alpha_+(\gamma),-}(x)$ goes to zero for all $x \in W$. This is in contradiction with (4-29). Hence (6-6) does not hold, and therefore (6-5) holds.

A straightforward consequence of (6-5) and Lemma 5.2 is that there exists $l \in \mathbb{R}$ such that

$$\lim_{x\to 0} \frac{u(x)}{d(x,\partial\Omega)|x|^{-\alpha_+(\gamma)}} = l. \tag{6-12}$$

We now show the following lemma:

Lemma 6.3. *If*

$$\lim_{x\to 0} \frac{u(x)}{d(x,\partial\Omega)|x|^{-\alpha_+(\gamma)}} = 0,$$

then $u \in D^{1,2}(\Omega)_{\mathrm{loc},0}$ *and there exists* $K > 0$ *such that* $u(x) \sim_{x\to 0} K d(x,\partial\Omega)/|x|^{\alpha_-(\gamma)}$.

Proof. We shall use Theorem 4.1. Take W as in (4-26) and let $\eta \in C^\infty(\mathbb{R}^n)$ be such that $\eta(x) = 0$ for $x \in B_{\delta/4}(0)$ and $\eta(x) = 1$ for $x \in \mathbb{R}^n \setminus B_{\delta/3}(0)$. Define

$$f(x) := \left(-\Delta - \frac{\gamma + O(|x|^\tau)}{|x|^2}\right)(\eta u) \quad \text{for } x \in W.$$

The function $f \in C^\infty(\overline{W})$ vanishes around 0. Let $v \in D^{1,2}(\Omega)$ be such that

$$\begin{cases} -\Delta v - \dfrac{\gamma + O(|x|^\tau)}{|x|^2} v = f & \text{in } W, \\ v = 0 & \text{on } \partial W. \end{cases}$$

Note again that for $r > 0$ small enough, $-\Delta - (\gamma + O(|x|^\tau))|x|^{-2}$ is coercive on W, and therefore, the existence of v is ensured for small r. Define

$$\tilde{u} := u - \eta u + v.$$

The properties of W and the definition of η and v yield

$$\begin{cases} -\Delta \tilde{u} - \dfrac{\gamma + O(|x|^\tau)}{|x|^2} \tilde{u} = 0 & \text{in } W, \\ \tilde{u} = 0 & \text{in } \partial W \setminus \{0\}. \end{cases}$$

Moreover, since $-\Delta v - (\gamma + O(|x|^\tau))|x|^{-2} v = 0$ around 0 and $v \in D^{1,2}(W)$, it follows from Theorem 4.1 that there exists $C > 0$ such that $|v(x)| \le C d(x, W)|x|^{-\alpha_-(\gamma)}$ for all $x \in W$. Therefore, we have that

$$\lim_{x\to 0} \frac{\tilde{u}(x)}{d(x,\partial\Omega)|x|^{-\alpha_+(\gamma)}} = 0. \tag{6-13}$$

It then follows from Lemma 5.2 that

$$\lim_{x\to 0} |x|^{\alpha_+(\gamma)} |\nabla \tilde{u}(x)| = 0. \tag{6-14}$$

Let $\psi \in C_c^\infty(W)$ and $w \in D^{1,2}(W)$ be such that

$$\begin{cases} -\Delta w - \dfrac{\gamma + O(|x|^\tau)}{|x|^2} w = \psi & \text{in } W, \\ w = 0 & \text{on } \partial W. \end{cases}$$

Since ψ vanishes around 0, it follows from Theorem 4.1 and Lemma 5.2 that

$$w(x) = O(d(x, \partial W)|x|^{-\alpha_-(\gamma)}) \quad \text{and} \quad |\nabla w(x)| = O(|x|^{-\alpha_-(\gamma)}) \quad \text{as } x \to 0. \tag{6-15}$$

Fix $\epsilon > 0$ small and integrate by parts, using that both $\tilde{u}$ and w vanish on ∂W, to get

$$\begin{aligned} 0 &= \int_{W\setminus B_\epsilon(0)} \left(-\Delta\tilde{u} - \frac{\gamma + O(|x|^\tau)}{|x|^2}\tilde{u}\right) w\, dx \\ &= \int_{W\setminus B_\epsilon(0)} \left(-\Delta w - \frac{\gamma + O(|x|^\tau)}{|x|^2} w\right)\tilde{u}\, dx + \int_{\partial(W\setminus B_\epsilon(0))} (-w\partial_\nu\tilde{u} + \tilde{u}\partial_\nu w)\, d\sigma \\ &= \int_{W\setminus B_\epsilon(0)} \psi\tilde{u}\, dx - \int_{\Omega\cap\partial B_\epsilon(0)} (-w\partial_\nu\tilde{u} + \tilde{u}\partial_\nu w)\, d\sigma. \end{aligned}$$

Using the limits and estimates (6-13), (6-14) and (6-15), and that ψ vanishes around 0, we get

$$0 = \int_{W\setminus B_\epsilon(0)} \psi\tilde{u}\, dx + o\big(\epsilon^{n-1}(\epsilon^{1-\alpha_-(\gamma)}\epsilon^{-\alpha_+(\gamma)} + \epsilon^{1-\alpha_+(\gamma)}\epsilon^{-\alpha_-(\gamma)})\big) = \int_{W\setminus B_\epsilon(0)} \psi\tilde{u}\, dx + o(1), \quad \text{as } \epsilon \to 0.$$

Therefore, we have $\int_W \psi\tilde{u}\, dx = 0$ for all $\psi \in C_c^\infty(W)$. Since $\tilde{u} \in L^p$ is smooth outside 0, we then get that $\tilde{u} \equiv 0$, and therefore $u = \eta u + v$. In particular, $u \in D^{1,2}(\Omega)_{\mathrm{loc},0}$ is a distributional positive solution to

$$-\Delta u - \frac{\gamma + O(|x|^\tau)}{|x|^2} u = 0$$

on W. It then follows from Theorem 4.1 that there exists $K > 0$ such that $u(x) \sim_{x\to 0} K d(x, \partial\Omega)/|x|^{\alpha_-(\gamma)}$. This proves Lemma 6.3. □

Combining Lemma 6.3 with (6-12) completes the proof of Theorem 6.1. □

As a consequence of Theorem 6.1, we improve Lemma 4.2 as follows.

Proposition 6.4. *Let $u \in C^2(\overline{\mathbb{R}^n_+} \setminus \{0\})$ be a nonnegative function such that*

$$\begin{cases} -\Delta u - \dfrac{\gamma}{|x|^2} u = 0 & \text{in } \mathbb{R}^n_+, \\ u = 0 & \text{on } \partial\mathbb{R}^n_+. \end{cases} \tag{6-16}$$

Then there exist $\lambda_-, \lambda_+ \geq 0$ such that

$$u(x) = \lambda_- x_1 |x|^{-\alpha_-(\gamma)} + \lambda_+ x_1 |x|^{-\alpha_+(\gamma)} \quad \text{for all } x \in \mathbb{R}^n_+.$$

Proof. Without loss of generality, we assume that $u \not\equiv 0$, so that $u > 0$. We consider the Kelvin transform of u defined by $\hat{u}(x) := |x|^{2-n} u(x/|x|^2)$ for all $x \in \mathbb{R}^n_+$. Both u and $\hat{u}$ are then nonnegative solutions of (6-16). It follows from Theorem 6.1 that, after performing back the Kelvin transform, there exist $\alpha_1, \alpha_2 \in \{\alpha_+(\gamma), \alpha_-(\gamma)\}$ such that

$$\lim_{x\to 0} \frac{u(x)}{x_1|x|^{-\alpha_1}} = l_1 > 0 \quad \text{and} \quad \lim_{|x|\to\infty} \frac{u(x)}{x_1|x|^{-\alpha_2}} = l_2 > 0.$$

If $\alpha_1 \le \alpha_2$, then $u(x) \le Cx_1|x|^{-\alpha_1}$ for all $x \in \mathbb{R}^n_+$. The result then follows from Lemma 4.2. If $\alpha_1 > \alpha_2$, then $\alpha_1 = \alpha_+(\gamma)$ and $\alpha_2 = \alpha_-(\gamma)$. We define

$$\tilde{u}(x) := u(x) - l_1 x_1 |x|^{-\alpha_+(\gamma)} \quad \text{for all } x \in \mathbb{R}^n_+.$$

to obtain that $-\Delta\tilde{u} - \tilde{u}\gamma/|x|^2 = 0$ in $\mathbb{R}^n_+$, $\tilde{u} = 0$ on $\partial\mathbb{R}^n_+$, and $\tilde{u}(x) = o(x_1|x|^{-\alpha_+(\gamma)})$ as $x \to 0$. Arguing as in the proof of Lemma 6.3, we get that $\tilde{u} \in D^{1,2}(\mathbb{R}^n_+)_{\mathrm{loc},0}$ and $\tilde{u}(x) = O(x_1|x|^{-\alpha_-(\gamma)})$ as $x \to 0$. Moreover, we have that $\tilde{u}(x) = (l_2 + o(1))x_1|x|^{-\alpha_-(\gamma)}$ as $|x| \to +\infty$; therefore $\tilde{u}(x) > 0$ for $|x| \gg 1$. Since $\tilde{u} \in D^{1,2}(\mathbb{R}^n_+)_{\mathrm{loc},0}$, the comparison principle then yields $\tilde{u} > 0$ everywhere. We also have that $\tilde{u}(x) \le Cx_1|x|^{-\alpha_-(\gamma)}$ for all $x \in \mathbb{R}^n_+$. It then follows from Lemma 4.2 that there exists $\lambda_- \ge 0$ such that $\tilde{u}(x) = \lambda_- x_1|x|^{-\alpha_-(\gamma)}$ for all $x \in \mathbb{R}^n_+$, from which Proposition 6.4 follows. □

7. The Hardy singular boundary mass of a domain Ω when 0 ∈ ∂Ω

We shall proceed in the following theorem to define the mass of a smooth bounded domain Ω of $\mathbb{R}^n$ such as $0 \in \partial\Omega$. It will involve the expansion of positive singular solutions of the Dirichlet boundary problem $L_\gamma u = 0$.

Theorem 7.1. *Let Ω be a smooth bounded domain Ω of $\mathbb{R}^n$ such as $0 \in \partial\Omega$, and assume that $\frac{1}{4}(n^2-1) < \gamma < \gamma_H(\Omega)$. Then, up to multiplication by a positive constant, there exists a unique function $H \in C^2(\overline{\Omega}\setminus\{0\})$ such that*

$$\begin{cases} -\Delta H - \dfrac{\gamma}{|x|^2} H = 0 & \text{in } \Omega, \\ H > 0 & \text{in } \Omega, \\ H = 0 & \text{on } \partial\Omega \setminus \{0\}. \end{cases} \tag{7-1}$$

Moreover, there exists $c_1 > 0$ and $c_2 \in \mathbb{R}$ such that

$$H(x) = c_1 \frac{d(x,\partial\Omega)}{|x|^{\alpha_+(\gamma)}} + c_2 \frac{d(x,\partial\Omega)}{|x|^{\alpha_-(\gamma)}} + o\left(\frac{d(x,\partial\Omega)}{|x|^{\alpha_-(\gamma)}}\right) \quad \text{as } x \to 0. \tag{7-2}$$

The quantity $m_\gamma(\Omega) := c_2/c_1 \in \mathbb{R}$, which is independent of the choice of H satisfying (7-1)*, will be called the Hardy b-mass of Ω associated to L_γ.*

Proof. First, we start by constructing a singular solution H_0 for (7-1). For that, consider $u_{\alpha_+(\gamma)}$ as in (4-14) and let $\eta \in C^\infty_c(\mathbb{R}^n)$ be such that $\eta(x) = 1$ for $x \in B_{\delta/2}(0)$ and $\eta(x) = 0$ for $x \in \mathbb{R}^n \setminus B_\delta(0)$. Set

$$f := -\Delta(\eta u_{\alpha_+(\gamma)}) - \frac{\gamma}{|x|^2}(\eta u_{\alpha_+(\gamma)}) \quad \text{in } \overline{\Omega} \setminus \{0\}.$$

It follows from (4-19) and (4-5) that f is smooth outside 0 and that

$$f(x) = O\big(d(x,\partial\Omega)|x|^{-\alpha_+(\gamma)-1}\big) = O(|x|^{-\alpha_+(\gamma)}) \quad \text{in } \Omega \cap B_{\delta/2}(0).$$

Since $\gamma > \frac{1}{4}(n^2-1)$, we have that $\alpha_+(\gamma) < \frac{1}{2}(n+1)$, and therefore $f \in L^{2n/(n+2)}(\Omega) = (L^{2^\star}(\Omega))' \subset (D^{1,2}(\Omega))'$. It then follows from the coercivity assumption $\gamma < \gamma_H(\Omega)$ that there exists $v \in D^{1,2}(\Omega)$ such that

$$-\Delta v - \frac{\gamma}{|x|^2} v = f \quad \text{in } (D^{1,2}(\Omega))'.$$

Let $v_1, v_2 \in D^{1,2}(\Omega)$ be such that

$$-\Delta v_1 - \frac{\gamma}{|x|^2} v_1 = f_+ \quad \text{and} \quad -\Delta v_2 - \frac{\gamma}{|x|^2} v_2 = f_- \quad \text{in } (D^{1,2}(\Omega))'. \tag{7-3}$$

In particular, $v = v_1 - v_2$ and $v_1, v_2 \in C^1(\overline{\Omega} \setminus \{0\})$, and they vanish on $\partial\Omega \setminus \{0\}$.

Assume that $f_+ \not\equiv 0$. Since $f_+ \geq 0$, the comparison principle yields $v_1 > 0$ on $\Omega \setminus \{0\}$ and $\partial_\nu v_1 < 0$ on $\partial\Omega \setminus \{0\}$. Therefore, for any $\delta > 0$ small enough, there exists $C(\delta) > 0$ such that $v_1(x) \geq C(\delta) d(x, \partial\Omega)$ for all $x \in \partial B_\delta(0) \cap \Omega$. Let $u_{\alpha_-(\gamma),-}$ be the subsolution defined in (4-4). It follows from the asymptotic (4-5) that there exists $C'(\delta) > 0$ such that $v_1 \geq C'(\delta) u_{\alpha_-(\gamma),-}$ in $\partial B_\delta(0) \cap \Omega$. Since this inequality also holds on $\partial(B_\delta(0) \cap \Omega)$ and

$$\left(-\Delta - \frac{\gamma}{|x|^2}\right)(v_1 - C'(\delta) u_{\alpha_-(\gamma),-}) \geq 0 \quad \text{in } B_\delta(0) \cap \Omega,$$

coercivity and the maximum principle yield $v_1 \geq C'(\delta) u_{\alpha_-(\gamma),-}$ in $B_\delta(0) \cap \Omega$. It then follows from (4-5) that there exists $c > 0$ such that

$$v_1(x) \geq c \cdot d(x, \partial\Omega) |x|^{-\alpha_-(\gamma)} \quad \text{in } B_\delta(0) \cap \Omega.$$

Therefore, we have for $x \in B_\delta(0) \cap \Omega$,

$$f_+(x) \leq C d(x, \partial\Omega) |x|^{-\alpha_+(\gamma)-1} \leq \frac{C}{c} |x|^{\alpha_-(\gamma)-\alpha_+(\gamma)-1} v_1(x) \leq \frac{C}{c} |x|^{\alpha_-(\gamma)-\alpha_+(\gamma)+1} \frac{v_1(x)}{|x|^2}.$$

Therefore, (7-3) yields

$$-\Delta v_1 + \frac{\gamma + O(|x|^{\alpha_-(\gamma)-\alpha_+(\gamma)+1})}{|x|^2} v_1 = 0 \quad \text{in } B_\delta(0) \cap \Omega.$$

Since $\gamma > \frac{1}{4}(n^2 - 1)$, we have that $\alpha_-(\gamma) - \alpha_+(\gamma) + 1 > 0$. Since $v_1 \in D^{1,2}(\Omega)$, $v_1 \geq 0$ and $v_1 \not\equiv 0$, it follows from Theorem 4.1 that there exists $K_1 > 0$ such that

$$v_1(x) = K_1 \frac{d(x, \partial\Omega)}{|x|^{\alpha_-(\gamma)}} + o\left(\frac{d(x, \partial\Omega)}{|x|^{\alpha_-(\gamma)}}\right) \quad \text{as } x \to 0. \tag{7-4}$$

If $f_+ \equiv 0$, then $v_1 \equiv 0$ and (7-4) holds with $K_1 = 0$. Arguing similarly for f_-, and using that $v = v_1 - v_2$, we then get that there exists $K \in \mathbb{R}$ such that

$$v(x) = -K \frac{d(x, \partial\Omega)}{|x|^{\alpha_-(\gamma)}} + o\left(\frac{d(x, \partial\Omega)}{|x|^{\alpha_-(\gamma)}}\right) \quad \text{as } x \to 0. \tag{7-5}$$

Set

$$H_0(x) := \eta(x) u_{\alpha_+(\gamma)}(x) - v(x) \quad \text{for all } x \in \overline{\Omega} \setminus \{0\}. \tag{7-6}$$

It follows from the definition of v and the regularity outside 0 that

$$-\Delta H_0 - \frac{\gamma}{|x|^2} H_0 = 0 \quad \text{in } \Omega, \qquad H_0(x) = 0 \quad \text{in } \partial\Omega \setminus \{0\}.$$

Moreover, the asymptotics (4-5) and (7-5) yield $H_0(x) > 0$ on $\Omega \cap B_{\delta'}(0)$ for some $\delta' > 0$ small enough. It follows from the comparison principle that $H_0 > 0$ in Ω.

We now perform an expansion of H_0. First note that from the definition (4-14) of $u_{\alpha_+(\gamma)}$, the asymptotic (7-5) of v and the fact that $\alpha_+(\gamma) - \alpha_-(\gamma) < 1$, we have

$$H_0(x) = \frac{d(x,\partial\Omega)}{|x|^{\alpha_+(\gamma)}}(1+O(|x|)) + K\frac{d(x,\partial\Omega)}{|x|^{\alpha_-(\gamma)}} + o\left(\frac{d(x,\partial\Omega)}{|x|^{\alpha_-(\gamma)}}\right) = \frac{d(x,\partial\Omega)}{|x|^{\alpha_+(\gamma)}} + K\frac{d(x,\partial\Omega)}{|x|^{\alpha_-(\gamma)}} + o\left(\frac{d(x,\partial\Omega)}{|x|^{\alpha_-(\gamma)}}\right)$$

as $x \to 0$. In particular, since in addition $H_0 > 0$ in Ω, there exists $c > 1$ such that

$$\frac{1}{c}\frac{d(x,\partial\Omega)}{|x|^{\alpha_+(\gamma)}} \le H_0(x) \le c\frac{d(x,\partial\Omega)}{|x|^{\alpha_+(\gamma)}} \quad \text{for all } x \in \Omega. \tag{7-7}$$

Finally, we establish the uniqueness. For that, we let $H \in C^2(\bar{\Omega} \setminus \{0\})$ be as in (7-1) and set

$$\lambda_0 := \max\{\lambda \ge 0 : H \ge \lambda H_0\}.$$

The number λ_0 is clearly defined, and so we set $\widetilde{H} := H - \lambda_0 H_0 \ge 0$. Assume that $\widetilde{H} \not\equiv 0$. Since $-\Delta\widetilde{H} - \gamma|x|^{-2}\widetilde{H} = 0$, it follows from Theorem 6.1 that there exists $\alpha \in \{\alpha_+(\gamma), \alpha_-(\gamma)\}$ and $K > 0$ such that

$$H(x) \sim_{x\to 0} K\frac{d(x,\partial\Omega)}{|x|^{\alpha}}. \tag{7-8}$$

If $\alpha = \alpha_-(\gamma)$, then $\widetilde{H} \in D^{1,2}(\Omega)$ is a variational solution to $-\Delta\widetilde{H} - \widetilde{H}\gamma/|x|^2 = 0$ in Ω. The coercivity then yields that $\widetilde{H} \equiv 0$, contradicting the initial hypothesis.

Therefore $\alpha = \alpha_+(\gamma)$. Since $\widetilde{H} > 0$ vanishes on $\partial\Omega \setminus \{0\}$, we have that for any $\delta > 0$, there exists $c(\delta) > 0$ such that

$$\widetilde{H}(x) \ge c(\delta)\, d(x,\partial\Omega) \quad \text{for } x \in \Omega \setminus B_\delta(0). \tag{7-9}$$

Therefore, (7-8), (7-9) and (7-7) yield the existence of $c > 0$ such that $\widetilde{H} \ge cH_0$, and then $H \ge (\lambda_0 + c)H_0$, contradicting the definition of λ_0. It follows that $\widetilde{H} \equiv 0$, which means that $H = \lambda_0 H_0$ for some $\lambda_0 > 0$. This proves uniqueness and completes the proof of Theorem 7.1. □

Now we establish the monotonicity of the mass with respect to set inclusion.

Proposition 7.2. *The mass m_γ is a strictly increasing set-function in the following sense*: *Assume Ω_1, Ω_2 are two smooth bounded domains such that $0 \in \partial\Omega_1 \cap \partial\Omega_2$, and $\frac{1}{4}(n^2-1) < \gamma < \min\{\gamma_H(\Omega_1), \gamma_H(\Omega_2)\}$. Then*

$$\Omega_1 \subsetneq \Omega_2 \quad \Longrightarrow \quad m_\gamma(\Omega_1) < m_\gamma(\Omega_2). \tag{7-10}$$

Moreover, if $\Omega \subsetneq \mathbb{R}^n_+$ and $\frac{1}{4}(n^2-1) < \gamma < \frac{1}{4}n^2$, then $m_\gamma(\Omega) < 0$.

Proof. It follows from the definition of the mass that for $i = 1, 2$, there exists $H_i \in C^2(\bar{\Omega}_i \setminus \{0\})$ such that

$$\begin{cases} -\Delta H_i - \dfrac{\gamma}{|x|^2}H_i = 0 & \text{in } \Omega_i, \\ H_i > 0 & \text{in } \Omega_i, \\ H_i = 0 & \text{on } \partial\Omega_i, \end{cases} \tag{7-11}$$

with

$$H_i(x) = \frac{d(x, \partial\Omega_i)}{|x|^{\alpha_+(\gamma)}} + m_\gamma(\Omega_i)\frac{d(x, \partial\Omega_i)}{|x|^{\alpha_-(\gamma)}} + o\left(\frac{d(x, \partial\Omega_i)}{|x|^{\alpha_-(\gamma)}}\right) \tag{7-12}$$

as $x \to 0,\ x \in \Omega_i$. Set $h := H_2 - H_1$ on Ω_1. Since $\Omega_1 \subsetneq \Omega_2$, we have that

$$\begin{cases} -\Delta h - \dfrac{\gamma}{|x|^2} h = 0 & \text{in } \Omega_1, \\ h \geq 0,\ h \not\equiv 0 & \text{on } \partial\Omega_1. \end{cases} \tag{7-13}$$

First, we claim that $h \in H^{1,2}(\Omega_1)$. Indeed, it follows from the construction of the singular function in (7-6) that there exists $w \in H^{1,2}(\Omega_1)$ such that

$$h(x) = \frac{d(x, \partial\Omega_2) - d(x, \partial\Omega_1)}{|x|^{\alpha_+(\gamma)}} + w(x) \quad \text{for all } x \in \Omega_1. \tag{7-14}$$

Since $\Omega_1 \subset \Omega_2$ and 0 is on the boundary of both domains, the tangent spaces at 0 of Ω_1 and Ω_2 are equal, and one gets that $d(x, \partial\Omega_1) - d(x, \partial\Omega_2) = O(|x|^2)$ as $x \to 0$. Since $\alpha_+(\gamma) - \alpha_-(\gamma) < 1$, we then get that

$$\tilde{h}(x) := \frac{d(x, \partial\Omega_2) - d(x, \partial\Omega_1)}{|x|^{\alpha_+(\gamma)}} = O(|x|^{1-\alpha_-(\gamma)}) \quad \text{as } x \to 0.$$

Similarly, $|\nabla\tilde{h}(x)| = O(|x|^{-\alpha_-(\gamma)})$ as $x \to 0$. Therefore, we deduce that $\tilde{h} \in H^{1,2}(\Omega_1)$. It then follows from (7-14) that $h \in H^{1,2}(\Omega_1)$.

To prove the monotonicity, note first that since $\gamma < \gamma_H(\Omega_1)$ and $h \in H^{1,2}(\Omega_1)$, it follows from (7-13) and the comparison principle that $h \geq 0$ in Ω_1 (indeed, this is obtained by multiplying (7-13) by $h_- \in D_1^2(\Omega)$ and integrating; therefore, coercivity yields $h_- \equiv 0$). Since $h \not\equiv 0$, it follows from Hopf's maximum principle that for any $\delta > 0$ small, there exists $C(\delta) > 0$ such that $h(x) \geq C(\delta)\, d(x, \partial\Omega_1)$ for all $x \in \partial B_\delta(0) \cap \Omega_1$. We define the subsolution $u_{\alpha_-(\gamma),-}$ as in Proposition 4.3. It then follows from the inequality above and the asymptotics in (4-5) that there exists $\epsilon_0 > 0$ such that $h(x) \geq 2\epsilon_0 u_{\alpha_-(\gamma),-}(x)$ for all $x \in \partial B_\delta(0) \cap \Omega_1$. This inequality also holds on $B_\delta(0) \cap \partial\Omega_1$ since $u_{\alpha_-(\gamma),-}$ vanishes on $\partial\Omega_1$. It then follows from the maximum principle that $h(x) \geq 2\epsilon_0 u_{\alpha_-(\gamma),-}(x)$ for all $x \in B_\delta(0) \cap \Omega_1$. With the definition of h and the asymptotic (4-5), we then have that for $\delta' > 0$ small enough

$$H_2(x) - H_1(x) \geq \epsilon_0 \frac{d(x, \partial\Omega_1)}{|x|^{\alpha_-(\gamma)}} \quad \text{for all } x \in B_{\delta'}(0) \cap \Omega_1. \tag{7-15}$$

We let $\vec{\nu}$ be the inner unit normal vector of $\partial\Omega_1$ at 0. This is also the inner unit normal vector of $\partial\Omega_2$ at 0. Therefore, for any $t > 0$ small enough, we have that $d(t\vec{\nu}, \partial\Omega_i) = t$ for $i = 1, 2$. It then follows from the expressions (7-12) and (7-15) that

$$(m_\gamma(\Omega_2) - m_\gamma(\Omega_1))\frac{t}{t^{\alpha_-(\gamma)}} + o\left(\frac{t}{t^{\alpha_-(\gamma)}}\right) \geq \epsilon_0 \frac{t}{t^{\alpha_-(\gamma)}} \quad \text{as } t \downarrow 0.$$

We then get that $m_\gamma(\Omega_2) - m_\gamma(\Omega_1) \geq \epsilon_0$, and therefore $m_\gamma(\Omega_2) > m_\gamma(\Omega_1)$. This proves (7-10) and ends the first part of Proposition 7.2.

The proof of the second part is similar. Indeed, we take $\Omega_2 := \mathbb{R}^n_+$ and we define $H_2(x) := x_1/|x|^{\alpha_+(\gamma)}$. Arguing as above, we get that $0 > m_\gamma(\Omega)$, which completes the proof of Proposition 7.2. □

Note that we have used above that the mass $m_\gamma(\mathbb{R}^n_+)$ is 0 even though we had only defined the mass for bounded sets. In the rest of the section, we shall extend the notion of mass to certain unbounded sets that include $\mathbb{R}^n_+$. For that, we shall use the Kelvin transformation, defined as follows: for any $x_0 \in \mathbb{R}^n$, let

$$i_{x_0}(x) := x_0 + |x_0|^2 \frac{x - x_0}{|x - x_0|^2} \quad \text{for all } x \in \mathbb{R}^n \setminus \{x_0\}. \tag{7-16}$$

The inversion i_{x_0} is clearly the identity map on $\partial B_{|x_0|}(x_0)$ (the ball of center x_0 and of radius $|x_0|$), and in particular $i_{x_0}(0) = 0$.

Definition 7.3. We say that a domain $\Omega \subset \mathbb{R}^n$ $(0 \in \partial\Omega)$ is *conformally bounded* if there exists $x_0 \notin \overline{\Omega}$ such that $i_{x_0}(\Omega)$ is a smooth bounded domain of $\mathbb{R}^n$ having both 0 and x_0 on its boundary $\partial(i_{x_0}(\Omega))$.

One can easily check that $\mathbb{R}^n_+$ is a smooth domain at infinity. For instance, take $x_0 := (-1, 0, \ldots, 0)$. The following proposition shows that the notion of mass extends to unbounded domains that are conformally bounded.

Proposition 7.4. *Let Ω be a conformally bounded domain in $\mathbb{R}^n$ such that $0 \in \partial\Omega$. Assume that $\gamma_H(\Omega) > \frac{1}{4}(n^2-1)$ and that $\gamma \in \big(\frac{1}{4}(n^2-1), \gamma_H(\Omega)\big)$. Then, up to a multiplicative constant, there exists a unique function $H \in C^2(\overline{\Omega} \setminus \{0\})$ such that*

$$\begin{cases} -\Delta H - \dfrac{\gamma}{|x|^2} H = 0 & \text{in } \Omega, \\ H > 0 & \text{in } \Omega, \\ H = 0 & \text{on } \partial\Omega \setminus \{0\}, \\ H(x) \le C|x|^{1-\alpha_+(\gamma)} & \text{for } x \in \Omega. \end{cases} \tag{7-17}$$

Moreover, there exists $c_1 > 0$ and $c_2 \in \mathbb{R}$ such that

$$H(x) = c_1 \frac{d(x, \partial\Omega)}{|x|^{\alpha_+(\gamma)}} + c_2 \frac{d(x, \partial\Omega)}{|x|^{\alpha_-(\gamma)}} + o\left(\frac{d(x, \partial\Omega)}{|x|^{\alpha_-(\gamma)}}\right) \quad \text{as } x \to 0.$$

We define the mass $b_\gamma(\Omega) := c_2/c_1$, which is independent of the choice of H in (7-17).

Proof. For convenience, up to a rotation and a dilation, we can assume that $x_0 := (-1, 0, \ldots, 0) \in \mathbb{R}^n$ so that the inversion becomes

$$i(x) := x_0 + \frac{x - x_0}{|x - x_0|^2} \quad \text{for all } x \in \mathbb{R}^n \setminus \{x_0\}.$$

For any $u \in C^2(U)$, with $U \subset \mathbb{R}^n$, we define its Kelvin transform $\hat{u} : \widehat{U} \to \mathbb{R}$ by

$$\hat{u}(x) := |x - x_0|^{2-n} u(i(x)) \quad \text{for all } x \in \widehat{U} := i^{-1}(U \setminus \{x_0\}).$$

This transform leaves the Laplacian invariant in the following sense:

$$-\Delta\hat{u}(x) = |x - x_0|^{-(n+2)}(-\Delta u)(i(x)) \quad \text{for all } x \in \widehat{U}. \tag{7-18}$$

Define $\widetilde{\Omega} := i(\Omega)$ and suppose $u \in C^2(\overline{\Omega} \setminus \{0\})$ is such that

$$\begin{cases} -\Delta u - \dfrac{\gamma}{|x|^2} u = 0 & \text{in } \Omega, \\ u > 0 & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases}$$

The Kelvin transform $\tilde{u}$ of u then satisfies

$$-\Delta \tilde{u} - V\tilde{u} = 0 \quad \text{in } \widetilde{\Omega},$$

where

$$V(x) := \frac{\gamma}{|x|^2 |x - x_0|^2} \quad \text{for } x \in \mathbb{R}^n \setminus \{0, x_0\}. \tag{7-19}$$

It is easy to check that

$$V(x) = \frac{\gamma + O(|x|)}{|x|^2} \quad \text{as } x \to 0 \qquad \text{and} \qquad V(x) = \frac{\gamma + O(|x - x_0|)}{|x - x_0|^2} \quad \text{as } x \to x_0.$$

In other words, the Kelvin transform allows us to reduce the study of the Hardy-singular boundary mass of a conformally bounded domain Ω into defining a notion of mass for the Schrödinger operator $-\Delta + V$ on $\widetilde{\Omega}$.

Note that the coercivity of $-\Delta - \gamma|x|^{-2}$ on Ω (since $\gamma < \gamma_H(\Omega)$) yields the coercivity of $-\Delta - V$ on $\widetilde{\Omega}$; that is, there exists $c_0 > 0$ such that

$$\int_{\widetilde{\Omega}} (|\nabla u|^2 - V(x)u^2)\, dx \geq c_0 \int_{\widetilde{\Omega}} |\nabla u|^2\, dx \quad \text{for all } u \in D^{1,2}(\widetilde{\Omega}).$$

Arguing as is Section 4, we get for $\delta > 0$ small enough, a function u_{α_+} satisfying

$$\begin{cases} (-\Delta - V)u_{\alpha_+} = O(d(x, \partial\widetilde{\Omega})|x|^{-\alpha_+(\gamma)-1}) & \text{in } \widetilde{\Omega} \cap \widetilde{B}_\delta, \\ u_{\alpha_+} > 0 & \text{in } \widetilde{\Omega} \cap \widetilde{B}_\delta, \\ u_{\alpha_+} = 0 & \text{on } \partial\widetilde{\Omega} \setminus \{0\}, \end{cases}$$

and

$$u_{\alpha_+}(x) = \frac{d(x, \partial\widetilde{\Omega})}{|x|^{\alpha_+(\gamma)}}(1 + O(|x|) \quad \text{as } x \to 0.$$

The function $f_0 := -\Delta u_{\alpha_+} - V u_{\alpha_+}$ then satisfies for all $x \in \widetilde{\Omega} \cap \widetilde{B}_\delta$,

$$|f_0(x)| \leq C d(x, \partial\widetilde{\Omega})|x|^{-\alpha_+(\gamma)-1} \leq C|x|^{-\alpha_+(\gamma)},$$

where C is a positive constant. Since $\gamma > \frac{1}{4}(n^2 - 1)$, it follows that $f_0 \in L^{2n/(n+2)}(\widetilde{\Omega})$. Let now $v_0 \in D^{1,2}(\widetilde{\Omega})$ be such that

$$-\Delta v_0 - V v_0 = f_0 \quad \text{weakly in } D^{1,2}(\widetilde{\Omega}). \tag{7-20}$$

The existence follows from the coercivity of $-\Delta - V$ on $\widetilde{\Omega}$, and the proof of Theorem 7.1 yields that $|v_0(x)|$ is bounded by $|x|^{1-\alpha_-(\gamma)}$ around 0. Note that around x_0, we have $-\Delta v_0 - V v_0 = 0$ and the regularity theorem, Theorem 5.1, yields a control by $|x - x_0|^{1-\alpha_-(\gamma)}$, which means that there exists $C > 0$ such that

$$|v_0(x)| \leq C d(x, \partial\widetilde{\Omega})\big(|x|^{-\alpha_-(\gamma)} + |x - x_0|^{-\alpha_-(\gamma)}\big) \quad \text{for all } x \in \widetilde{\Omega}.$$

The construction of the mass, Theorem 7.1, and the regularity theorem, Theorem 5.1, then yield that there exists $K_0 \in \mathbb{R}$ such that

$$v_0(x) = K_0 \frac{d(x, \partial\widetilde{\Omega})}{|x|^{\alpha_-(\gamma)}} + o\left(\frac{d(x, \partial\widetilde{\Omega})}{|x|^{\alpha_-(\gamma)}}\right). \tag{7-21}$$

Define now $\widetilde{H}_0(x) := u_{\alpha_+(\gamma)}(x) - v_0(x)$ for all $x \in \overline{\widetilde{\Omega}} \setminus \{0, x_0\}$, and consider its Kelvin transform

$$H_0(x) := |x - x_0|^{2-n} \widetilde{H}_0(i(x)) = |x - x_0|^{2-n}(u_{\alpha_+(\gamma)} - v_0)(i(x)), \quad x \in \Omega. \tag{7-22}$$

It follows from (7-18) and the definitions of $u_{\alpha_+(\gamma)}$ and v_0 that H_0 satisfies the properties

$$\begin{cases} -\Delta H_0 - \dfrac{\gamma}{|x|^2} H_0 = 0 & \text{in } \Omega, \\ H_0 > 0 & \text{in } \Omega, \\ H_0 = 0 & \text{in } \partial\Omega \setminus \{0\}. \end{cases} \tag{7-23}$$

Concerning the pointwise behavior, we have that

$$H_0(x) = \frac{d(x, \partial\Omega)}{|x|^{\alpha_+}} - K_0 \frac{d(x, \partial\Omega)}{|x|^{\alpha_-}} + o\left(\frac{d(x, \partial\Omega)}{|x|^{\alpha_-}}\right) \quad \text{as } x \to 0,\ x \in \Omega, \tag{7-24}$$

and

$$H_0(x) \le C|x|^{1-\alpha_+} \quad \text{for all } x \in \Omega,\ |x| > 1. \tag{7-25}$$

This proves the existence part in Proposition 7.4. In order to show uniqueness, we let $H \in C^2(\overline{\Omega} \setminus \{0\})$ be as in Proposition 7.4, and consider its Kelvin transform $\widetilde{H}(x) := |x - x_0|^{2-n} H(i(x))$ for all $x \in \overline{\widetilde{\Omega}} \setminus \{0, x_0\}$. The transformation law (7-18) yields

$$\begin{cases} -\Delta \widetilde{H} - V\widetilde{H} = 0 & \text{in } \widetilde{\Omega}, \\ \widetilde{H} > 0 & \text{in } \widetilde{\Omega}, \\ \widetilde{H} = 0 & \text{in } \partial\widetilde{\Omega} \setminus \{0, x_0\}. \end{cases} \tag{7-26}$$

Moreover, we have that $\widetilde{H}(x) \le C|x|^{1-\alpha_+(\gamma)} + C|x - x_0|^{1-\alpha_-(\gamma)}$ for all $x \in \widetilde{\Omega}$. It then follows from Theorem 6.1 that there exist $C_1, C_2 > 0$ such that

$$\widetilde{H}(x) \sim_{x \to 0} C_1 \frac{d(x, \partial\widetilde{\Omega})}{|x|^{\alpha}} \quad \text{and} \quad \widetilde{H}(x) \sim_{x \to x_0} C_2 \frac{d(x, \partial\widetilde{\Omega})}{|x - x_0|^{\alpha_-(\gamma)}}, \tag{7-27}$$

where $\alpha \in \{\alpha_-(\gamma), \alpha_+(\gamma)\}$. We claim that $\alpha = \alpha_+(\gamma)$. Indeed, otherwise, we would have $\widetilde{H} \in D^{1,2}(\widetilde{\Omega})$ (see Theorem 6.1) and then (7-26) and coercivity would yield $\widetilde{H} \equiv 0$, which is a contradiction. Therefore $\alpha = \alpha_+(\gamma)$. By the same reasoning, the estimates (7-27) hold for $\widetilde{H}_0$ (with different constants C_1, C_2). Arguing as in the proof of Theorem 7.1, we get that there exists $\lambda > 0$ such that $\widetilde{H} = \lambda \widetilde{H}_0$, and therefore $H = \lambda H_0$. This proves uniqueness and completes the proof of Proposition 7.4. □

Note that as a consequence of (7-24), the mass $m_\gamma(\Omega)$ is well-defined and is equal to $-K_0$.

8. Test functions and the existence of extremals

Let Ω be a domain of $\mathbb{R}^n$ such that $0 \in \partial\Omega$. For $\gamma \in \mathbb{R}$ and $s \in [0, 2)$, recall that

$$\mu_{\gamma,s}(\Omega) := \inf_{u \in D^{1,2}(\Omega)\setminus\{0\}} J^{\Omega}_{\gamma,s}(u), \tag{8-1}$$

where

$$J^{\Omega}_{\gamma,s}(u) := \frac{\int_\Omega (|\nabla u|^2 - u^2\gamma/|x|^2)\,dx}{\left(\int_\Omega |u|^{2^\star}/|x|^s\,dx\right)^{2/2^\star}}.$$

Note that critical points $u \in D^{1,2}(\Omega)$ of $J^{\Omega}_{\gamma,s}$ are weak solutions to the PDE

$$-\Delta u - \frac{\gamma}{|x|^2} = \lambda \frac{|u|^{2^\star-2}u}{|x|^s} \quad \text{for some } \lambda \in \mathbb{R}, \tag{8-2}$$

which can be rescaled to be equal to 1 if $\lambda > 0$ and to be -1 if $\lambda < 0$. In this section, we investigate the existence of minimizers for $J^{\Omega}_{\gamma,s}$. We start with the following easy case, where we do not have extremals.

Proposition 8.1. *Let $\Omega \subset \mathbb{R}^n$ be a smooth domain such that $0 \in \partial\Omega$ (no boundedness is assumed). When $s = 0$ and $\gamma \le 0$, we have that $\mu_{\gamma,0}(\Omega) = 1/K(n,2)^2$ (where $1/K(n,2)^2 = \mu_{0,0}(\mathbb{R}^n)$ is the best constant in the Sobolev inequality* (1-19)) *and there is no extremal.*

Proof. Note that $2^\star(s) = 2^\star(0) = 2^\star$. Since $\gamma \le 0$, we have for any $u \in C^\infty_c(\Omega)\setminus\{0\}$,

$$\frac{\int_\Omega (|\nabla u|^2 - u^2\gamma/|x|^2)\,dx}{\left(\int_\Omega |u|^{2^\star}\,dx\right)^{2/2^\star}} \ge \frac{\int_\Omega |\nabla u|^2\,dx}{\left(\int_\Omega |u|^{2^\star}\,dx\right)^{2/2^\star}} \ge \frac{1}{K(n,2)^2}, \tag{8-3}$$

and therefore $\mu_{\gamma,0}(\Omega) \ge 1/K(n,2)^2$. Fix now $x_0 \in \Omega$ and let $\eta \in C^\infty_c(\Omega)$ be such that $\eta(x) = 1$ around x_0. Set

$$u_\epsilon(x) := \eta(x)\left(\frac{\epsilon}{\epsilon^2 + |x - x_0|^2}\right)^{(n-2)/2}$$

for all $x \in \Omega$ and $\epsilon > 0$. Since $x_0 \neq 0$, it is classical (see, for example, [Aubin 1976]) that $\lim_{\epsilon\to 0} J^{\Omega}_{0,0}(u_\epsilon) = 1/K(n,2)^2$. It follows that $\mu_{\gamma,0}(\Omega) \le 1/K(n,2)^2$. This proves that $\mu_{\gamma,0}(\Omega) = 1/K(n,2)^2$.

Assume now that there exists an extremal u_0 for $\mu_{\gamma,0}(\Omega)$ in $D^{1,2}(\Omega)\setminus\{0\}$. It then follows from (8-3) that $u_0 \in D^{1,2}(\Omega) \subset D^{1,2}(\mathbb{R}^n)$ is an extremal for the classical Sobolev inequality on $\mathbb{R}^n$. But these extremals are known (see [Aubin 1976]) and their support is the whole of $\mathbb{R}^n$, which is a contradiction since u_0 has bounded support in Ω. It follows that there is no extremal for $\mu_{\gamma,0}(\Omega)$. □

The remainder of the section is devoted to the proof of the following.

Theorem 8.2. *Let Ω be a smooth bounded domain in $\mathbb{R}^n$ ($n \ge 3$) such that $0 \in \partial\Omega$ and let $0 \le s < 2$ and $\gamma < \frac{1}{4}n^2$. Assume that either $s > 0$, or that $\{s = 0,\ n \ge 4$ and $\gamma > 0\}$. There are then extremals for $\mu_{\gamma,s}(\Omega)$ under one of the following two conditions*:

(1) $\gamma \le \frac{1}{4}(n^2-1)$ *and the mean curvature of $\partial\Omega$ at 0 is negative.*

(2) $\gamma > \frac{1}{4}(n^2-1)$ *and the mass $m_\gamma(\Omega)$ of Ω is positive.*

Moreover, if $\gamma < \gamma_H(\Omega)$ (*resp.*, $\gamma \geq \gamma_H(\Omega)$), *then such extremals are positive solutions for* (8-2) *with* $\lambda > 0$ (*resp.*, $\lambda \leq 0$).

The remaining case $n = 3$, $s = 0$ and $\gamma > 0$ will be dealt with in Section 10.

According to Theorem 3.6, in order to establish existence of extremals, it suffices to show that $\mu_{\gamma,s}(\Omega) < \mu_{\gamma,s}(\mathbb{R}^n_+)$. The rest of the section consists in showing that the above-mentioned geometric conditions lead to such a gap. The existence of extremals on $\mathbb{R}^n_+$ as described in Proposition 1.3 is essential here.

In the sequel, $h_\Omega(0)$ will denote the mean curvature of $\partial\Omega$ at 0. The orientation is chosen such that the mean curvature of the canonical sphere (as the boundary of the ball) is positive. Since $\{s > 0\}$, or $\{s = 0$, $n \geq 4$ and $\gamma > 0\}$, it follows from Proposition 1.3 that there are extremals for $\mu_{\gamma,s}(\mathbb{R}^n_+)$. The following proposition combined with Theorem 3.6 clearly yield the claims in Theorem 8.2.

Proposition 8.3. *We fix* $\gamma < \frac{1}{4}n^2$. *Assume that there are extremals for* $\mu_{\gamma,s}(\mathbb{R}^n_+)$. *There exist then two families* $(u^1_\epsilon)_{\epsilon>0}$ *and* $(u^2_\epsilon)_{\epsilon>0}$ *in* $D^{1,2}(\Omega)$, *and two positive constants* $c^1_{\gamma,s}$ *and* $c^2_{\gamma,s}$ *such that*:

(1) *For* $\gamma < \frac{1}{4}(n^2-1)$, *we have that*

$$J(u^1_\epsilon) = \mu_{\gamma,s}(\mathbb{R}^n_+)\big(1 + c^1_{\gamma,s} \cdot h_\Omega(0) \cdot \epsilon + o(\epsilon)\big) \quad \textit{when } \epsilon \to 0. \tag{8-4}$$

(2) *For* $\gamma = \frac{1}{4}(n^2-1)$, *we have that*

$$J(u^1_\epsilon) = \mu_{\gamma,s}(\mathbb{R}^n_+)\Big(1 + c^1_{\gamma,s} \cdot h_\Omega(0) \cdot \epsilon \ln\frac{1}{\epsilon} + o\Big(\epsilon \ln\frac{1}{\epsilon}\Big)\Big) \quad \textit{when } \epsilon \to 0. \tag{8-5}$$

(3) *For* $\gamma > \frac{1}{4}(n^2-1)$, *we have as* $\epsilon \to 0$, *that*

$$J(u^2_\epsilon) = \mu_{\gamma,s}(\mathbb{R}^n_+)\big(1 - c^2_{\gamma,s} \cdot m_\gamma(\Omega) \cdot \epsilon^{\alpha_+(\gamma)-\alpha_-(\gamma)} + o(\epsilon^{\alpha_+(\gamma)-\alpha_-(\gamma)})\big). \tag{8-6}$$

Remark. When $\gamma < \frac{1}{4}(n^2-1)$, this result is due to Chern and Lin [2010]. Actually, they stated the result for $\gamma < \frac{1}{4}(n-2)^2$, but their proof works for $\gamma < \frac{1}{4}(n^2-1)$. However, when $\gamma \geq \frac{1}{4}(n^2-1)$, we need the exact asymptotic profile of U that was described by Corollary 5.3.

Proof. By assumption, there exists $U \in D^{1,2}(\mathbb{R}^n_+) \setminus \{0\}$, $U \geq 0$, that is a minimizer for $\mu_{\gamma,s}(\mathbb{R}^n_+)$. In other words,

$$J^{\mathbb{R}^n_+}_{\gamma,s}(U) = \frac{\int_{\mathbb{R}^n_+}(|\nabla U|^2 - U^2\gamma/|x|^2)\,dx}{\big(\int_{\mathbb{R}^n_+}|U|^{2^\star(s)}/|x|^s\,dx\big)^{2/2^\star(s)}} = \mu_{\gamma,s}(\mathbb{R}^n_+).$$

Therefore, there exists $\lambda > 0$ such that

$$\begin{cases} -\Delta U - \dfrac{\gamma}{|x|^2}U = \lambda\dfrac{U^{2^\star(s)-1}}{|x|^s} & \text{in } \mathbb{R}^n_+, \\ U > 0 & \text{in } \mathbb{R}^n_+, \\ U = 0 & \text{in } \partial\mathbb{R}^n_+, \end{cases} \tag{8-7}$$

and there exist $K_1, K_2 > 0$ such that

$$U(x) \sim_{x\to 0} K_1\frac{x_1}{|x|^{\alpha_-}} \quad \text{and} \quad U(x) \sim_{|x|\to+\infty} K_2\frac{x_1}{|x|^{\alpha_+}}, \tag{8-8}$$

where here and in the sequel, we write for convenience

$$\alpha_+ := \alpha_+(\gamma) \quad \text{and} \quad \alpha_- := \alpha_-(\gamma).$$

In particular, it follows from Lemma 5.2 (after reducing all limits to happen at 0 via the Kelvin transform) that there exists $C > 0$ such that

$$U(x) \le C x_1 |x|^{-\alpha_+} \quad \text{and} \quad |\nabla U(x)| \le C|x|^{-\alpha_+} \quad \text{for all } x \in \mathbb{R}^n_+. \tag{8-9}$$

We shall now construct a suitable test function for each range of γ. First note that

$$\gamma < \tfrac{1}{4}(n^2-1) \iff \alpha_+ - \alpha_- > 1,$$
$$\gamma = \tfrac{1}{4}(n^2-1) \iff \alpha_+ - \alpha_- = 1.$$

Concerning terminology, here and in the sequel, we define as in (4-6)

$$\widetilde{B}_r := (-r, r) \times B_r^{(n-1)}(0) \subset \mathbb{R} \times \mathbb{R}^{n-1}$$

for all $r > 0$ and

$$V_+ := V \cap \mathbb{R}^n_+$$

for all $V \subset \mathbb{R}^n$. Since Ω is smooth, up to a rotation, there exist $\delta > 0$ and $\varphi_0 : B_\delta^{(n-1)}(0) \to \mathbb{R}$ such that $\varphi_0(0) = |\nabla \varphi_0(0)| = 0$ and

$$\begin{aligned} &\varphi : \widetilde{B}_{3\delta} \to \mathbb{R}^n, \\ &(x_1, x') \mapsto (x_1 + \varphi_0(x'), x'), \end{aligned} \tag{8-10}$$

that realizes a diffeomorphism onto its image and such that

$$\varphi(\widetilde{B}_{3\delta} \cap \mathbb{R}^n_+) = \varphi(\widetilde{B}_{3\delta}) \cap \Omega \quad \text{and} \quad \varphi(\widetilde{B}_{3\delta} \cap \partial\mathbb{R}^n_+) = \varphi(\widetilde{B}_{3\delta}) \cap \partial\Omega.$$

Let $\eta \in C_c^\infty(\mathbb{R}^n)$ be such that $\eta(x) = 1$ for all $x \in \widetilde{B}_\delta$ and $\eta(x) = 0$ for all $x \notin \widetilde{B}_{2\delta}$.

Case 1: $\gamma \le \frac{1}{4}(n^2-1)$. As in [Chern and Lin 2010], for any $\epsilon > 0$, we define

$$u_\epsilon(x) := \big(\eta \epsilon^{-(n-2)/2} U(\epsilon^{-1}x)\big) \circ \varphi^{-1}(x) \quad \text{for } x \in \varphi(\widetilde{B}_{2\delta}) \cap \Omega \text{ and 0 elsewhere.}$$

This case is devoted to giving a Taylor expansion of $J^\Omega_{\gamma,s}(u_\epsilon)$ as $\epsilon \to 0$. In the sequel, we adopt the following notation: given $(a_\epsilon)_{\epsilon>0} \in \mathbb{R}$, let $\Theta_\gamma(a_\epsilon)$ denote a quantity such that, as $\epsilon \to 0$,

$$\Theta_\gamma(a_\epsilon) := \begin{cases} o(a_\epsilon) & \text{if } \gamma < \frac{1}{4}(n^2-1), \\ O(a_\epsilon) & \text{if } \gamma = \frac{1}{4}(n^2-1). \end{cases}$$

A. Estimate of $\int_\Omega |\nabla u_\epsilon|^2\, dx$. It follows from (8-9) that

$$|\nabla u_\epsilon(x)| \le C\epsilon^{\alpha_+ - n/2}|x|^{-\alpha_+} \quad \text{for all } x \in \Omega \text{ and } \epsilon > 0. \tag{8-11}$$

Therefore, $\int_{\varphi((\widetilde{B}_{3\delta} \setminus \widetilde{B}_\delta) \cap \mathbb{R}^n_+)} |\nabla u_\epsilon|^2\, dx = \Theta_\gamma(\epsilon)$ as $\epsilon \to 0$. It follows that

$$\int_\Omega |\nabla u_\epsilon|^2\, dx = \int_{\widetilde{B}_{\delta,+}} |\nabla(u_\epsilon \circ \varphi)|^2_{\varphi^\star \mathrm{Eucl}} |\mathrm{Jac}\,\varphi|\, dx + \Theta_\gamma(\epsilon) \quad \text{as } \epsilon \to 0,$$

where $\widetilde{B}_{\delta,+} := \widetilde{B}_\delta \cap \mathbb{R}^n_+$. The definition (8-10) of φ yields $\operatorname{Jac}\varphi = 1$. Moreover, for any $\theta \in (0,1)$, we have as $x \to 0$,

$$\varphi^\star \mathrm{Eucl} := \begin{pmatrix} 1 & \partial_j\varphi_0 \\ \partial_i\varphi_0 & \delta_{ij} + \partial_i\varphi_0\,\partial_j\varphi_0 \end{pmatrix} = \mathrm{Id} + H + O(|x|^{1+\theta}),$$

where

$$H := \begin{pmatrix} 0 & \partial_j\varphi_0 \\ \partial_i\varphi_0 & 0 \end{pmatrix}.$$

It follows that

$$\int_\Omega |\nabla u_\epsilon|^2\,dx = \int_{\widetilde{B}_{\delta,+}} |\nabla(u_\epsilon\circ\varphi)|^2_{\mathrm{Eucl}}\,dx - \int_{\widetilde{B}_{\delta,+}} H^{ij}\,\partial_i(u_\epsilon\circ\varphi)\partial_j(u_\epsilon\circ\varphi)\,dx + O\left(\int_{\widetilde{B}_{\delta,+}} |x|^{1+\theta}|\nabla(u_\epsilon\circ\varphi)|^2\,dx\right) + \Theta_\gamma(\epsilon) \quad \text{as } \epsilon\to 0. \tag{8-12}$$

We have that

$$\begin{aligned}\int_{\widetilde{B}_{\delta,+}} & H^{ij}\,\partial_i(u_\epsilon\circ\varphi)\partial_j(u_\epsilon\circ\varphi)\,dx \\ &= 2\sum_{i\ge 2}\int_{\widetilde{B}_{\delta,+}} H^{1i}\,\partial_1(u_\epsilon\circ\varphi)\partial_i(u_\epsilon\circ\varphi)\,dx = 2\sum_{i\ge 2}\int_{\widetilde{B}_{\delta,+}} \partial_i\varphi_0(x')\partial_1(u_\epsilon\circ\varphi)\partial_i(u_\epsilon\circ\varphi)\,dx \\ &= 2\sum_{i,j\ge 2}\int_{\widetilde{B}_{\delta,+}} \partial_{ij}\varphi_0(0)(x')^j\,\partial_1(u_\epsilon\circ\varphi)\partial_i(u_\epsilon\circ\varphi)\,dx + O\left(\int_{\widetilde{B}_{\delta,+}} |x|^2|\nabla(u_\epsilon\circ\varphi)|^2\,dx\right) \quad \text{as } \epsilon\to 0. \end{aligned} \tag{8-13}$$

We let II be the second fundamental form at 0 of the oriented boundary $\partial\Omega$. By definition, for any $X, Y \in T_0\partial\Omega$, we have that

$$\mathrm{II}(X,Y) := (d\vec{\nu}_0(X), Y)_{\mathrm{Eucl}},$$

where $\vec{\nu} : \partial\Omega \to \mathbb{R}^n$ is the outer unit normal vector of $\partial\Omega$. In particular, we have that $\vec{\nu}(0) = (-1, 0, \cdot, 0)$. For any $i, j \ge 2$, we have that

$$\mathrm{II}_{ij} := \mathrm{II}(\partial_i\varphi(0), \partial_j\varphi(0)) = \big(\partial_i(\vec{\nu}\circ\varphi)(0), \partial_j\varphi(0)\big) = -(\vec{\nu}(0), \partial_{ij}\varphi(0)) = \partial_{ij}\varphi_0(0).$$

Plugging (8-13) in (8-12), and using a change of variables, we get that

$$\int_\Omega |\nabla u_\epsilon|^2\,dx = \int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} |\nabla U|^2\,dx - 2\mathrm{II}_{ij}\sum_{i,j\ge 2}\int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} (x')^j\,\partial_1 U\,\partial_i U\,dx + O\left(\int_{\widetilde{B}_{\delta,+}} |x|^{1+\theta}|\nabla(u_\epsilon\circ\varphi)|^2\,dx\right) + \Theta_\gamma(\epsilon) \quad \text{as } \epsilon\to 0. \tag{8-14}$$

We now choose θ in the following way:

(i) If $\gamma < \frac{1}{4}(n^2-1)$, then take θ in $(0, \alpha_+ - \alpha_- - 1)$.

(ii) If $\gamma = \frac{1}{4}(n^2-1)$, take $\theta \in (0,1)$.

In both cases, we get by using (8-11), that

$$\int_{\widetilde{B}_{\delta,+}} |x|^{1+\theta} |\nabla(u_\epsilon \circ \varphi)|^2 \, dx = \Theta_\gamma(\epsilon) \quad \text{as } \epsilon \to 0. \tag{8-15}$$

Moreover, using (8-9), we have that

$$\int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} |\nabla U|^2 \, dx = \int_{\mathbb{R}^n_+} |\nabla U|^2 \, dx + \Theta_\gamma(\epsilon) \quad \text{as } \epsilon \to 0. \tag{8-16}$$

Plugging together (8-14)–(8-16) yields

$$\int_\Omega |\nabla u_\epsilon|^2 \, dx = \int_{\mathbb{R}^n_+} |\nabla U|^2 \, dx - 2I I_{ij} \sum_{i,j\geq 2} \int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} (x')^j \partial_1 U \partial_i U \, dx + \Theta_\gamma(\epsilon). \tag{8-17}$$

B. Estimate for $\int_\Omega |u_\epsilon|^{2^\star(s)}/|x|^s \, dx$. Fix $\sigma \in [0, 2]$. We will apply the estimates below to $\sigma = s \in [0, 2)$ or to $\sigma := 2$. The first estimate in (8-9) yields

$$|u_\epsilon(x)| \leq C\epsilon^{\alpha_+ - n/2} d(x, \partial\Omega)\, |x|^{-\alpha_+} \leq C\epsilon^{\alpha_+ - n/2} |x|^{1-\alpha_+} \tag{8-18}$$

for all $\epsilon > 0$ and all $x \in \Omega$. Since $\operatorname{Jac} \varphi = 1$, this estimate then yields

$$\begin{aligned} \int_\Omega \frac{|u_\epsilon|^{2^\star(\sigma)}}{|x|^\sigma} \, dx &= \int_{\varphi(\widetilde{B}_{\delta,+})} \frac{|u_\epsilon|^{2^\star(\sigma)}}{|x|^\sigma} \, dx + \Theta_\gamma(\epsilon) \\ &= \int_{\widetilde{B}_{\delta,+}} \frac{|u_\epsilon \circ \varphi|^{2^\star(\sigma)}}{|\varphi(x)|^\sigma} \, dx + \Theta_\gamma(\epsilon) \quad \text{as } \epsilon \to 0. \end{aligned} \tag{8-19}$$

If $\gamma < \frac{1}{4}(n^2-1)$ or if $\gamma = \frac{1}{4}(n^2-1)$ and $\sigma < 2$, we choose $\theta \in \big(0, (\alpha_+ - \alpha_-)2^\star(\sigma)/2 - 1\big) \cap (0, 1)$. If $\gamma = \frac{1}{4}(n^2-1)$ and $\sigma = 2$, we choose any $\theta \in (0, 1)$. Using the expression of $\varphi(x_1, x')$, a Taylor expansion yields

$$|\varphi(x)|^{-\sigma} = |x|^{-\sigma} \left(1 - \frac{\sigma}{2} \frac{x_1}{|x|^2} \sum_{i,j\geq 2} \partial_{ij}\varphi_0(0)(x')^i (x')^j + O(|x|^{1+\theta})\right) \quad \text{as } \epsilon \to 0. \tag{8-20}$$

The choice of θ yields

$$\int_{\widetilde{B}_{\delta,+}} \frac{|u_\epsilon \circ \varphi|^{2^\star(\sigma)}}{|\varphi(x)|^\sigma} |x|^{1+\theta} \, dx = \Theta_\gamma(\epsilon) \quad \text{as } \epsilon \to 0. \tag{8-21}$$

Putting together (8-19)–(8-21), using a change of variable and (8-9), we get as $\epsilon \to 0$ that

$$\int_\Omega \frac{|u_\epsilon|^{2^\star(\sigma)}}{|x|^\sigma} \, dx = \int_{\mathbb{R}^n_+} \frac{|U|^{2^\star(\sigma)}}{|x|^\sigma} \, dx - \frac{\sigma}{2} \sum_{i,j\geq 2} \epsilon \, \mathrm{II}_{ij} \int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \frac{|U|^{2^\star(\sigma)}}{|x|^\sigma} \frac{x_1}{|x|^2} (x')^i (x')^j \, dx + \Theta_\gamma(\epsilon). \tag{8-22}$$

We now compute the terms in U by using its symmetry property established in [Chern and Lin 2010]. Indeed, there exists $\widetilde{U} : (0, +\infty) \times \mathbb{R}$ such that $U(x_1, x') = \widetilde{U}(x_1, |x'|)$ for all $(x_1, x') \in \mathbb{R}^n_+$. Therefore,

for any $i, j \geq 2$, we get that

$$\int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \frac{|U|^{2^\star(\sigma)}}{|x|^\sigma} \frac{x_1}{|x|^2} (x')^i (x')^j \, dx = \frac{\delta_{ij}}{n-1} \int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \frac{|U|^{2^\star(\sigma)}}{|x|^\sigma} \frac{x_1}{|x|^2} |x'|^2 \, dx$$

and that

$$\int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} (x')^j \, \partial_1 U \, \partial_i U \, dx = \frac{\delta_{ij}}{n-1} \int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \partial_1 U (x', \nabla U) \, dx,$$

where $x = (x_1, x') \in \mathbb{R}^n_+$. Therefore, the identities (8-17) and (8-22) can be rewritten as

$$\int_\Omega |\nabla u_\epsilon|^2 \, dx = \int_{\mathbb{R}^n_+} |\nabla U|^2 \, dx - \frac{2h_\Omega(0)}{n-1} \epsilon \int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \partial_1 U (x', \nabla U) \, dx + \Theta_\gamma(\epsilon) \tag{8-23}$$

and

$$\int_\Omega \frac{|u_\epsilon|^{2^\star(\sigma)}}{|x|^\sigma} \, dx = \int_{\mathbb{R}^n_+} \frac{|U|^{2^\star(\sigma)}}{|x|^\sigma} \, dx - \frac{\sigma h_\Omega(0)}{2(n-1)} \epsilon \int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \frac{|U|^{2^\star(\sigma)}}{|x|^\sigma} \frac{x_1}{|x|^2} |x'|^2 \, dx + \Theta_\gamma(\epsilon) \tag{8-24}$$

as $\epsilon \to 0$, where $h_\Omega(0) = \sum_i \mathrm{II}_{ii}$ is the mean curvature at 0.

C. An intermediate identity. We now claim that as $\epsilon \to 0$,

$$\begin{aligned} &\int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \partial_1 U (x', \nabla U) \, dx \\ &\qquad = \int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \frac{|x'|^2 x_1}{2|x|^2} \left(\lambda \frac{s}{2^\star(s)} \frac{U^{2^\star(s)}}{|x|^s} + \gamma \frac{U^2}{|x|^2} \right) dx - \int_{\partial \mathbb{R}^n_+ \cap \widetilde{B}_{\epsilon^{-1}\delta}} \frac{|x'|^2 (\partial_1 U)^2}{4} \, dx + \Theta_\gamma(1), \end{aligned} \tag{8-25}$$

where $\lambda > 0$ is as in (8-7). This was shown by Chern and Lin [2010], and we include it for the sake of completeness. Here and in the sequel, ν_i denotes the i-th coordinate of the direct outward normal vector on the boundary of the relevant domain (for instance, on $\partial \mathbb{R}^n_+$, we have that $\nu_i = -\delta_{1i}$). We write

$$\begin{aligned} \int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \partial_1 U (x', \nabla U) \, dx &= \sum_{j \geq 2} \int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \partial_1 U (x')^j \, \partial_j U \, dx \\ &= \sum_{j \geq 2} \int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \partial_1 U \, \partial_j \left(\frac{|x'|^2}{2} \right) \partial_j U \, dx \\ &= \sum_{j \geq 2} \int_{\partial(\widetilde{B}_{\epsilon^{-1}\delta,+})} \partial_1 U \, \frac{|x'|^2}{2} \, \partial_j U \nu_j \, d\sigma - \sum_{j \geq 2} \int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \frac{|x'|^2}{2} \partial_j (\partial_1 U \, \partial_j U) \, dx \\ &= \sum_{j \geq 2} \int_{\partial \mathbb{R}^n_+ \cap \widetilde{B}_{\epsilon^{-1}\delta}} \partial_1 U \, \frac{|x'|^2}{2} \, \partial_j U \nu_j \, d\sigma + O\left(\int_{\mathbb{R}^n_+ \cap \partial \widetilde{B}_{\epsilon^{-1}\delta}} |x'|^2 \, |\nabla U|^2(x) \, d\sigma \right) \\ &\qquad - \sum_{j \geq 2} \int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \frac{|x'|^2}{2} (\partial_{1j} U \, \partial_j U + \partial_1 U \, \partial_{jj} U) \, dx. \end{aligned} \tag{8-26}$$

Since $U(0, x') = 0$ for all $x' \in \mathbb{R}^{n-1}$, using the upper-bound (8-9) and writing $\nabla' = (\partial_2, \dots, \partial_n)$, we get that

$$\begin{aligned}\int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \partial_1 U(x', \nabla U)\,dx &= -\sum_{j\geq 2}\int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \frac{|x'|^2}{2}(\partial_{1j}U\,\partial_j U + \partial_1 U\,\partial_{jj}U)\,dx + \Theta_\gamma(1)\\ &= -\int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \frac{|x'|^2}{4}\partial_1(|\nabla' U|^2)\,dx + \int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \frac{|x'|^2}{2}\partial_1 U(-\Delta U + \partial_{11}U)\,dx + \Theta_\gamma(1)\\ &= -\int_{\partial(\widetilde{B}_{\epsilon^{-1}\delta,+})} \frac{|x'|^2|\nabla' U|^2}{4}\nu_1\,dx + \int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \frac{|x'|^2}{2}\partial_1 U(-\Delta U)\,dx\\ &\quad + \int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \partial_1\left(\frac{|x'|^2(\partial_1 U)^2}{4}\right)dx + \Theta_\gamma(1). \qquad (8\text{-}27)\end{aligned}$$

Using again that U vanishes on $\partial\mathbb{R}^n_+$ and the bound (8-9), we get as $\epsilon \to 0$,

$$\begin{aligned}&\int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \partial_1 U(x', \nabla U)\,dx\\ &= \int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \frac{|x'|^2}{2}\partial_1 U(-\Delta U)\,dx + \int_{\partial\mathbb{R}^n_+\cap\widetilde{B}_{\epsilon^{-1}\delta}} \frac{|x'|^2(\partial_1 U)^2}{4}\nu_1\,dx + O\left(\int_{\partial(\widetilde{B}_{\epsilon^{-1}\delta})\cap\mathbb{R}^n_+} |x'|^2|\nabla U|^2\,dx\right) + \Theta_\gamma(1)\\ &= \int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \frac{|x'|^2}{2}\partial_1 U(-\Delta U)\,dx - \int_{\partial\mathbb{R}^n_+\cap\widetilde{B}_{\epsilon^{-1}\delta}} \frac{|x'|^2(\partial_1 U)^2}{4}\,dx + \Theta_\gamma(1). \qquad (8\text{-}28)\end{aligned}$$

Now use equation (8-7) to get that

$$\int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \frac{|x'|^2}{2}\partial_1 U(-\Delta U)\,dx = \int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \frac{|x'|^2}{2}\partial_1 U\left(\lambda\frac{U^{2^\star(s)-1}}{|x|^s} + \gamma\frac{U}{|x|^2}\right)dx. \qquad (8\text{-}29)$$

Integrating by parts, using that U vanishes on $\partial\mathbb{R}^n_+$ and the upper-bound (8-9), for $\sigma \in [0, 2]$, we get that

$$\begin{aligned}\int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} |x'|^2\,\partial_1 U\,\frac{U^{2^\star(\sigma)-1}}{|x|^\sigma}\,dx &= \int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} |x'|^2|x|^{-\sigma}\partial_1\left(\frac{U^{2^\star(\sigma)}}{2^\star(\sigma)}\right)dx\\ &= \int_{\partial(\widetilde{B}_{\epsilon^{-1}\delta,+})} |x'|^2|x|^{-\sigma}\frac{U^{2^\star(\sigma)}}{2^\star(\sigma)}\nu_1\,dx - \int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \partial_1(|x'|^2|x|^{-\sigma})\left(\frac{U^{2^\star(\sigma)}}{2^\star(\sigma)}\right)dx\\ &= O\left(\int_{\mathbb{R}^n_+\cap\partial\widetilde{B}_{\epsilon^{-1}\delta,+}} |x|^{2-\sigma}U^{2^\star(\sigma)}\,d\sigma\right) + \frac{\sigma}{2^\star(s)}\int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \frac{|x'|^2x_1}{|x|^{\sigma+2}}U^{2^\star(\sigma)}\,dx\\ &= \frac{\sigma}{2^\star(s)}\int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \frac{|x'|^2x_1}{|x|^{\sigma+2}}U^{2^\star(\sigma)}\,dx + \Theta_\gamma(1) \quad \text{as } \epsilon \to 0. \qquad (8\text{-}30)\end{aligned}$$

Putting together (8-28)–(8-30) yields (8-25).

D. Estimate for $J^\Omega_{\gamma,s}(u_\epsilon)$. Since $U \in D^{1,2}(\mathbb{R}^n)$, it follows from (8-7) that

$$\int_{\mathbb{R}^n_+}\left(|\nabla U|^2 - \frac{\gamma}{|x|^2}U^2\right)dx = \lambda\int_{\mathbb{R}^n_+}\frac{U^{2^\star(s)}}{|x|^s}\,dx. \qquad (8\text{-}31)$$

This equality, combined with (8-23) and (8-24) gives

$$J^{\Omega}_{\gamma,s}(u_\epsilon) = \frac{\int_\Omega \big(|\nabla u_\epsilon|^2 - u_\epsilon^2\gamma/|x|^2\big)\,dx}{\big(\int_\Omega |u_\epsilon|^{2^\star(s)}/|x|^s\,dx\big)^{2/2^\star(s)}}$$
$$= \frac{\int_{\mathbb{R}^n_+} \big(|\nabla U|^2 - U^2\gamma/|x|^2\big)\,dx}{\big(\int_{\mathbb{R}^n_+} |U|^{2^\star(s)}/|x|^s\,dx\big)^{2/2^\star(s)}}\bigg(1 + \epsilon\frac{h_\Omega(0)}{(n-1)\lambda\int_{\mathbb{R}^n_+}|U|^{2^\star(s)}/|x|^s\,dx}C_\epsilon + \Theta_\gamma(\epsilon)\bigg), \tag{8-32}$$

where for all $\epsilon > 0$,

$$C_\epsilon := -2\int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \partial_1 U(x', \nabla U)\,dx + \gamma\int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \frac{|x'|^2x_1}{|x|^2}\frac{U^2}{|x|^2}\,dx + \lambda\frac{s}{2^\star(s)}\int_{\widetilde{B}_{\epsilon^{-1}\delta,+}} \frac{|x'|^2x_1}{|x|^2}\frac{U^{2^\star(s)}}{|x|^s}\,dx.$$

The identity (8-25) then yields as $\epsilon \to 0$,

$$C_\epsilon = \int_{\partial\mathbb{R}^n_+\cap\widetilde{B}_{\epsilon^{-1}\delta}} \frac{|x'|^2(\partial_1 U)^2}{2}\,dx + \Theta_\gamma(1).$$

Therefore, (8-32) yields that as $\epsilon \to 0$,

$$J^{\Omega}_{\gamma,s}(u_\epsilon) = \mu_{\gamma,s}(\mathbb{R}^n_+)\bigg(1 + \epsilon\frac{h_\Omega(0)\int_{\partial\mathbb{R}^n_+\cap\widetilde{B}_{\epsilon^{-1}\delta}}|x'|^2(\partial_1 U)^2\,dx'}{2(n-1)\lambda\int_{\mathbb{R}^n_+}|U|^{2^\star(s)}/|x|^s\,dx} + \Theta_\gamma(\epsilon)\bigg). \tag{8-33}$$

We now distinguish two cases:

(i) $\gamma < \frac{1}{4}(n^2-1)$. The bound (8-9) then yields $x' \mapsto |x'|^2|\partial_1 U(x')|^2$ is in $L^1(\partial\mathbb{R}^n_+)$ and so we get from (8-33) that

$$J^{\Omega}_{\gamma,s}(u_\epsilon) = \mu_{\gamma,s}(\mathbb{R}^n_+)\big(1 + C_0\cdot h_\Omega(0)\cdot\epsilon + o(\epsilon)\big) \quad \text{as } \epsilon \to 0, \tag{8-34}$$

with

$$C_0 := \frac{\int_{\partial\mathbb{R}^n_+}|x'|^2(\partial_1 U)^2\,dx'}{2(n-1)\lambda\int_{\mathbb{R}^n_+}|U|^{2^\star(s)}/|x|^s\,dx} > 0.$$

(ii) $\gamma = \frac{1}{4}(n^2-1)$. From (8-8), Lemma 5.2 and the Kelvin transform, we have that

$$\lim_{|x'|\to+\infty} |x'|^{\alpha_+}|\partial_1 U(0,x')| = K_2 > 0.$$

Since $2\alpha_+ - 2 = n-1$, we get that

$$\int_{\partial\mathbb{R}^n_+\cap\widetilde{B}_{\epsilon^{-1}\delta}} |x'|^2(\partial_1 U)^2\,dx' = \omega_{n-1}K_2^2\ln\frac{1}{\epsilon} + o\Big(\ln\frac{1}{\epsilon}\Big)$$

as $\epsilon \to 0$. Therefore, (8-33) yields

$$J^{\Omega}_{\gamma,s}(u_\epsilon) = \mu_{\gamma,s}(\mathbb{R}^n_+)\Big(1 + C_0'h_\Omega(0)\epsilon\ln\frac{1}{\epsilon} + o\Big(\ln\frac{1}{\epsilon}\Big)\Big) \quad \text{as } \epsilon \to 0, \tag{8-35}$$

where

$$C_0' := \frac{\omega_{n-1}K_2^2}{2(n-1)\lambda\int_{\mathbb{R}^n_+}|U|^{2^\star(s)}/|x|^s\,dx} > 0.$$

Cases (i) and (ii) prove Proposition 8.3 when $\gamma \le \frac{1}{4}(n^2-1)$.

Case 2: $\gamma > \frac{1}{4}(n^2-1)$. In this case, the construction of test functions is more subtle. First, use Theorem 7.1 to obtain $H \in C^2(\bar{\Omega} \setminus \{0\})$ such that (7-1) holds and

$$H(x) = \frac{d(x,\partial\Omega)}{|x|^{\alpha_+}} + m_\gamma(\Omega)\frac{d(x,\partial\Omega)}{|x|^{\alpha_-}} + o\left(\frac{d(x,\partial\Omega)}{|x|^{\alpha_-}}\right) \quad \text{when } x \to 0. \tag{8-36}$$

As above, we fix $\eta \in C_c^\infty(\mathbb{R}^n)$ such that $\eta(x) = 1$ for all $x \in \widetilde{B}_\delta$ and $\eta(x) = 0$ for all $x \notin \widetilde{B}_{2\delta}$. We then define β such that

$$H(x) = \left(\eta\frac{x_1}{|x|^{\alpha_+}}\right) \circ \varphi^{-1}(x) + \beta(x) \quad \text{for all } x \in \Omega.$$

Here φ is as in (4-7)–(4-12). Note that $\beta \in D^{1,2}(\Omega)$ and

$$\beta(x) = m_\gamma(\Omega)\frac{d(x,\partial\Omega)}{|x|^{\alpha_-}} + o\left(\frac{d(x,\partial\Omega)}{|x|^{\alpha_-}}\right) \quad \text{as } x \to 0. \tag{8-37}$$

Indeed, since $\alpha_+ - \alpha_- < 1$, an essential point underlying all of this case is that

$$|x| = o(|x|^{\alpha_+ - \alpha_-}) \quad \text{as } x \to 0.$$

We choose U as in (8-7). By multiplying by a constant if necessary, we assume that $K_2 = 1$; that is,

$$U(x) \sim_{x\to 0} K_1\frac{x_1}{|x|^{\alpha_-}} \quad \text{and} \quad U(x) \sim_{|x|\to+\infty} \frac{x_1}{|x|^{\alpha_+}}. \tag{8-38}$$

Now define

$$u_\epsilon(x) := (\eta\epsilon^{-(n-2)/2}U(\epsilon^{-1}\,\cdot)) \circ \varphi^{-1}(x) + \epsilon^{(\alpha_+-\alpha_-)/2}\beta(x) \quad \text{for } x \in \Omega \text{ and } \epsilon > 0. \tag{8-39}$$

We start by showing that for any $k \geq 0$,

$$\lim_{\epsilon\to 0} \frac{u_\epsilon}{\epsilon^{(\alpha_+-\alpha_-)/2}} = H \quad \text{in } C^k_{\text{loc}}(\bar{\Omega} \setminus \{0\}). \tag{8-40}$$

Indeed, the convergence in $C^0_{\text{loc}}(\bar{\Omega} \setminus \{0\})$ is a consequence of the definition of u_ϵ, the choice $K_2 = 1$ and the asymptotic behavior (8-38). For convergence in C^k, we need in addition that $\nabla^i(U - x_1|x|^{-\alpha_+}) = o(|x|^{1-\alpha_+-i})$ as $x \to +\infty$ for all $i \geq 0$. This estimate follows from (8-38) and Lemma 5.2.

In the sequel, we adopt the following notation: θ_c^ϵ will denote any quantity such that there exists $\theta : \mathbb{R} \to \mathbb{R}$ such that $\lim_{c\to 0}\lim_{\epsilon\to 0}\theta_c^\epsilon = 0$.

We first claim that for any $c > 0$, we have that

$$\int_{\Omega\setminus\varphi(B_c(0)_+)} \left(|\nabla u_\epsilon|^2 - \frac{\gamma}{|x|^2}u_\epsilon^2\right) dx = \epsilon^{\alpha_+-\alpha_-}\left((\alpha_+ - 1)c^{n-2\alpha_+}\frac{\omega_{n-1}}{2n} + m_\gamma(\Omega)\frac{(n-2)\omega_{n-1}}{2n}\right) + \theta_c^\epsilon\epsilon^{\alpha_+-\alpha_-}. \tag{8-41}$$

Indeed, it follows from (8-40) that

$$\lim_{\epsilon\to 0} \frac{\int_{\Omega\setminus\varphi(B_c(0)_+)}\left(|\nabla u_\epsilon|^2 - u_\epsilon^2\gamma/|x|^2\right)dx}{\epsilon^{\alpha_+-\alpha_-}} = \int_{\Omega\setminus\varphi(B_c(0)_+)}\left(|\nabla H|^2 - \frac{\gamma}{|x|^2}H^2\right)dx. \tag{8-42}$$

Since H vanishes on $\partial\Omega\setminus\{0\}$ and satisfies $-\Delta H - H\gamma/|x|^2 = 0$, integrating by parts yields

$$\int_{\Omega\setminus\varphi(B_c(0)_+)}\left(|\nabla H|^2 - \frac{\gamma}{|x|^2}H^2\right)dx = -\int_{\varphi(\mathbb{R}^n_+\cap\partial B_c(0))} H\,\partial_\nu H\,d\sigma$$
$$= -\int_{\mathbb{R}^n_+\cap\partial B_c(0)} H\circ\varphi\,\partial_{\varphi_\star\nu}(H\circ\varphi)\,d(\varphi^\star\sigma), \tag{8-43}$$

where in the two last equalities, $\nu(x)$ is the outer normal vector of $B_c(0)$ at $x\in\partial B_c(0)$.

We now estimate $H\circ\varphi\,\partial_{\varphi_\star\nu}H\circ\varphi$. Since $\varphi_\star\nu(x) = x/|x| + O(|x|)$ as $x\to 0$, it follows from (8-36) that

$$H\circ\varphi\,\partial_{\varphi_\star\nu}(H\circ\varphi) = \frac{(\alpha_+-1)x_1^2}{|x|^{2\alpha_+ +1}} + (n-2)m_\gamma(\Omega)\frac{x_1^2}{|x|^{n+1}} + o(|x|^{1-n}) \quad \text{as } x\to 0.$$

Integrating this expression on $B_c(0)_+ = \mathbb{R}^n_+\cap\partial B_c(0)$ and plugging into (8-43) yields

$$\int_{\Omega\setminus\varphi(B_c(0)_+)}\left(|\nabla H|^2 - \frac{\gamma}{|x|^2}H^2\right)dx = \frac{(\alpha_+-1)c^{n-2\alpha_+}\omega_{n-1}}{2n} + (n-2)m_\gamma(\Omega)\frac{\omega_{n-1}}{2n} + \theta_c,$$

where $\lim_{c\to 0}\theta_c = 0$. Here, we have used that

$$\int_{\mathbb{S}^{n-1}_+} x_1^2\,d\sigma = \frac{1}{2}\int_{\mathbb{S}^{n-1}} x_1^2\,d\sigma = \frac{1}{2n}\int_{\mathbb{S}^{n-1}}|x|^2\,d\sigma = \frac{\omega_{n-1}}{2n}, \quad \omega_{n-1} := \int_{\mathbb{S}^{n-1}} d\sigma.$$

This equality and (8-42) prove (8-41).

We now claim that

$$\int_\Omega\left(|\nabla u_\epsilon|^2 - \frac{\gamma}{|x|^2}u_\epsilon^2\right)dx = \lambda\int_{\mathbb{R}^n_+}\frac{U^{2^\star(s)}}{|x|^s}\,dx + m_\gamma(\Omega)\frac{(n-2)\omega_{n-1}}{2n}\epsilon^{\alpha_+-\alpha_-} + o(\epsilon^{\alpha_+-\alpha_-}) \quad \text{as } \epsilon\to 0. \tag{8-44}$$

Indeed, define $U_\epsilon(x) := \epsilon^{-(n-2)/2}U(\epsilon^{-1}x)$ for all $x\in\mathbb{R}^n_+$. The definition (8-39) of u_ϵ can be rewritten as

$$u_\epsilon\circ\varphi(x) = U_\epsilon(x) + \epsilon^{(\alpha_+-\alpha_-)/2}\beta\circ\varphi(x) \quad \text{for all } x\in\mathbb{R}^n_+\cap\widetilde{B}_\delta.$$

Fix $c\in(0,\delta)$, which we will eventually let go to 0. Since $d\varphi_0$ is an isometry, we get that

$$\begin{aligned}
&\int_{\varphi(B_c(0)_+)}\left(|\nabla u_\epsilon|^2 - \frac{\gamma}{|x|^2}u_\epsilon^2\right)dx\\
&\quad= \int_{B_c(0)_+}\left(|\nabla(u_\epsilon\circ\varphi)|^2_{\varphi^\star\mathrm{Eucl}} - \frac{\gamma}{|\varphi(x)|^2}(u_\epsilon\circ\varphi)^2\right)|\mathrm{Jac}\,\varphi|\,dx\\
&\quad= \int_{B_c(0)_+}\left(|\nabla U_\epsilon|^2_{\varphi^\star\mathrm{Eucl}} - \frac{\gamma}{|\varphi(x)|^2}U_\epsilon^2\right)|\mathrm{Jac}\,\varphi|\,dx\\
&\qquad+ 2\epsilon^{(\alpha_+-\alpha_-)/2}\int_{B_c(0)_+}\left((\nabla U_\epsilon,\nabla(\beta\circ\varphi))_{\varphi^\star\mathrm{Eucl}} - \frac{\gamma}{|\varphi(x)|^2}U_\epsilon(u_\epsilon\circ\varphi)\right)|\mathrm{Jac}\,\varphi|\,dx\\
&\qquad+ \epsilon^{\alpha_+-\alpha_-}\int_{B_c(0)_+}\left(|\nabla(\beta\circ\varphi)|^2_{\varphi^\star\mathrm{Eucl}} - \frac{\gamma}{|\varphi(x)|^2}(\beta\circ\varphi)^2\right)|\mathrm{Jac}\,\varphi|\,dx.
\end{aligned} \tag{8-45}$$

Since $\varphi^\star\mathrm{Eucl} = \mathrm{Eucl} + O(|x|)$, $|\varphi(x)| = |x| + O(|x|^2)$ and $\beta \in D^{1,2}(\Omega)$, we get that

$$\begin{aligned}
&\int_{\varphi(B_c(0)_+)} \left(|\nabla u_\epsilon|^2 - \frac{\gamma}{|x|^2} u_\epsilon^2 \right) dx \\
&\quad = \int_{B_c(0)_+} \left(|\nabla U_\epsilon|^2_{\mathrm{Eucl}} - \frac{\gamma}{|x|^2} U_\epsilon^2 \right) dx + O\left(\int_{B_c(0)_+} |x| \left(|\nabla U_\epsilon|^2_{\mathrm{Eucl}} + \frac{U_\epsilon^2}{|x|^2} \right) dx \right) \\
&\qquad + 2\epsilon^{(\alpha_+ - \alpha_-)/2} \int_{B_c(0)_+} \left((\nabla U_\epsilon, \nabla(\beta \circ \varphi))_{\mathrm{Eucl}} - \frac{\gamma}{|x|^2} U_\epsilon (\beta \circ \varphi) \right) dx \\
&\qquad\quad + O\left(\epsilon^{(\alpha_+ - \alpha_-)/2} \int_{B_c(0)_+} |x| \left(|\nabla U_\epsilon| \cdot |\nabla(\beta \circ \varphi)| + \frac{U_\epsilon |\beta \circ \varphi|}{|x|^2} \right) dx \right) + \epsilon^{\alpha_+ - \alpha_-} \theta_c^\epsilon
\end{aligned} \tag{8-46}$$

as $\epsilon \to 0$. The pointwise estimates (8-38) yield

$$\begin{aligned}
\int_{\varphi(B_c(0)_+)} \left(|\nabla u_\epsilon|^2 - \frac{\gamma}{|x|^2} u_\epsilon^2 \right) dx &= \int_{B_c(0)_+} \left(|\nabla U_\epsilon|^2_{\mathrm{Eucl}} - \frac{\gamma}{|x|^2} U_\epsilon^2 \right) dx \\
&\quad + 2\epsilon^{(\alpha_+ - \alpha_-)/2} \int_{B_c(0)_+} \left((\nabla U_\epsilon, \nabla(\beta \circ \varphi))_{\mathrm{Eucl}} - \frac{\gamma}{|x|^2} U_\epsilon (\beta \circ \varphi) \right) dx + \epsilon^{\alpha_+ - \alpha_-} \theta_c^\epsilon
\end{aligned}$$

as $\epsilon \to 0$. Integrating by parts yields

$$\begin{aligned}
&\int_{\varphi(B_c(0)_+)} \left(|\nabla u_\epsilon|^2 - \frac{\gamma}{|x|^2} u_\epsilon^2 \right) dx \\
&\quad = \int_{B_c(0)_+} \left(-\Delta U_\epsilon - \frac{\gamma}{|x|^2} U_\epsilon \right) U_\epsilon \, dx + \int_{\partial(B_c(0)_+)} U_\epsilon \, \partial_\nu U_\epsilon \, d\sigma \\
&\qquad + 2\epsilon^{(\alpha_+ - \alpha_-)/2} \left(\int_{B_c(0)_+} \left(-\Delta U_\epsilon - \frac{\gamma}{|x|^2} U_\epsilon \right) \beta \circ \varphi \, dx + \int_{\partial(B_c(0)_+)} \beta \circ \varphi \, \partial_\nu U_\epsilon \, d\sigma \right) + \epsilon^{\alpha_+ - \alpha_-} \theta_c^\epsilon
\end{aligned}$$

as $\epsilon \to 0$. Since both U and $\beta \circ \varphi$ vanish on $\partial \mathbb{R}^n_+ \setminus \{0\}$, we get that

$$\begin{aligned}
&\int_{\varphi(B_c(0)_+)} \left(|\nabla u_\epsilon|^2 - \frac{\gamma}{|x|^2} u_\epsilon^2 \right) dx \\
&\quad = \int_{B_c(0)_+} \left(-\Delta U_\epsilon - \frac{\gamma}{|x|^2} U_\epsilon \right) U_\epsilon \, dx + \int_{\mathbb{R}^n_+ \cap \partial B_c(0)} U_\epsilon \, \partial_\nu U_\epsilon \, d\sigma \\
&\qquad + 2\epsilon^{(\alpha_+ - \alpha_-)/2} \left(\int_{B_c(0)_+} \left(-\Delta U_\epsilon - \frac{\gamma}{|x|^2} U_\epsilon \right) \beta \circ \varphi \, dx + \int_{\mathbb{R}^n_+ \cap \partial B_c(0)} \beta \circ \varphi \, \partial_\nu U_\epsilon \, d\sigma \right) + \epsilon^{\alpha_+ - \alpha_-} \theta_c^\epsilon
\end{aligned} \tag{8-47}$$

as $\epsilon \to 0$. The asymptotic estimate (8-38) of U and Lemma 5.2 yield (after a Kelvin transform)

$$\partial_\nu U_\epsilon = -(\alpha_+ - 1) \epsilon^{(\alpha_+ - \alpha_-)/2} x_1 |x|^{-\alpha_+ - 1} + o(\epsilon^{(\alpha_+ - \alpha_-)/2} |x|^{-\alpha_+})$$

as $\epsilon \to 0$ uniformly on compact subsets of $\overline{\mathbb{R}^n_+} \setminus \{0\}$. We then get that

$$\beta \circ \varphi \, \partial_\nu U_\epsilon = \epsilon^{(\alpha_+ - \alpha_-)/2} \big(-m_\gamma(\Omega)(\alpha_+ - 1) x_1^2 |x|^{-n-1} + o(|x|^{1-n}) \big)$$

and

$$U_\epsilon\,\partial_\nu U_\epsilon = \epsilon^{\alpha_+-\alpha_-}\big(-(\alpha_+-1)x_1^2|x|^{-2\alpha_+-1}+o(|x|^{1-2\alpha_+})\big)$$

as $\epsilon\to 0$ uniformly on compact subsets of $\overline{\mathbb{R}^n_+}\setminus\{0\}$. Plugging these identities into (8-47) and using equation (8-7) yields, as $\epsilon\to 0$,

$$\begin{aligned}\int_{\varphi(B_c(0)_+)}\left(|\nabla u_\epsilon|^2-\frac{\gamma}{|x|^2}u_\epsilon^2\right)dx &= \int_{B_c(0)_+}\lambda\frac{U_\epsilon^{2^\star(s)}}{|x|^s}\,dx-(\alpha_+-1)\frac{\omega_{n-1}}{2n}c^{n-2\alpha_+}\epsilon^{\alpha_+-\alpha_-}\\ &\quad+2\epsilon^{(\alpha_+-\alpha_-)/2}\int_{B_c(0)_+}\lambda\frac{U_\epsilon^{2^\star(s)-1}}{|x|^s}\beta\circ\varphi\,dx\\ &\quad-(\alpha_+-1)\frac{\omega_{n-1}}{n}m_\gamma(\Omega)\epsilon^{\alpha_+-\alpha_-}+\epsilon^{\alpha_+-\alpha_-}\theta_c^\epsilon.\end{aligned}\tag{8-48}$$

As $\epsilon\to 0$, we have that

$$\int_{B_c(0)_+}\lambda\frac{U_\epsilon^{2^\star(s)}}{|x|^s}\,dx=\int_{\mathbb{R}^n_+}\lambda\frac{U_\epsilon^{2^\star(s)}}{|x|^s}\,dx+o(\epsilon^{\alpha_+-\alpha_-}).\tag{8-49}$$

The expansion (8-37) and the change of variable $x:=\epsilon y$ yield as $\epsilon\to 0$,

$$\int_{B_c(0)_+}\lambda\frac{U_\epsilon^{2^\star(s)-1}}{|x|^s}\beta\circ\varphi\,dx=\lambda m_\gamma(\Omega)\epsilon^{(\alpha_+-\alpha_-)/2}\int_{\mathbb{R}^n_+}\frac{U^{2^\star(s)-1}}{|y|^s}\frac{y_1}{|y|^{\alpha_-}}\,dy+\epsilon^{(\alpha_+-\alpha_-)/2}\theta_\epsilon^c.\tag{8-50}$$

Integrating by parts, and using the asymptotics (8-38) for U, we have

$$\begin{aligned}&\lambda\int_{\mathbb{R}^n_+}\frac{U^{2^\star(s)-1}}{|y|^s}\frac{y_1}{|y|^{\alpha_-}}\,dy\\ &\quad=\lim_{R\to+\infty}\int_{B_R(0)_+}\lambda\frac{U^{2^\star(s)-1}}{|y|^s}\frac{y_1}{|y|^{\alpha_-}}\,dy=\lim_{R\to+\infty}\int_{B_R(0)_+}\left(-\Delta U-\frac{\gamma}{|y|^2}U\right)\frac{y_1}{|y|^{\alpha_-}}\,dy\\ &\quad=\lim_{R\to+\infty}\int_{B_R(0)_+}U\left(-\Delta-\frac{\gamma}{|y|^2}\right)\left(\frac{y_1}{|y|^{\alpha_-}}\right)dy-\int_{\partial B_R(0)_+}\partial_\nu U\frac{y_1}{|y|^{\alpha_-}}\,d\sigma=(\alpha_+-1)\frac{\omega_{n-1}}{2n}.\end{aligned}\tag{8-51}$$

Putting together (8-49)–(8-51) gives

$$\int_\Omega\left(|\nabla u_\epsilon|^2-\frac{\gamma}{|x|^2}u_\epsilon^2\right)dx=\lambda\int_{\mathbb{R}^n_+}\frac{U^{2^\star(s)}}{|x|^s}\,dx+m_\gamma(\Omega)\frac{(n-2)\omega_{n-1}}{2n}\epsilon^{\alpha_+-\alpha_-}+o(\epsilon^{\alpha_+-\alpha_-})$$

as $\epsilon\to 0$. This finally yields (8-44).

We finally claim that

$$\int_\Omega\frac{u_\epsilon^{2^\star(s)}}{|x|^s}\,dx=\int_{\mathbb{R}^n_+}\frac{U^{2^\star(s)}}{|x|^s}\,dx+\frac{2^\star(s)}{\lambda}m_\gamma(\Omega)\frac{(\alpha_+-1)\omega_{n-1}}{2n}\epsilon^{\alpha_+-\alpha_-}+o(\epsilon^{\alpha_+-\alpha_-})\quad\text{as }\epsilon\to 0.\tag{8-52}$$

Indeed, fix $c > 0$. Due to estimates (8-37) and (8-38), we have that

$$\begin{aligned}\int_\Omega \frac{u_\epsilon^{2^\star(s)}}{|x|^s}\,dx &= \int_{\varphi(B_c(0)_+)} \frac{u_\epsilon^{2^\star(s)}}{|x|^s}\,dx + o(\epsilon^{\alpha_+-\alpha_-})\\ &= \int_{B_c(0)_+} \frac{|U_\epsilon + \epsilon^{(\alpha_+-\alpha_-)/2}\beta\circ\varphi|^{2^\star(s)}}{|\varphi(x)|^s}\,|\mathrm{Jac}\,\varphi|\,dx + o(\epsilon^{\alpha_+-\alpha_-})\\ &= \int_{B_c(0)_+} \frac{|U_\epsilon + \epsilon^{(\alpha_+-\alpha_-)/2}\beta\circ\varphi|^{2^\star(s)}}{|x|^s}\,(1+O(|x|))\,dx + o(\epsilon^{\alpha_+-\alpha_-})\end{aligned}$$

as $\epsilon \to 0$. As one checks, there exists $C > 0$ such that for all $X, Y \in \mathbb{R}$,

$$\big||X+Y|^{2^\star(s)} - |X|^{2^\star(s)} - 2^\star(s)|X|^{2^\star(s)-2}XY\big| \le C(|X|^{2^\star(s)-2}|Y|^2 + |Y|^{2^\star(s)}). \tag{8-53}$$

Therefore, using the asymptotics (8-37) and (8-38) of U and β, we get that

$$\begin{aligned}\int_\Omega \frac{u_\epsilon^{2^\star(s)}}{|x|^s}\,dx &= \int_{B_c(0)_+} \frac{U_\epsilon^{2^\star(s)}}{|x|^s}|(1+O(|x|))\,dx\\ &\qquad +2^\star(s)\epsilon^{(\alpha_+-\alpha_-)/2}\int_{B_c(0)_+}\frac{U_\epsilon^{2^\star(s)-1}}{|x|^s}\beta\circ\varphi(1+O(|x|))\,dx+\epsilon^{(\alpha_+-\alpha_-)/2}\theta_\epsilon^c\\ &= \int_{B_c(0)_+} \frac{U_\epsilon^{2^\star(s)}}{|x|^s}\,dx+2^\star(s)\epsilon^{(\alpha_+-\alpha_-)/2}\int_{B_c(0)_+}\frac{U_\epsilon^{2^\star(s)-1}}{|x|^s}\beta\circ\varphi\,dx+\epsilon^{(\alpha_+-\alpha_-)/2}\theta_\epsilon^c \quad \text{as } \epsilon\to 0.\end{aligned}$$

Then (8-52) follows from this latest identity, combined with (8-49)–(8-51).

We finally use (8-31), (8-44) and (8-52) to get

$$J^\Omega_{\gamma,s}(u_\epsilon) = J^{\mathbb{R}^n_+}_{\gamma,s}(U)\bigg(1 - \frac{\big(\alpha_+ - \frac{1}{2}n\big)\omega_{n-1}}{n\lambda\int_{\mathbb{R}^n_+} U^{2^\star(s)}/|x|^s\,dx}m_\gamma(\Omega)\epsilon^{\alpha_+-\alpha_-} + o(\epsilon^{\alpha_+-\alpha_-})\bigg) \quad \text{as } \epsilon\to 0,$$

which proves (8-6). This completes Proposition 8.3 and therefore Theorem 8.2. □

9. Domains with positive mass and an arbitrary geometry at 0

In this section, we construct smooth bounded domains in $\mathbb{R}^n$ with positive or negative mass, regardless of the local geometry of $\partial\Omega$ at 0. This is illustrated by the following result.

Theorem 9.1. *Let ω be a smooth open set of $\mathbb{R}^n$. Then, there exist $r_0 > 0$ and two smooth bounded domains Ω_+, Ω_- of $\mathbb{R}^n$ such that*

$$\Omega_+ \cap B_{r_0}(0) = \Omega_- \cap B_{r_0}(0) = \omega \cap B_{r_0}(0), \tag{9-1}$$

$$\min\{\gamma_H(\Omega_+), \gamma_H(\Omega_-)\} > \tfrac{1}{4}(n^2-1), \tag{9-2}$$

$$m_\gamma(\Omega_+) > 0 > m_\gamma(\Omega_-), \tag{9-3}$$

whenever $\frac{1}{4}(n^2-1) < \gamma < \min\{\gamma_H(\Omega_+), \gamma_H(\Omega_-)\}$.

We shall need the following stability result for the mass under continuous deformations and truncations.

Proposition 9.2. *Let $\Omega \subset \mathbb{R}^n$ be a conformally bounded domain such that $0 \in \partial\Omega$. Assume that $\gamma_H(\Omega) > \frac{1}{4}(n^2-1)$ and fix $\gamma \in \big(\frac{1}{4}(n^2-1), \gamma_H(\Omega)\big)$. For any $R > 0$, let D_R be a smooth domain of $\mathbb{R}^n$ such that*

- $B_R(x_0) \subset D_R \subset B_{2R}(x_0)$,
- $\Omega \cap D_R$ *is a smooth domain of* $\mathbb{R}^n$.

Let $\Phi \in C^\infty(\mathbb{R} \times \mathbb{R}^n, \mathbb{R}^n)$ be such that

- $\Phi_t := \Phi(t, \cdot)$ *is a smooth diffeomorphism of* $\mathbb{R}^n$,
- $\Phi_t(x) = x$ *for all* $|x| > \frac{1}{2}$ *and all* $t \in \mathbb{R}$,
- $\Phi_t(0) = 0$ *for all* $t \in \mathbb{R}$,
- $\Phi_0 = \mathrm{Id}_{\mathbb{R}^n}$.

Set $\Omega_{t,R} := \Phi_t(\Omega) \cap D_R$. Then as $t \to 0$, $R \to +\infty$, we have that $\gamma_H(\Omega_{t,R}) > \frac{1}{4}(n^2-1)$ and $m_\gamma(\Omega_{t,R})$ is well defined. In addition,

$$\lim_{t\to 0,\, R\to+\infty} m_\gamma(\Omega_{t,R}) = m_\gamma(\Omega).$$

As a preliminary remark, we claim that if Ω is a conformally bounded domain of $\mathbb{R}^n$ such that $0 \in \partial\Omega$, then

$$\liminf_{t\to 0, R\to\infty} \gamma_H(\Omega_{t,R}) \geq \gamma_H(\Omega), \tag{9-4}$$

where $\Omega_{t,R}$ are defined as in Proposition 9.2. Indeed, by definition, $\gamma_H(\Omega_{t,R}) \geq \gamma_H(\Omega_t) = \gamma_H(\Phi_t(\Omega))$. Inequality (9-4) then follows from (3-7) of Lemma 3.2.

We shall use the same approach as in the proof of Proposition 7.4. Assuming $x_0 := (-1, 0, \ldots, 0) \in \mathbb{R}^n$, and denoting the corresponding Kelvin inversion by i, this transformation allows us to map the operator $-\Delta - \gamma/|x|^2$ on a conformally bounded domain Ω into the Schrödinger operator $-\Delta + V$ on the bounded domain $\widetilde{\Omega}$, where V is the potential defined in (7-19).

Set now $\widetilde{\Omega} := i(\Omega)$, $\widetilde{\Phi}(t, x) := i \circ \Phi(t, i(x))$ for $(t, x) \in \mathbb{R} \times \mathbb{R}^n$, and $\widetilde{D}_r := \mathbb{R}^n \setminus i(D_{r^{-1}})$ in $\mathbb{R}^n$. Observe that $R \to +\infty$ in Proposition 9.2 is equivalent to $r \to 0$ in here. Note that $\widetilde{\Phi} \in C^\infty(\mathbb{R} \times \mathbb{R}^n, \mathbb{R}^n)$ is such that:

- For any $t \in (-2, 2)$, the map $\widetilde{\Phi}_t := \widetilde{\Phi}(t, \cdot)$ is a C^∞-diffeomorphism onto its open image $\widetilde{\Phi}_t(\mathbb{R}^n)$.
- $\widetilde{\Phi}_0 = \mathrm{Id}$.
- $\widetilde{\Phi}_t(0) = 0$ for all $t \in (-2, 2)$.
- $\widetilde{\Phi}_t(x) = x$ for all $t \in (-2, 2)$ and all $x \in B_{2\delta}(x_0)$ with $\delta < \frac{1}{4}$.

Set $\widetilde{\Omega}_t := \widetilde{\Phi}_t(\widetilde{\Omega})$ and note that the sets $\widetilde{D}_r$ satisfy the following properties:

- $B_{r/2}(x_0) \subset \widetilde{D}_r \subset B_r(x_0)$.
- $\widetilde{\Omega}_{t,r} := \widetilde{\Omega}_t \setminus \widetilde{D}_r$ is a smooth domain of $\mathbb{R}^n$.

In particular, we have that $\widetilde{\Omega}_{t,r} = i(\Omega_{t,r^{-1}})$. Let $u \in C^2(\overline{\Omega}_{t,r} \setminus \{0\})$ be such that

$$\begin{cases} -\Delta u - \dfrac{\gamma}{|x|^2} u = 0 & \text{in } \Omega_{t,r}, \\ u > 0 & \text{in } \Omega_{t,r}, \\ u = 0 & \text{on } \partial\Omega_{t,r}. \end{cases}$$

We shall need the following.

Lemma 9.3. *For any $t \in (-1, 1)$, there exists $u_t \in C^2(\overline{\widetilde{\Omega}}_t \setminus \{0, x_0\})$ such that*

$$\begin{cases} -\Delta u_t - V u_t = 0 & \text{in } \widetilde{\Omega}_t, \\ u_t > 0 & \text{in } \widetilde{\Omega}_t, \\ u_t = 0 & \text{on } \partial\widetilde{\Omega}_t \setminus \{0, x_0\}, \\ u_t(x) \le C|x|^{1-\alpha_+(\gamma)} + C|x - x_0|^{1-\alpha_-(\gamma)} & \text{for } x \in \widetilde{\Omega}_t. \end{cases} \tag{9-5}$$

Moreover, we have that

$$u_t(x) = \frac{d(x, \partial\widetilde{\Omega}_t)}{|x|^{\alpha_+(\gamma)}}(1 + O(|x|^{\alpha_+(\gamma)-\alpha_-(\gamma)})) \tag{9-6}$$

as $x \to 0$, uniformly with respect to $t \in (-1, 1)$.

Proof. We construct approximate singular solutions as in Section 4. For all $t \in (-2, 2)$, there exists a chart φ_t that satisfies (4-7)–(4-12) for $\widetilde{\Omega}_t$. Without restriction, we assume that $\lim_{t\to 0} \varphi_t = \varphi_0$ in $C^k(\widetilde{B}_{2\delta}, \mathbb{R}^n)$. We define a cut-off function η_δ such that $\eta_\delta(x) = 1$ for $x \in \widetilde{B}_\delta$ and $\eta_\delta(x) = 0$ for $x \notin \widetilde{B}_{2\delta}$. As in (4-14), we define $u_{\alpha_+(\gamma),t} \in C^2(\overline{\widetilde{\Omega}}_t \setminus \{0\})$ with compact support in $\varphi_t(\widetilde{B}_{2\delta})$ such that

$$u_{\alpha_+,t} \circ \varphi_t(x_1, x') := \eta_\delta(x_1, x') x_1 |x|^{-\alpha_+}(1 + \Theta_t(x)) \quad \text{for all } (x_1, x') \in \widetilde{B}_{2\delta} \setminus \{0\}, \tag{9-7}$$

where $\Theta_t(x_1, x') := e^{-x_1 H_t(x')/2} - 1$ for all $x = (x_1, x') \in \widetilde{B}_{2\delta}$ and all $t \in (-2, 2)$. Here, $H_t(x')$ is the mean curvature of $\partial\widetilde{\Omega}_t$ at the point $\varphi_t(0, x')$. Note that $\lim_{t\to 0} \Theta_t = \Theta_0$ in $C^k(U)$. Arguing as in Section 4, we get that

$$\begin{cases} (-\Delta - V) u_{\alpha_+,t} = O(d(x, \partial\widetilde{\Omega}_t)|x|^{-\alpha_+(\gamma)-1}) & \text{in } \widetilde{\Omega}_t \cap \widetilde{B}_\delta, \\ u_{\alpha_+,t} > 0 & \text{in } \widetilde{\Omega}_t \cap \widetilde{B}_\delta, \\ u_{\alpha_+,t} = 0 & \text{on } \partial\widetilde{\Omega}_t \setminus \{0\}, \end{cases}$$

and

$$u_{\alpha_+,t}(x) = \frac{d(x, \partial\widetilde{\Omega}_t)}{|x|^{\alpha_+(\gamma)}}(1 + O(|x|) \quad \text{as } x \to 0.$$

The construction in Section 4 also yields

$$\lim_{t\to 0} u_{\alpha_+,t} \circ \Phi_t = u_{\alpha_+,0} \quad \text{in } C^2_{\text{loc}}(\overline{\widetilde{\Omega}} \setminus \{0\}). \tag{9-8}$$

Note also that all these estimates are uniform in $t \in (-1, 1)$. In particular, defining

$$f_t := -\Delta u_{\alpha_+,t} - V u_{\alpha_+,t}, \tag{9-9}$$

there exists $C > 0$ such that

$$|f_t(x)| \le Cd(x, \partial\widetilde{\Omega}_t)|x|^{-\alpha_+(\gamma)-1} \le C|x|^{-\alpha_+(\gamma)} \tag{9-10}$$

for all $t \in (-1, 1)$ and all $x \in \widetilde{\Omega}_t \cap \widetilde{B}_\delta$. Therefore, since $\gamma > \frac{1}{4}(n^2-1)$, it follows from (9-8) and this pointwise control that $f_t \in L^{2n/(n+2)}(\widetilde{\Omega}_t)$ for all $t \in (-1, 1)$ and that

$$\lim_{t\to 0} \|f_t \circ \Phi_t - f_0\|_{L^{2n/(n+2)}(\widetilde{\Omega})} = 0. \tag{9-11}$$

For any $t \in (-1, 1)$, we let $v_t \in D^{1,2}(\widetilde{\Omega}_t)$ be such that

$$-\Delta v_t - V v_t = f_t \quad \text{weakly in } D^{1,2}(\widetilde{\Omega}_t).$$

The existence follows from the coercivity of $-\Delta - V$ on $\widetilde{\Omega}_t$, which follows itself from the coercivity on $\widetilde{\Omega} = \widetilde{\Omega}_0$. We then get from (9-11) and the uniform coercivity on $\widetilde{\Omega}_t$ that

$$\lim_{t\to 0} v_t \circ \Phi_t = v_0 \quad \text{in } D^{1,2}(\widetilde{\Omega}) \text{ and } C^1_{\text{loc}}(\overline{\widetilde{\Omega}} \setminus \{0, x_0\}).$$

It follows from the construction of the mass in Section 7 (see the proof of Theorem 7.1) that around 0, $|v_t(x)|$ is bounded by $|x|^{1-\alpha_-(\gamma)}$. Around x_0, we know $-\Delta v_t - V v_t = 0$ and the regularity theorem, Theorem 4.1, yields a control by $|x - x_0|^{1-\alpha_-(\gamma)}$. These controls are uniform with respect to $t \in (-1, 1)$. Therefore, there exists $C > 0$ such that

$$|v_t(x)| \le Cd(x, \partial\widetilde{\Omega}_t)\big(|x|^{-\alpha_-(\gamma)} + |x - x_0|^{-\alpha_-(\gamma)}\big)$$

for all $t \in (-1, 1)$ and all $x \in \widetilde{\Omega}_t$. Now define $u_t(x) := u_{\alpha_+,t}(x) - v_t(x)$ for all $t \in (-1, 1)$ and $x \in \widetilde{\Omega}_t$. This function satisfies all the requirements of Lemma 9.3. □

Proof of Proposition 9.2. Let $\widetilde{\Omega}_{t,r} = \widetilde{\Omega}_t \setminus \widetilde{D}_r$, and note that for $r \in \big(0, \frac{1}{2}\delta\big)$, we have $\widetilde{\Omega}_{t,r} \cap B_\delta(0) = \widetilde{\Omega} \cap B_\delta(0)$. We shall define a mass associated to the potential V as in Proposition 7.4 and prove its continuity.

Step 1: The function $f_t : \widetilde{\Omega}_t \to \mathbb{R}$ defined in (9-9) has compact support in $B_{2\delta}(0)$; therefore, it is well defined also on $\widetilde{\Omega}_{t,r}$. Let $v_{t,r} \in D^{1,2}(\widetilde{\Omega}_{t,r})$ be such that

$$-\Delta v_{t,r} - V v_{t,r} = f_t \quad \text{weakly in } D^{1,2}(\widetilde{\Omega}_{t,r}). \tag{9-12}$$

Since the operator $-\Delta - V$ is uniformly coercive on $\widetilde{\Omega}_t$, it is also uniformly coercive on $\widetilde{\Omega}_{t,r}$ with respect to (t, r), so that the definition of $v_{t,r}$ via (9-12) makes sense. The uniform coercivity and (9-9)–(9-10) yield the existence of $C > 0$ such that $\|v_{t,r}\|_{D^{1,2}(\widetilde{\Omega}_{t,r})} \le C$ for all t, r. Since $x_0 \notin \widetilde{\Omega}_{t,r}$, (9-9)–(9-10) and regularity theory yield $v_{t,r} \in C^1(\overline{\widetilde{\Omega}}_{t,r} \setminus \{0\})$ and for all $\rho > 0$, there exists $C(\rho) > 0$ independent of t and r such that

$$\|v_{t,r}\|_{C^1(\widetilde{\Omega}_{t,r} \setminus (B_\rho(0) \cup B_\rho(x_0)))} \le C(\rho). \tag{9-13}$$

Step 2: There exists $C > 0$ such that for all $t \in (-1, 1)$ and all $x \in \widetilde{\Omega}_{t,r}$,

$$|v_{t,r}(x)| \le Cd(x, \partial\Omega_t)(|x|^{-\alpha_-(\gamma)} + |x - x_0|^{-\alpha_-(\gamma)}). \tag{9-14}$$

Indeed, around 0, we know $\widetilde{\Omega}_{t,r}$ coincides with $\widetilde{\Omega}_t$, and the proof of the control goes as in the construction of the mass in Section 7 (see the proof of Theorem 7.1). The argument is different around x_0. We let $r_0 > 0$ be such that $\widetilde{\Omega}_t \cap B_{2r_0}(x_0) = \widetilde{\Omega} \cap B_{2r_0}(x_0)$. Therefore, for $r \in (0, r_0)$, we have that

$$\widetilde{\Omega}_{t,r} \cap B_{2r_0}(x_0) = (\widetilde{\Omega} \setminus \widetilde{D}_r) \cap B_{2r_0}(x_0).$$

Arguing as in the proof of Proposition 4.3, there exists $\tilde{u}_{\alpha_-} \in C^\infty(\overline{\widetilde{\Omega}} \setminus \{0\})$ and $\tau' > 0$ such that

$$\begin{cases} \tilde{u}_{\alpha_-} > 0 & \text{in } \widetilde{\Omega} \cap B_{2r_0}(x_0), \\ \tilde{u}_{\alpha_-} = 0 & \text{in } (\partial\widetilde{\Omega}) \cap B_{2r_0}(x_0), \\ -\Delta\tilde{u}_{\alpha_-} - V\tilde{u}_{\alpha_-} > 0 & \text{in } \widetilde{\Omega} \cap B_{2r_0}(x_0). \end{cases}$$

Moreover, we have that

$$\tilde{u}_{\alpha_-}(x) = \frac{d(x, \partial\widetilde{\Omega})}{|x - x_0|^{\alpha_-}}(1 + O(|x - x_0|)) \quad \text{as } x \to x_0,\ x \in \widetilde{\Omega}. \tag{9-15}$$

Therefore, since $v_{t,r}$ vanishes on $B_{2r_0}(x_0) \cap \partial(\widetilde{\Omega} \setminus \widetilde{D}_r)$, it follows from (9-13) and the properties of $\tilde{u}_{\alpha_-}$ that there exists $C > 0$ such that $v_{t,r} \le C\tilde{u}_{\alpha_-}$ on the boundary of $(\widetilde{\Omega} \cap \widetilde{D}_r) \cap B_{2r_0}(x_0)$. Since in addition $(-\Delta - V)v_{t,r} = 0 < (-\Delta - V)(C\tilde{u}_{\alpha_-})$, it follows from the comparison principle that $v_{t,r} \le C\tilde{u}_{\alpha_-}$ in $(\widetilde{\Omega} \setminus \widetilde{D}_r) \cap B_{2r_0}(x_0)$. Arguing similarly with $-v_{t,r}$ and using the asymptotic (9-15), we get (9-14).

Step 3: We have

$$\lim_{t,r\to 0} v_{t,r} \circ \Phi_t = v_0 \quad \text{in } D^{1,2}(\widetilde{\Omega})_{\mathrm{loc},\{x_0\}^c} \cap C^1_{\mathrm{loc}}(\overline{\widetilde{\Omega}} \setminus \{0, x_0\}), \tag{9-16}$$

where v_0 was defined in (7-20), and the convergence in $D^{1,2}(\widetilde{\Omega})_{\mathrm{loc},\{x_0\}^c}$ means that $\lim_{t,r\to 0} \eta v_{t,r} \circ \Phi_t = \eta v_0$ in $D^{1,2}(\widetilde{\Omega})$ for all $\eta \in C^\infty(\mathbb{R}^n)$ vanishing around x_0. Indeed, $v_{t,r} \circ \Phi_t \in D^{1,2}(\widetilde{\Omega} \setminus \widetilde{D}_r) \subset D^{1,2}(\widetilde{\Omega})$. Uniform coercivity yields weak convergence in $D^{1,2}(\widetilde{\Omega})$ to $\tilde{v} \in D^{1,2}(\widetilde{\Omega})$. Passing to the limit, one gets $(-\Delta - V)\tilde{v} = f_0$, so that $\tilde{v} = v_0$. Uniqueness then yields convergence in $C^1_{\mathrm{loc}}(\overline{\widetilde{\Omega}} \setminus \{0, x_0\})$. With a change of variable, (9-12) yields an elliptic equation for $v_{t,r} \circ \Phi_t$. Multiplying this equation by $\eta^2 \cdot (v_{t,r} \circ \Phi_t - v_0)$ for $\eta \in C^\infty(\mathbb{R}^n)$ vanishing around x_0, one gets convergence of $\eta v_{t,r} \circ \Phi_t$ to ηv_0 in $D^{1,2}(\widetilde{\Omega})$. This proves the claim.

It follows from the construction of the mass (see Theorem 7.1) and the regularity theorem, Theorem 4.1, that there exists $K_0 \in \mathbb{R}$ and for all (t, r) small, there exists $K_{t,r} \in \mathbb{R}$ such that

$$v_{t,r}(x) = K_{t,r}\frac{d(x, \partial\widetilde{\Omega}_t)}{|x|^{\alpha_-(\gamma)}} + o\left(\frac{d(x, \partial\widetilde{\Omega}_t)}{|x|^{\alpha_-(\gamma)}}\right) \quad \text{and} \quad v_0(x) = K_0\frac{d(x, \partial\widetilde{\Omega})}{|x|^{\alpha_-(\gamma)}} + o\left(\frac{d(x, \partial\widetilde{\Omega})}{|x|^{\alpha_-(\gamma)}}\right) \tag{9-17}$$

as $x \in \widetilde{\Omega}$ goes to 0. Note that around 0, we know $\widetilde{\Omega}_{t,r}$ coincides with $\widetilde{\Omega}_t$.

Step 4: We claim that

$$\lim_{t,r\to 0} K_{t,r} = K_0. \tag{9-18}$$

We only give a sketch. Noting $\tilde{v}_{t,r} := v_{t,r} \circ \Phi_t$, the proof relies on (9-16) and the fact that

$$-\Delta_{\Phi_t^* \mathrm{Eucl}}\tilde{v}_{t,r} - V \circ \Phi_t \tilde{v}_{t,r} = f_t \circ \Phi_t \quad \text{in } \widetilde{\Omega} \cap B_\delta(0).$$

The comparison principle and the definitions (9-17) then yield (9-18).

Note that

$$m_\gamma(\Omega) = -K_0, \tag{9-19}$$

where the mass of a conformally bounded Ω is defined as in Proposition 7.4.

Step 5: convergence of the mass. We claim that

$$\lim_{t\to 0, R\to\infty} m_\gamma(\Omega_{t,R}) = m_\gamma(\Omega). \tag{9-20}$$

We define $\widetilde{H}_{t,r} := u_{\alpha_+,t} - v_{t,r}$ so that

$$-\Delta \widetilde{H}_{t,r} - V\widetilde{H}_{t,r} = 0 \quad \text{in } \widetilde{\Omega}_{t,r}.$$

It follows from (9-6) and (9-17) that $\widetilde{H}_{t,r} > 0$ around 0. From the maximum principle, we deduce that $\widetilde{H}_{t,r} > 0$ on $\widetilde{\Omega}_{t,r}$ and that it vanishes on $\partial\widetilde{\Omega}_{t,r} \setminus \{0, x_0\}$.

It follows from (9-6) and (9-17) that

$$\widetilde{H}_{t,r}(x) = \frac{d(x, \partial\widetilde{\Omega}_{t,r})}{|x|^{\alpha_+}} - K_{t,r}\frac{d(x, \partial\widetilde{\Omega}_{t,r})}{|x|^{\alpha_-}} + o\left(\frac{d(x, \partial\widetilde{\Omega}_{t,r})}{|x|^{\alpha_-}}\right)$$

as $x \to 0$, $x \in \widetilde{\Omega}_{t,r}$. Coming back to $\Omega_{t,R}$ with $R = r^{-1}$ via the inversion i with

$$H_{t,R}(x) := |x - x_0|^{2-n}\widetilde{H}_{t,r}(i(x))$$

for all $x \in \Omega_{t,R}$, we get that

$$\begin{cases} -\Delta H_{t,R} - \dfrac{\gamma}{|x|^2}H_{t,R} = 0 & \text{in } \Omega_{t,R}, \\ H_{t,R} > 0 & \text{in } \Omega_{t,R}, \\ H_{t,R} = 0 & \text{in } \partial\Omega_{t,R} \setminus \{0\} \end{cases}$$

and

$$H_{t,R}(x) = \frac{d(x, \partial\Omega_{t,R})}{|x|^{\alpha_+}} - K_{t,r}\frac{d(x, \partial\Omega_{t,R})}{|x|^{\alpha_-}} + o\left(\frac{d(x, \partial\Omega_{t,R})}{|x|^{\alpha_-}}\right)$$

as $x \to 0$, $x \in \Omega_{t,R}$. Therefore, it follows from the definition of the mass (see Theorem 7.1) that $m_\gamma(\Omega_{t,R}) = -K_{t,r}$ for all t, r, $R = r^{-1}$. Claim (9-20) then follows from (9-18) and (9-19). □

In order to prove Theorem 9.1, we need to exhibit prototypes of unbounded domains with either positive or negative mass.

Proposition 9.4. *Let Ω be a domain such that $0 \in \partial\Omega$ and Ω is conformally bounded. Assume that $\gamma_H(\Omega) > \frac{1}{4}(n^2-1)$ and fix $\gamma \in \left(\frac{1}{4}(n^2-1), \gamma_H(\Omega)\right)$. Then $m_\gamma(\Omega) > 0$ if $\mathbb{R}^n_+ \subsetneq \Omega$, and $m_\gamma(\Omega) < 0$ if $\Omega \subsetneq \mathbb{R}^n_+$.*

Proof. With H_0 defined as in (7-22), we set

$$\mathcal{U}(x) := H_0(x) - x_1|x|^{-\alpha_+} \quad \text{for all } x \in \Omega.$$

We first assume that $\mathbb{R}^n_+ \subsetneq \Omega$. We then have that

$$\begin{cases} -\Delta \mathcal{U} - \dfrac{\gamma}{|x|^2}\mathcal{U} = 0 & \text{in } \mathbb{R}^n_+, \\ \mathcal{U} \gneqq 0 & \text{in } \partial\mathbb{R}^n_+ \setminus \{0\}. \end{cases} \tag{9-21}$$

We claim that

$$\int_{\mathbb{R}^n_+} |\nabla \mathcal{U}|^2 \, dx < +\infty. \tag{9-22}$$

Indeed, at infinity, this is the consequence of the fact that $|\nabla \mathcal{U}|(x) \le C|x|^{-\alpha_+}$ for all $x \in \mathbb{R}^n_+$ large, this latest bound being a consequence of (7-25) combined with elliptic regularity theory. At 0, the argument is different. Indeed, one first notes that $d(x, \partial\Omega') = x_1 + O(|x|^2)$ for $x \in \mathbb{R}^n_+$ close to 0, and therefore, $\mathcal{U}(x) = O(|x|^{1-\alpha_-})$ for $x \to 0$. The control on the gradient $|\nabla \mathcal{U}|(x) \le C|x|^{-\alpha_-}$ at 0 follows from the construction of $\widetilde{H}_0$. This yields integrability at 0 and proves (9-22).

We claim that $\mathcal{U} > 0$ in $\mathbb{R}^n_+$. Indeed, it follows from (9-21) and (9-22) that $\mathcal{U}_- \in D^{1,2}(\mathbb{R}^n_+)$. Multiplying equation (7-23) by $\mathcal{U}_-$, integrating by parts on $(B_R(0) \setminus B_\epsilon(0)) \cap \mathbb{R}^n_+$, and letting $\epsilon \to 0$ and $R \to +\infty$ by using (9-22), one gets $\mathcal{U}_- \equiv 0$, and then $\mathcal{U} \ge 0$. The result follows from Hopf's maximum principle.

We now claim that

$$m_\gamma(\Omega) > 0. \tag{9-23}$$

Indeed, since $\mathcal{U} > 0$ in $\mathbb{R}^n_+$, there exists $c_0 > 0$ such that $\mathcal{U}(x) \ge c_0 x_1 |x|^{-\alpha_-}$ for all $x \in \partial(B_1(0)_+)$. It then follows from (9-22), (9-21) and the comparison principle that $\mathcal{U}(x) \ge c_0 x_1 |x|^{-\alpha_-}$ for all $x \in B_1(0)_+$. The expansion (7-24) then yields $-K_0 \ge c_0 > 0$. This combined with (9-19) proves the claim.

When $\Omega \subset \mathbb{R}^n_+$, the argument is similar except that one works on Ω (and not $\mathbb{R}^n_+$) and that $\mathcal{U} \lneqq 0$ in $\partial\Omega \setminus \{0\}$. This ends the proof of Proposition 9.4. □

Proof of Theorem 9.1. Let ω be a smooth domain of $\mathbb{R}^n$ such that $0 \in \partial\Omega$. Up to a rotation, there exists $\varphi \in C^\infty(\mathbb{R}^{n-1})$ such that $\varphi(0) = 0$, $\nabla\varphi(0) = 0$ and there exists $\delta_0 > 0$ such that

$$\omega \cap B_{\delta_0}(0) = \{x_1 > \varphi(x') : (x_1, x') \in B_{\delta_0}(0)\}.$$

Let $\eta \in C^\infty_c(B_{\delta_0}(0))$ be such that $\eta(x) = 1$ for all $x \in B_{\delta_0/2}(0)$, and define

$$\Phi_t(x) := \left(x_1 + \eta(x)\frac{\varphi(tx')}{t}, x'\right) \quad \text{for all } t > 0 \text{ and } x \in \mathbb{R}^n,$$

and $\Phi_0 := \mathrm{Id}_{\mathbb{R}^n}$. It is easy to see that Φ_t satisfies the hypotheses of Proposition 9.2. Moreover, for $0 < t < 1$, we have that

$$\frac{\omega}{t} \cap \Phi_t(B_{\delta_0/2}(0)) = \Phi_t(\mathbb{R}^n_+ \cap B_{\delta_0/2}(0)).$$

We let Ω be a smooth domain at infinity such that

$$\Omega \cap B_1(0) = \mathbb{R}^n_+ \cap B_1(0) \quad \text{and} \quad \gamma_H(\Omega) > \tfrac{1}{4}(n^2 - 1), \tag{9-24}$$

(for example, $\mathbb{R}^n_+$), and let $\Omega_{t,R}$ be as in Proposition 9.2. It is easy to see that

$$\omega \cap t\Phi_t(B_{\delta_0/2}(0)) = t\Omega_{t,R} \cap t\Phi_t(B_{\delta_0/2}(0)).$$

Therefore, for $t > 0$ small enough, we have that

$$\omega \cap B_{t\delta_0/3}(0) = t\Omega_{t,R} \cap B_{t\delta_0/3}(0).$$

Moreover, $\gamma_H(t\Omega_{t,R}) = \gamma_H(\Omega_{t,R}) > \frac{1}{4}(n^2-1)$ as $t \to 0$ and $R \to +\infty$; see (9-4). Concerning the mass, we have

$$t^{\alpha_+(\gamma)-\alpha_-(\gamma)} m_\gamma(t\Omega_{t,R}) = m_\gamma(\Omega_{t,R}) \to m_\gamma(\Omega) \quad \text{as } t \to 0,\ R \to +\infty.$$

We now choose Ω appropriately.

To get a negative mass, we choose Ω smooth at infinity such that $\Omega \cap B_1(0) = \mathbb{R}^n_+ \cap B_1(0)$ and $\Omega \subsetneq \mathbb{R}^n_+$. Then $\gamma_H(\Omega) = \frac{1}{4}n^2$, (9-24) holds and Proposition 9.4 yields $m_\gamma(\Omega) < 0$. With this choice of Ω, we take $\Omega_- := \Omega_{t,R}$ for t small and R large.

To get a positive mass, we choose $\mathbb{R}^n_+ \subsetneq \Omega$ such that (9-24) holds (this is possible for any value of $\gamma_H(\Omega)$ arbitrarily close to $\frac{1}{4}n^2$, see point (5) of Proposition 3.1). Then Proposition 9.4 yields $m_\gamma(\Omega) > 0$. With this choice of Ω, we take $\Omega_+ := \Omega_{t,R}$ for t small and R large. This proves Theorem 9.1. □

10. The Hardy singular interior mass and the remaining cases

The remaining situation not covered by Proposition 8.1 and Theorem 8.2 is $s = 0$, $n = 3$ and $\gamma \in \left(0, \frac{1}{4}n^2\right)$. If $\gamma \geq \gamma_H(\Omega)$, then Proposition 3.3 and Theorem 3.6 yield $\mu_{\gamma,0}(\Omega) \leq 0 < \mu_{\gamma,0}(\mathbb{R}^n_+)$ and the existence of extremals is guaranteed. When $\mu_{\gamma,0}(\mathbb{R}^n_+)$ does have an extremal U, then Proposition 8.3 and Theorem 3.6 provide sufficient conditions for the existence of extremals. The rest of this section addresses the remaining case, that is, when $\gamma \in (0, \gamma_H(\Omega))$ and when $\mu_{\gamma,0}(\mathbb{R}^n_+)$ has no extremal, and therefore $\mu_{\gamma,0}(\mathbb{R}^3_+) = 1/K(3,2)^2$ according to Proposition 1.3.

We first define the "interior" mass in the spirit of Schoen and Yau [1988].

Proposition 10.1. *Let $\Omega \subset \mathbb{R}^3$ be an open smooth bounded domain such that $0 \in \partial\Omega$. Fix $x_0 \in \Omega$. If $\gamma \in (0, \gamma_H(\Omega))$, then the equation*

$$\begin{cases} -\Delta G - \dfrac{\gamma}{|x|^2} G = 0 & \text{in } \Omega \setminus \{x_0\}, \\ G > 0 & \text{in } \Omega \setminus \{x_0\}, \\ G = 0 & \text{on } \partial\Omega \setminus \{0\} \end{cases}$$

has a solution $G \in C^2(\overline{\Omega} \setminus \{0, x_0\}) \cap D^2_1(\Omega \setminus \{x_0\})_{\mathrm{loc},0}$ that is unique up to multiplication by a constant. Moreover, for any $x_0 \in \Omega$, there exists a unique $R_\gamma(x_0) \in \mathbb{R}$ independent of the choice of G and $c_G > 0$ such that

$$G(x) = c_G\left(\frac{1}{|x - x_0|} + R_\gamma(x_0)\right) + o(1) \quad \text{as } x \to x_0.$$

Proof. Since $\gamma < \gamma_H(\Omega)$, the operator $-\Delta - \gamma|x|^{-2}$ is coercive and we can consider G to be its Green's function at x_0 on Ω with Dirichlet boundary condition. In particular, for any $\varphi \in C_c^\infty(\Omega)$, we have that

$$\varphi(x) = \int_\Omega G_x(y)\left(-\Delta\varphi(y) - \gamma\frac{\varphi(y)}{|y|^2}\right)dy \quad \text{for } x \in \Omega,$$

where $G_x := G(x, \cdot)$. Fix $x_0 \in \Omega$ and let $\eta \in C_c^\infty(\Omega)$ be such that $\eta(x) = 1$ around x_0. Define the distribution $\beta_{x_0} : \Omega \to \mathbb{R}$ as

$$G_{x_0}(x) = \frac{1}{\omega_2}\left(\frac{\eta(x)}{|x - x_0|} + \beta_{x_0}(x)\right) \quad \text{for all } x \in \Omega,$$

where $\omega_2 := 4\pi$ is the volume of the canonical 2-sphere. As one checks,

$$\left(-\Delta - \frac{\gamma}{|x|^2}\right)\beta_{x_0} = -\left(-\Delta - \frac{\gamma}{|x|^2}\right)\left(\frac{\eta(x)}{|x - x_0|}\right) := f = O(|x - x_0|^{-1})$$

in the distributional sense. Since $f \in L^2(\Omega)$ and, by uniqueness of the Green's function (since the operator is coercive), we have that $\beta_{x_0} \in D^{1,2}(\Omega)$. It follows from standard elliptic theory that

$$\beta_{x_0} \in C^\infty(\bar\Omega \setminus \{0, x_0\}) \cap C^{0,\theta}(\bar\Omega \setminus B_\delta(0))$$

for all $\theta \in (0, 1)$ and $\delta > 0$. Since f vanishes around 0, it follows from Theorem 4.1 and Lemma 5.2 that

$$\beta_{x_0}(x) = O(|x|^{1-\alpha_-(\gamma)}) \quad \text{and} \quad |\nabla\beta_{x_0}(x)| = O(|x|^{-\alpha_-(\gamma)}) \quad \text{when } x \to 0. \tag{10-1}$$

We can therefore define the *mass of* Ω at x_0 associated to the operator L_γ by $R_\gamma(\Omega, x_0) := \beta_{x_0}(x_0)$. As one checks, $\beta_{x_0}(x_0)$ is independent of the choice of η.

The uniqueness is proved as in Theorem 7.1. The behavior on the boundary is given by Theorem 4.1 and the interior behavior around x_0 is classical. □

Lemma 10.2. *Let $\Omega \subset \mathbb{R}^3$ be an open smooth bounded domain such that $0 \in \partial\Omega$ and $x_0 \in \Omega$. Assume that $\gamma \in (0, \gamma_H(\Omega))$ and that $\mu_{\gamma,0}(\mathbb{R}^3_+) = 1/K(3, 2)^2$. Then, there exists a family $(u_\epsilon)_\epsilon$ in $D^{1,2}(\Omega)$ such that*

$$J^\Omega_{\gamma,0}(u_\epsilon) = \frac{1}{K(n, 2)^2}\left(1 - \frac{\omega_2 R_\gamma(x_0)}{3\int_{\mathbb{R}^3} U^{2^\star}\,dx}\epsilon + o(\epsilon)\right) \quad \text{as } \epsilon \to 0, \tag{10-2}$$

where $U(x) := (1 + |x|^2)^{-1/2}$ for all $x \in \mathbb{R}^3$ and $2^\star = 2^\star(0) = 2n/(n - 2)$.

Proof. The proof is very similar to what was performed by Schoen [1984] (see [Druet 2002a; 2002b; Jaber 2014]). For $\epsilon > 0$, define the functions

$$u_\epsilon(x) := \eta(x)\left(\frac{\epsilon}{\epsilon^2 + |x - x_0|^2}\right)^{1/2} + \epsilon^{1/2}\beta_{x_0}(x) \quad \text{for all } x \in \Omega.$$

As one checks, $u_\epsilon \in D^{1,2}(\Omega)$. Proceeding as in the case $\gamma > \frac{1}{4}(n^2 - 1)$ of Section 8, we get (10-2). We omit the details that are standard. This proves Lemma 10.2. □

We finally get the following.

Theorem 10.3. *Let Ω be a bounded smooth domain of $\mathbb{R}^3$ such that $0 \in \partial\Omega$.*

(1) *If $\gamma \geq \gamma_H(\Omega)$, then there are extremals for $\mu_{\gamma,0}(\Omega)$.*

(2) *If $\gamma \leq 0$, then there are no extremals for $\mu_{\gamma,0}(\Omega)$.*

(3) *If $0 < \gamma < \gamma_H(\Omega)$ and there are extremals for $\mu_{\gamma,0}(\mathbb{R}^n_+)$, then there are extremals for $\mu_{\gamma,0}(\Omega)$ under either one of the following conditions*:
   - *$\gamma \leq \frac{1}{4}(n^2-1)$ and the mean curvature of $\partial\Omega$ at 0 is negative.*
   - *$\gamma > \frac{1}{4}(n^2-1)$ and the mass $m_\gamma(\Omega)$ is positive.*

(4) *If $0 < \gamma < \gamma_H(\Omega)$ and there are no extremals for $\mu_{\gamma,0}(\mathbb{R}^n_+)$, then there are extremals for $\mu_{\gamma,0}(\Omega)$ if there exists $x_0 \in \Omega$ such that $R_\gamma(\Omega, x_0) > 0$.*

Proof. The two first points of the theorem follow from Proposition 8.1 and Theorem 3.6. The third point follows from Proposition 8.3. For the fourth point, in this situation, it follows from Proposition 1.3 that $\mu_{\gamma,0}(\mathbb{R}^n_+) = 1/K(n,2)^2$, and then Lemma 10.2 gives $\mu_{\gamma,0}(\Omega) < \mu_{\gamma,0}(\mathbb{R}^n_+)$, which yields the existence of extremals by Theorem 3.6. This proves Theorem 10.3. □

References

[Attar, Merchán and Peral 2015] A. Attar, S. Merchán, and I. Peral, "A remark on the existence properties of a semilinear heat equation involving a Hardy–Leray potential", *J. Evol. Equ.* **15**:1 (2015), 239–250. MR Zbl

[Aubin 1976] T. Aubin, "Problèmes isopérimétriques et espaces de Sobolev", *J. Differential Geometry* **11**:4 (1976), 573–598. MR Zbl

[Bartsch, Peng and Zhang 2007] T. Bartsch, S. Peng, and Z. Zhang, "Existence and non-existence of solutions to elliptic equations related to the Caffarelli–Kohn–Nirenberg inequalities", *Calc. Var. Partial Differential Equations* **30**:1 (2007), 113–136. MR Zbl

[Caffarelli, Kohn and Nirenberg 1984] L. Caffarelli, R. Kohn, and L. Nirenberg, "First order interpolation inequalities with weights", *Compositio Math.* **53**:3 (1984), 259–275. MR Zbl

[Chern and Lin 2003] J.-L. Chern and C.-S. Lin, "The symmetry of least-energy solutions for semilinear elliptic equations", *J. Differential Equations* **187**:2 (2003), 240–268. MR Zbl

[Chern and Lin 2010] J.-L. Chern and C.-S. Lin, "Minimizers of Caffarelli–Kohn–Nirenberg inequalities with the singularity on the boundary", *Arch. Ration. Mech. Anal.* **197**:2 (2010), 401–432. MR Zbl

[Cowan 2010] C. Cowan, "Optimal Hardy inequalities for general elliptic operators with improvements", *Commun. Pure Appl. Anal.* **9**:1 (2010), 109–140. MR Zbl

[Dávila and Peral 2011] J. Dávila and I. Peral, "Nonlinear elliptic problems with a singular weight on the boundary", *Calc. Var. Partial Differential Equations* **41**:3-4 (2011), 567–586. MR Zbl

[Druet 2002a] O. Druet, "Elliptic equations with critical Sobolev exponents in dimension 3", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **19**:2 (2002), 125–142. MR Zbl

[Druet 2002b] O. Druet, "Optimal Sobolev inequalities and extremal functions: the three-dimensional case", *Indiana Univ. Math. J.* **51**:1 (2002), 69–88. MR Zbl

[Fall 2012] M. M. Fall, "On the Hardy–Poincaré inequality with boundary singularities", *Commun. Contemp. Math.* **14**:3 (2012), art. id. 1250019. MR Zbl

[Fall and Musina 2012] M. M. Fall and R. Musina, "Hardy–Poincaré inequalities with boundary singularities", *Proc. Roy. Soc. Edinburgh Sect. A* **142**:4 (2012), 769–786. MR Zbl

[Ghoussoub and Kang 2004] N. Ghoussoub and X. S. Kang, "Hardy–Sobolev critical elliptic equations with boundary singularities", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **21**:6 (2004), 767–793. MR Zbl

[Ghoussoub and Moradifam 2013] N. Ghoussoub and A. Moradifam, *Functional inequalities: new perspectives and new applications*, Mathematical Surveys and Monographs **187**, American Mathematical Society, Providence, RI, 2013. MR Zbl

[Ghoussoub and Robert 2006a] N. Ghoussoub and F. Robert, "Concentration estimates for Emden–Fowler equations with boundary singularities and critical growth", *IMRP Int. Math. Res. Pap.* (2006), art. id. 21867. MR Zbl

[Ghoussoub and Robert 2006b] N. Ghoussoub and F. Robert, "The effect of curvature on the best constant in the Hardy–Sobolev inequalities", *Geom. Funct. Anal.* **16**:6 (2006), 1201–1245. MR Zbl

[Ghoussoub and Robert 2016] N. Ghoussoub and F. Robert, "Sobolev inequalities for the Hardy–Schrödinger operator: extremals and critical dimensions", *Bull. Math. Sci.* **6**:1 (2016), 89–144. MR Zbl

[Gilbarg and Trudinger 1998] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, revised 2nd ed., Grundlehren der Math. Wissenschaften **224**, Springer, Berlin, 1998. MR Zbl

[Gmira and Véron 1991] A. Gmira and L. Véron, "Boundary singularities of solutions of some nonlinear elliptic equations", *Duke Math. J.* **64**:2 (1991), 271–324. MR Zbl

[Hebey 1997] E. Hebey, *Introduction à l'analyse non linéaire sur les variétés*, Diderot Editeur, Paris, 1997. Zbl

[Jaber 2014] H. Jaber, "Hardy–Sobolev equations on compact Riemannian manifolds", *Nonlinear Anal.* **103** (2014), 39–54. MR Zbl

[Lin and Wadade 2012] C.-S. Lin and H. Wadade, "Minimizing problems for the Hardy–Sobolev type inequality with the singularity on the boundary", *Tohoku Math. J.* (2) **64**:1 (2012), 79–103. MR Zbl

[Pinchover 1994] Y. Pinchover, "On positive Liouville theorems and asymptotic behavior of solutions of Fuchsian type elliptic operators", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **11**:3 (1994), 313–341. MR Zbl

[Pinchover and Tintarev 2005] Y. Pinchover and K. Tintarev, "Existence of minimizers for Schrödinger operators under domain perturbations with application to Hardy's inequality", *Indiana Univ. Math. J.* **54**:4 (2005), 1061–1074. MR Zbl

[Robert 2010] F. Robert, "Existence et asymptotiques optimales des fonctions de Green des opérateurs elliptiques d'ordre deux", unpublished notes, 2010, available at http://www.iecl.univ-lorraine.fr/∼Frederic.Robert/ConstrucGreen.pdf.

[Schoen 1984] R. Schoen, "Conformal deformation of a Riemannian metric to constant scalar curvature", *J. Differential Geom.* **20**:2 (1984), 479–495. MR Zbl

[Schoen and Yau 1988] R. Schoen and S.-T. Yau, "Conformally flat manifolds, Kleinian groups and scalar curvature", *Invent. Math.* **92**:1 (1988), 47–71. MR Zbl

Received 4 Jan 2016. Revised 23 Feb 2017. Accepted 3 Apr 2017.

NASSIF GHOUSSOUB: nassif@math.ubc.ca
Department of Mathematics, 1984 Mathematics Road, University of British Columbia, Vancouver, BC V6T 1Z2, Canada

FRÉDÉRIC ROBERT: frederic.robert@univ-lorraine.fr
Institut Élie Cartan, Université de Lorraine, BP 70239, F-54506 Vandœuvre-lès-Nancy, France

ANALYSIS AND PDE
Vol. 10, No. 5, 2017
dx.doi.org/10.2140/apde.2017.10.1081

CONICAL MAXIMAL REGULARITY FOR ELLIPTIC OPERATORS VIA HARDY SPACES

YI HUANG

We give a technically simple approach to the maximal regularity problem in parabolic tent spaces for second-order, divergence-form, complex-valued elliptic operators. By using the associated Hardy space theory combined with certain L^2-L^2 off-diagonal estimates, we reduce the tent space boundedness in the upper half-space to the reverse Riesz inequalities in the boundary space. This way, we also improve recent results obtained by P. Auscher et al.

1. Introduction

Let $\mathbb{R}^{1+n}_+$ be the upper half-space $\mathbb{R}_+ \times \mathbb{R}^n$ with $\mathbb{R}_+ = (0, \infty)$ and $n \in \mathbb{N}_+ = \{1, 2, \dots\}$. Define the tent space T^p_{par}, $n/(n+1) < p < \infty$, as the space of all locally square-integrable functions on $\mathbb{R}^{1+n}_+$ such that

$$\|F\|_{T^p_{\mathrm{par}}} := \left(\int_{\mathbb{R}^n} \left(\iint_{\mathbb{R}^{1+n}_+} \frac{\mathbf{1}_{B(x,\, t^{1/2})}(y)}{t^{n/2}} |F(t,\, y)|^2 \, dt \, dy \right)^{p/2} dx \right)^{1/p} < \infty. \tag{1}$$

The scale T^p_{par}, $n/(n+1) < p < \infty$, is a parabolic analogue of the tent spaces introduced by R. R. Coifman, Y. Meyer and E. M. Stein [Coifman et al. 1985].

Let $A = A(x)$ be an $n \times n$ matrix of complex L^∞ coefficients, defined on $\mathbb{R}^n$, and satisfying the ellipticity (or "accretivity") condition

$$\lambda |\xi|^2 \le \operatorname{Re} A\xi \cdot \bar{\xi} \quad \text{and} \quad |A\xi \cdot \bar{\zeta}| \le \Lambda |\xi| |\zeta| \tag{2}$$

for $\xi, \zeta \in \mathbb{C}^n$ and for some λ and Λ such that $0 < \lambda \le \Lambda < \infty$. Let

$$L := -\operatorname{div} A \nabla$$

MSC2010: primary 42B37; secondary 47D06, 42B35, 42B20.
Keywords: maximal regularity operators, tent spaces, elliptic operators, Hardy spaces, off-diagonal decay, maximal L^p-regularity.

(its precise definition will be recalled in next section). Consider the associated forward maximal regularity operator $\boldsymbol{M}_L^+$ given by

$$\boldsymbol{M}_L^+(F)_t := \int_0^t Le^{-(t-s)L} F_s \, ds, \tag{3}$$

originally defined on $F \in L^2(\mathbb{R}_+; \boldsymbol{D}(L))$. Here $\boldsymbol{D}(L)$ is the domain of L in $L^2(\mathbb{R}^n)$ and $F_s = F(s, \cdot)$. By a classical result of L. de Simon [1964], $\boldsymbol{M}_L^+$ extends to a bounded operator on $L^2(\mathbb{R}_+; L^2(\mathbb{R}^n))$. By Fubini's theorem,

$$T^2_{\mathrm{par}}(\mathbb{R}^{1+n}_+) \simeq L^2(\mathbb{R}^n; L^2(\mathbb{R}_+)). \tag{4}$$

For p different from 2, the analogous equivalence of (4) between $T^p_{\mathrm{par}}(\mathbb{R}^{1+n}_+)$ and $L^p(\mathbb{R}^n; L^2(\mathbb{R}_+))$ breaks down. We shall refer to the maximal regularity (namely, the boundedness of $\boldsymbol{M}_L^+$) in T^p_{par} as *conical* maximal regularity for the reason that (parabolic) *cones* are involved in defining tent spaces in (1).

The maximal regularity operator $\boldsymbol{M}_L^+$ is a typical example of singular integral operators with operator-valued kernels. Let $1 \le p \le 2$. Let

$$\operatorname{dist}(E, E') := \inf\{|x - y| : x \in E,\ y \in E'\}.$$

We shall say that a class of uniformly $L^2 = L^2(\mathbb{R}^n)$ bounded kernels $\{T(t)\}_{t>0}$ satisfies the L^p-L^2 off-diagonal decay with some order $M \in \mathbb{N}_+$ if we have

$$\|\mathbf{1}_{E'} T(t) \mathbf{1}_E f\|_{L^2} \lesssim t^{-(n/2)(1/p-1/2)} \left(1 + \frac{\operatorname{dist}(E, E')^2}{t}\right)^{-M} \|\mathbf{1}_E f\|_{L^p} \tag{5}$$

for all Borel sets $E, E' \subset \mathbb{R}^n$, all $t > 0$ and all $f \in L^p \cap L^2$. We shall say $\{T(t)\}_{t>0}$ satisfies the L^p-L^2 off-diagonal decay if it satisfies the L^p-L^2 off-diagonal decay with any order $M \in \mathbb{N}_+$. Denote by $p_- = p_-(L)$ the infimum of p for which the heat semigroup $\{e^{-tL}\}_{t>0}$ satisfies the L^p-L^2 off-diagonal decay. Define the index

$$(p_-)_* := \frac{np_-}{n + p_-}. \tag{6}$$

For $L = -\Delta = -\operatorname{div} \nabla$, one has $p_- = 1$ and $1_* = n/(n+1)$.

Our main result in this letter reads as follows.

Theorem 1.1. *Let $L = -\operatorname{div} A\nabla$ with A satisfying* (2) *and p_- defined as in* (6). *Then for $p \in ((p_-)_*, 2]$, the maximal regularity operator $\boldsymbol{M}_L^+$ defined as in* (3) *extends to a bounded operator on T^p_{par}.*

We end the introduction with several remarks.

Remark 1.2. Under the assumption $(p_-)_* < 1$, Theorem 1.1 was first proved by Auscher et al. [2012a, Theorem 3.1] (with $m = 2$, $\beta = 0$ and q close to p_- in their statement). Indeed, we note that $(p_-)_* < 1$ is equivalent to $(p_-)' > n$, where $(p_-)'$ is the dual exponent of p_-. A threshold condition essentially the same as $(p_-)' > n$ is used in [Auscher et al. 2012a].

A general framework of singular integral operators on tent spaces is also presented by Auscher et al. [2012a]. Their method is heavily based on the L^p-L^2 off-diagonal decay of the family $\{tLe^{-tL}\}_{t>0}$

for $p \in (p_-, 2)$. Note that they already improved the previous result in [Auscher et al. 2012b], the T^p_{par}-boundedness of $\boldsymbol{M}_L^+$ for $p \in (2_*, 2]$, which assumes L^2-L^2 off-diagonal decay only.

Here we shall give a technically simple approach to Theorem 1.1 by using the well-established L-associated Hardy space theory combined (mainly) with L^2-L^2 off-diagonal decay of $\{tLe^{-tL}\}_{t>0}$.

Remark 1.3. The motivation of the reduction scheme

$$\text{(operator theory on tent spaces)} \to \text{(Hardy space theory)},$$

which is involved in our proof of Theorem 1.1, comes from the study of conical maximal regularity (in elliptic tent spaces) for first-order perturbed Dirac operators [Huang 2015, Chapter 5]. Furthermore, the motivation of considering such conical (elliptic) maximal regularity estimates is suggested by their applications to boundary-value elliptic problems (see [Auscher and Axelsson 2011] for example). In the parabolic case, the conical maximal regularity results have already proven to be useful in various settings (see for example [Auscher et al. 2014; Auscher and Frey 2015]).

Remark 1.4. Though the singularity of the integral operator $\boldsymbol{M}_L^+$ is at $s = t$, the most involved part turns out to be the estimation of tent space norms when $s \to 0$. For more explanations concerning the "singularity" pertaining to singular integral operators and maximal regularity operators on tent spaces, see [Auscher et al. 2012a, Remark 3.6; Auscher and Frey 2015, Remark 5.23].

Remark 1.5. Theorem 1.1 also extends to higher order elliptic operators. Then one changes correspondingly the homogeneity of tent spaces and off-diagonal decay in (5). We leave this issue to the interested reader.

2. Elliptic operators and Hardy spaces

We give some preliminary materials needed in the proof of Theorem 1.1.

Let A satisfy (2). We define the divergence-form elliptic operator

$$Lf := -\operatorname{div}(A\nabla f),$$

which we interpret in the sense of maximal-accretive operators via a sesquilinear form. That is, $\boldsymbol{D}(L)$ is the largest subspace contained in $W^{1,2}$ for which

$$\left| \int_{\mathbb{R}^n} A\nabla f \cdot \nabla g \right| \le C\|g\|_2$$

for all $g \in W^{1,2}$, and we set Lf by

$$\langle Lf, g\rangle = \int_{\mathbb{R}^n} A\nabla f \cdot \overline{\nabla g}$$

for $f \in \boldsymbol{D}(L)$ and $g \in W^{1,2}$. Thus defined, L is a maximal-accretive operator on L^2 and $\boldsymbol{D}(L)$ is dense in $W^{1,2}$. Furthermore, L has a square root, denoted by $L^{1/2}$ and defined as the unique maximal-accretive operator such that

$$L^{1/2}L^{1/2} = L \tag{7}$$

as unbounded operators [Kato 1976, p. 281].

For L as formulated above, the development of an L-associated Hardy space theory was taken in [Hofmann and Mayboroda 2009] (and independently in [Auscher et al. 2008] in a different geometric setting), in which the authors considered the model case $H^1_L(\mathbb{R}^n)$. In presence of pointwise heat kernel bounds, see [Duong and Yan 2005]. The definition of H^1_L given in [Hofmann and Mayboroda 2009; Auscher et al. 2008] can be extended immediately to $n/(n+1) < p \le 2$ [Hofmann et al. 2011]. To this end, consider the (conical) square function associated with the heat semigroup generated by L

$$\boldsymbol{S}_L(f)(x) := \left(\iint_{\Gamma(x)} |t^2 L e^{-t^2 L} f(y)|^2 \, \frac{dt\, dy}{t^{1+n}} \right)^{1/2}, \quad x \in \mathbb{R}^n,$$

where, as usual,

$$\Gamma(x) = \{(t, y) \in \mathbb{R}^{1+n}_+ : |x - y| < t\}$$

is a nontangential cone with vertex at $x \in \mathbb{R}^n$. As in [Hofmann and Mayboroda 2009; Hofmann et al. 2011], we define $H^p_L(\mathbb{R}^n)$ for $n/(n+1) < p \le 2$ as the completion of

$$\{f \in L^2(\mathbb{R}^n) : \boldsymbol{S}_L(f) \in L^p(\mathbb{R}^n)\}$$

in the quasinorm

$$\|f\|_{H^p_L(\mathbb{R}^n)} := \|\boldsymbol{S}_L(f)\|_{L^p(\mathbb{R}^n)}.$$

We will not get into the dual side ($p > 2$) of the Hardy space theory.

For L^2-L^2 off-diagonal decay related to $\{e^{-sL}, sLe^{-sL}, \sqrt{s}\nabla e^{-sL}\}_{s>0}$, and other holomorphic functions of L (for example $(I - e^{-sL})^\sigma$ with $\sigma > 0$), we refer to Chapter 2 of the memoir [Auscher 2007].

3. Proof of Theorem 1.1

Note that the extension of $\boldsymbol{M}^+_L$ will be divided into two steps: first from $F \in L^2(\mathbb{R}_+; \boldsymbol{D}(L))$ to T^2_{par} and then for $n/(n+1) < p < 2$ from $T^2_{\mathrm{par}} \cap T^p_{\mathrm{par}}$ to T^p_{par}.

First we split the operator $\boldsymbol{M}^+_L$: for $\ell \in \mathbb{N}_+$ large, set

$$\boldsymbol{R}^\ell_L := \boldsymbol{M}^+_L - \boldsymbol{V}^\ell_L, \tag{8}$$

where for $F \in L^2(\mathbb{R}_+; \boldsymbol{D}(L))$ the singular part $\boldsymbol{R}^\ell_L$ is given formally by

$$\boldsymbol{R}^\ell_L(F)_t = \int_0^t Le^{-(t-s)L}(I - e^{-2sL})^\ell F_s \, ds \tag{9}$$

and the regular part is defined by

$$\boldsymbol{V}^\ell_L = \sum_{k=1}^{\ell} \binom{\ell}{k} \boldsymbol{V}_{L,k}$$

with

$$\boldsymbol{V}_{L,k}(F)_t := \int_0^t Le^{-(t+(2k-1)s)L} F_s \, ds, \quad t \in \mathbb{R}_+.$$

For the above binomial sum $\boldsymbol{V}^\ell_L$, it suffices to consider $\boldsymbol{V}_L := \boldsymbol{V}_{L,1}$.

Let $2\mathbb{N}_+ = \{2, 4, \dots\}$. We make the following observation.

Lemma 3.1. *For $\ell \in 2\mathbb{N}_+$ and $\frac{1}{2}\ell > \frac{1}{2} + \frac{1}{4}n$, the operator $\boldsymbol{R}_L^\ell$, as given in* (9) *through* (8), *extends to a bounded operator on $T^p_{\rm par}$ for any $n/(n+1) < p \le 2$.*

Proof. The $T^2_{\rm par}$-boundedness is de Simon's theorem plus the uniform L^2-boundedness of $\{(I-e^{-2sL})^\ell\}_{s>0}$. By interpolation it suffices to consider $n/(n+1) < p \le 1$, and this follows from Lemmata 3.4 and 3.5 of [Auscher et al. 2012a] in the particular case $m = 2$, $\beta = 0$ and $q = 2$.[1] Indeed, first we can decompose the operator $\boldsymbol{R}_L^\ell$ as in [Auscher et al. 2012a] in the way

$$\boldsymbol{R}_L^\ell(F)_t = \int_{t/2}^t Le^{-(t-s)L}(I-e^{-2sL})^\ell F_s\,ds + \int_0^{t/2} Le^{-(t-s)L}(I-e^{-2sL})^\ell F_s\,ds =: \mathrm{I} + \mathrm{II}.$$

Here we view $\mathcal{T}_1 = \{(I-e^{-2sL})^\ell\}_{s>0}$ as an operator on $T^p_{\rm par}$ given by

$$\mathcal{T}_1 : F \mapsto \mathcal{T}_1(F)_s := (I-e^{-2sL})^\ell F_s,$$

with the similar interpretation for $\mathcal{T}_2 = \{(I-e^{-2sL})^\ell/(sL)^{\ell/2}\}_{s>0}$ in

$$Le^{-(t-s)L}(I-e^{-2sL})^\ell = \Big(\frac{s}{t-s}\Big)^{\ell/2} L((t-s)L)^{\ell/2}e^{-(t-s)L}\frac{(I-e^{-2sL})^\ell}{(sL)^{\ell/2}}.$$

Note that $t-s \sim t$ when $s < t/2$. Therefore, to obtain the $T^p_{\rm par}$-boundedness of $\boldsymbol{R}_L^\ell$ for $n/(n+1) < p \le 1$, we can use Lemma 3.4 of [Auscher et al. 2012a] together with the $T^p_{\rm par}$-boundedness of $\mathcal{T}_1$ to estimate I and use Lemma 3.5 of [Auscher et al. 2012a] together with the $T^p_{\rm par}$-boundedness of $\mathcal{T}_2$ to estimate II. The latter tent space boundedness results on $\mathcal{T}_i$, $i = 1, 2$, are implied by their L^2-L^2 off-diagonal decay with order at least $\frac{1}{2}\ell$, which satisfies the condition

$$\frac{\ell}{2} > \frac{1}{2} + \frac{n}{4} = \frac{n}{2}\Big(\frac{1}{n/(n+1)} - \frac{1}{2}\Big).$$

This implication can be easily verified via the extrapolation method on tent spaces through atomic decompositions. Note that we also need the condition $\frac{1}{2}\ell > \frac{1}{2} + \frac{1}{4}n$ in $(s/(t-s))^{\ell/2} \sim (s/t)^{\ell/2}$ when applying Lemma 3.5 of [Auscher et al. 2012a]. □

Next we rewrite the operator $\boldsymbol{V}_L$ in the following way:

$$\boldsymbol{V}_L(F)_t = -\widetilde{\boldsymbol{V}}_L(F)_t + \boldsymbol{I}_L(F)_t, \quad t \in \mathbb{R}_+, \tag{10}$$

where for $F \in L^2(\mathbb{R}_+; \boldsymbol{D}(L))$ the backward part $\widetilde{\boldsymbol{V}}_L$ is defined by

$$\widetilde{\boldsymbol{V}}_L(F)_t := \int_t^\infty Le^{-(t+s)L}F_s\,ds, \quad t \in \mathbb{R}_+, \tag{11}$$

and the trace part $\boldsymbol{I}_L$ is defined by

$$\boldsymbol{I}_L(F)_t := \int_0^\infty Le^{-(t+s)L}F_s\,ds = \sqrt{L}e^{-tL}\int_0^\infty \sqrt{L}e^{-sL}F_s\,ds.$$

We used the square root property $\sqrt{L}\sqrt{L} = L$ recalled in (7).

[1] We point out that one can also prove this lemma by adapting directly the arguments for Lemma 3.4 of [Auscher et al. 2012a] (see [Huang 2015] for details).

Lemma 3.2. *The integral operator $\widetilde{\boldsymbol{V}}_L$ as given in* (11) *extends to a bounded operator on T^p_{par} for any $n/(n+1) < p \le 2$.*[2]

Proof. This is a consequence of a more general claim by Auscher et al. [2012a, Proposition 3.7], again corresponding to the case $m = 2$, $\beta = 0$ and $q = 2$. Indeed, [Auscher et al. 2012a, Proposition 3.7] deals with a counterpart to $\boldsymbol{M}_L^+$, namely the backward maximal regularity operator

$$\boldsymbol{M}_L^-(F)_t := \int_t^\infty L e^{-(s-t)L} F_s \, ds,$$

where $F \in L^2(\mathbb{R}_+; \boldsymbol{D}(L))$, and they use the splitting

$$\boldsymbol{M}_L^-(F)_t = \int_t^{2t} L e^{-(s-t)L} F_s \, ds + \int_{2t}^\infty L e^{-(s-t)L} F_s \, ds =: \mathrm{III} + \mathrm{IV}.$$

We only need to use those arguments in proving [Auscher et al. 2012a, Proposition 3.7] with IV involved since $s - t \sim s$ when $s > 2t$, which is equivalent to $s + t \sim s$ when $s > t$ in our setting. We omit the details. □

Now we use the L-associated Hardy spaces, which we recalled in Section 2, to treat the trace part $\boldsymbol{I}_L$. First, from the conical square function estimates [Hofmann et al. 2011, Proposition 4.9], one has, for $n/(n+1) < p \le 2$,

$$\left\| \sqrt{L} e^{-tL} \int_0^\infty \sqrt{L} e^{-sL} F_s \, ds \right\|_{T^p_{\mathrm{par}}} \lesssim \left\| \int_0^\infty \sqrt{L} e^{-sL} F_s \, ds \right\|_{H^p_L}$$

for $F \in L^2(\mathbb{R}_+; \boldsymbol{D}(L))$. Next, from the reverse Riesz inequalities [Hofmann et al. 2011, Proposition 5.17], one has, for $p \in ((p_-)_*, 2]$,

$$\|\sqrt{L} f\|_{H^p_L} \lesssim \|\nabla f\|_{H^p}$$

for $f \in L^2$; hence, one further has, for $p \in ((p_-)_*, 2]$,

$$\left\| \int_0^\infty \sqrt{L} e^{-sL} F_s \, ds \right\|_{H^p_L} \lesssim \left\| \int_0^\infty \nabla e^{-sL} F_s \, ds \right\|_{H^p}.$$

Here, as usual, we use the convention $H^p = L^p$ for $p > 1$.[3]

For $F \in T^2_{\mathrm{par}}$, consider the sweeping operator

$$\pi_L(F) := \int_0^\infty \nabla e^{-sL} F_s \, ds.$$

An equivalent formulation of the Kato square root estimate for L^* [Auscher et al. 2002] is the square function estimate

$$\iint_{\mathbb{R}^{1+n}_+} |e^{-tL^*} \operatorname{div} \vec{F}(y)|^2 \, dt \, dy \lesssim \|\vec{F}\|_2^2$$

[2] As we will see in the proof, the lemma also holds for any $0 < p \le 2$. But that does not help in proving Theorem 1.1.

[3] We remark that in [Auscher and Frey 2015, Lemma 5.21] a variant of $\boldsymbol{I}_L$ is treated in a similar way, with informative connections to the Hardy space theory associated with the first-order perturbed Dirac operators as alluded to in Remark 1.3.

for all $\vec{F} \in L^2(\mathbb{R}^n; \mathbb{C}^n)$; hence, the mapping given by

$$\mathbb{Q}_{L^*} : \vec{F} \mapsto \mathbb{Q}_{L^*}(\vec{F})(t, y) := (e^{-tL^*} \operatorname{div} \vec{F})(y)$$

is bounded from $L^2(\mathbb{R}^n; \mathbb{C}^n)$ to T^2_{par}. Thereby, we see that $\pi_L : T^2_{\mathrm{par}} \to L^2$ is a bounded operator by duality with $\mathbb{Q}_{L^*}$.

Recall that a T^p_{par}-atom A supported in the parabolic Carleson cylinder

$$\mathrm{Cyl}(B) := (0, r_B^2) \times B$$

for some ball $B \subset \mathbb{R}^n$ (with radius r_B) satisfies the size estimate

$$\|A\|_{T^2_{\mathrm{par}}} \leq |B|^{-(1/p-1/2)}. \tag{12}$$

We have the following result on π_L.

Lemma 3.3. *For any $n/(n+1) < p \leq 1$ and any T^p_{par}-atom A with* $\operatorname{supp} A \subset \mathrm{Cyl}(B)$ *for some ball $B \subset \mathbb{R}^n$ (with radius r_B),*

$$m := \pi_L(A) = \int_0^{r_B^2} \nabla e^{-sL} A_s \, ds$$

satisfies the uniform estimate

$$\|m\|_{H^p} \lesssim 1. \tag{13}$$

Hence, π_L extends to a bounded operator from T^p_{par} to H^p for $n/(n+1) < p \leq 2$.

Proof. For $m = \pi_L(A)$ with A being T^p_{par}-atoms, $n/(n+1) < p \leq 1$, and by adapting [Coifman et al. 1983, Théorème 3; 1985, Theorem 6], (13) follows from the L^2-L^2 off-diagonal decay for the heat semigroup $\{e^{-sL}\}_{s>0}$ and the gradient family $\{\sqrt{s}\nabla e^{-sL}\}_{s>0}$, the size estimate (12) and the Coifman–Weiss molecular theory for H^p. Then for $n/(n+1) < p \leq 1$, π_L extends to a bounded operator from T^p_{par} to H^p, and by interpolation, π_L extends to a bounded operator from T^p_{par} to H^p for $n/(n+1) < p \leq 2$. □

With the splittings (8) and (10), together with the conditions $\ell \in 2\mathbb{N}_+$ and $\frac{1}{2}\ell > \frac{1}{2} + \frac{1}{4}n$, and using Lemmata 3.1, 3.2 and 3.3 in order, the proof of Theorem 1.1 (with $p \in ((p_-)_*, 2]$) is then concluded.

Acknowledgements

This research is supported in part by the Agence Nationale de la Recherche project "Harmonic Analysis at its Boundaries", ANR-12-BS01-0013-01. I am indebted to my Ph.D. advisor Pascal Auscher, who motivated me to work on this subject.

References

[Auscher 2007] P. Auscher, *On necessary and sufficient conditions for L^p-estimates of Riesz transforms associated to elliptic operators on $\mathbb{R}^n$ and related estimates*, Mem. Amer. Math. Soc. **871**, Amer. Math. Soc., Providence, RI, 2007.

[Auscher and Axelsson 2011] P. Auscher and A. Axelsson, "Weighted maximal regularity estimates and solvability of non-smooth elliptic systems, I", *Invent. Math.* **184**:1 (2011), 47–115.

[Auscher and Frey 2015] P. Auscher and D. Frey, "On the well-posedness of parabolic equations of Navier–Stokes type with BMO^{-1} data", *J. Inst. Math. Jussieu* (online publication April 2015).

[Auscher et al. 2002] P. Auscher, S. Hofmann, M. Lacey, A. McIntosh, and Ph. Tchamitchian, "The solution of the Kato square root problem for second order elliptic operators on $\mathbb{R}^n$", *Ann. of Math.* (2) **156**:2 (2002), 633–654.

[Auscher et al. 2008] P. Auscher, A. McIntosh, and E. Russ, "Hardy spaces of differential forms on Riemannian manifolds", *J. Geom. Anal.* **18**:1 (2008), 192–248.

[Auscher et al. 2012a] P. Auscher, C. Kriegler, S. Monniaux, and P. Portal, "Singular integral operators on tent spaces", *J. Evol. Equ.* **12**:4 (2012), 741–765.

[Auscher et al. 2012b] P. Auscher, S. Monniaux, and P. Portal, "The maximal regularity operator on tent spaces", *Commun. Pure Appl. Anal.* **11**:6 (2012), 2213–2219.

[Auscher et al. 2014] P. Auscher, J. van Neerven, and P. Portal, "Conical stochastic maximal L^p-regularity for $1 \leqslant p < \infty$", *Math. Ann.* **359**:3–4 (2014), 863–889.

[Coifman et al. 1983] R. R. Coifman, Y. Meyer, and E. M. Stein, "Un nouvel éspace fonctionnel adapté à l'étude des opérateurs définis par des intégrales singulières", pp. 1–15 in *Harmonic analysis* (Cortona, 1982), edited by F. Ricci and G. Weiss, Lecture Notes in Math. **992**, Springer, Berlin, 1983.

[Coifman et al. 1985] R. R. Coifman, Y. Meyer, and E. M. Stein, "Some new function spaces and their applications to harmonic analysis", *J. Funct. Anal.* **62**:2 (1985), 304–335.

[Duong and Yan 2005] X. T. Duong and L. Yan, "Duality of Hardy and BMO spaces associated with operators with heat kernel bounds", *J. Amer. Math. Soc.* **18**:4 (2005), 943–973.

[Hofmann and Mayboroda 2009] S. Hofmann and S. Mayboroda, "Hardy and BMO spaces associated to divergence form elliptic operators", *Math. Ann.* **344**:1 (2009), 37–116.

[Hofmann et al. 2011] S. Hofmann, S. Mayboroda, and A. McIntosh, "Second order elliptic operators with complex bounded measurable coefficients in L^p, Sobolev and Hardy spaces", *Ann. Sci. Éc. Norm. Supér.* (4) **44**:5 (2011), 723–800.

[Huang 2015] Y. Huang, *Operator theory on tent spaces*, Ph.D. thesis, Université Paris-Sud, Orsay, 2015, Available at https://tel.archives-ouvertes.fr/tel-01350629.

[Kato 1976] T. Kato, *Perturbation theory for linear operators*, 2nd ed., Grundlehren der math. Wissenschaften **132**, Springer, Berlin, 1976.

[de Simon 1964] L. de Simon, "Un'applicazione della teoria degli integrali singolari allo studio delle equazioni differenziali lineari astratte del primo ordine", *Rend. Sem. Mat. Univ. Padova* **34** (1964), 205–223.

Received 14 Apr 2016. Accepted 3 Apr 2017.

YI HUANG: yi.huang@njnu.edu.cn
Laboratoire de Mathématiques d'Orsay, Université Paris-Sud, Centre National de la Recherche Scientifique, Université Paris-Saclay, 91405 Orsay, France

ANALYSIS AND PDE
Vol. 10, No. 5, 2017
dx.doi.org/10.2140/apde.2017.10.1089

LOCAL EXPONENTIAL STABILIZATION FOR A CLASS OF KORTEWEG–DE VRIES EQUATIONS BY MEANS OF TIME-VARYING FEEDBACK LAWS

JEAN-MICHEL CORON, IVONNE RIVAS AND SHENGQUAN XIANG

We study the exponential stabilization problem for a nonlinear Korteweg-de Vries equation on a bounded interval in cases where the linearized control system is not controllable. The system has Dirichlet boundary conditions at the end-points of the interval and a Neumann nonhomogeneous boundary condition at the right end-point, which is the control. We build a class of time-varying feedback laws for which the solutions of the closed-loop systems with small initial data decay exponentially to 0. We present also results on the well-posedness of the closed-loop systems for general time-varying feedback laws.

1. Introduction

Let $L \in (0, +\infty)$. We consider the stabilization of the controlled Korteweg–de Vries (KdV) system

$$\begin{cases} y_t + y_{xxx} + y_x + yy_x = 0 & \text{for } (t, x) \in (s, +\infty) \times (0, L), \\ y(t, 0) = y(t, L) = 0 & \text{for } t \in (s, +\infty), \\ y_x(t, L) = u(t) & \text{for } t \in (s, +\infty), \end{cases} \tag{1-1}$$

where $s \in \mathbb{R}$ and where, at time $t \in [s, +\infty)$, the state is $y(t, \cdot) \in L^2(0, L)$ and the control is $u(t) \in \mathbb{R}$.

Boussinesq [1877] and Korteweg and de Vries [1895] introduced KdV equations for describing the propagation of small-amplitude long water waves. For a better understanding of KdV equations, one can see [Whitham 1974], in which different mathematical models of water waves are deduced. These equations have turned out to be good models, not only for water waves but also to describe other physical phenomena. For mathematical studies on these equations, let us mention [Bona and Smith 1975; Constantin and Saut 1988; Craig et al. 1992; Temam 1969], as well as the discovery of solitons and the inverse scattering method [Gardner et al. 1967; Murray 1978] to solve these equations. We also refer here to [Bona et al. 2003; 2009; Coron and Crépeau 2004; Rivas et al. 2011; Zhang 1999] for well-posedness results of initial-boundary-value problems of our KdV equation (1-1) or for other equations which are similar to (1-1). Finally, let us refer to [Cerpa 2014; Rosier and Zhang 2009] for reviews on recent progresses on the control of various KdV equations.

Coron and Rivas were supported by ERC advanced grant 266907 (CPDENL) of the 7th Research Framework Programme (FP7). Coron and Xiang were supported by ANR Project Finite4SoS (ANR 15-CE23-0007) and by LIASFMA.
MSC2010: 93D15, 93D20, 35Q53.
Keywords: Korteweg–de Vries, time-varying feedback laws, stabilization, controllability.

The controllability research on (1-1) began when Lionel Rosier [1997] showed that the linearized KdV control system (around 0 in $L^2(0, L)$)

$$\begin{cases} y_t + y_{xxx} + y_x = 0 & \text{in } (0, T) \times (0, L), \\ y(t, 0) = y(t, L) = 0 & \text{on } (0, L), \\ y_x(t, L) = u(t) & \text{on } (0, T) \end{cases} \tag{1-2}$$

is controllable if and only if $L \notin \mathcal{N}$, where $\mathcal{N}$ is called the set of critical lengths and is defined by

$$\mathcal{N} := \left\{ 2\pi \sqrt{\tfrac{1}{3}(l^2 + lk + k^2)} : l, k \in \mathbb{N}^* \right\}. \tag{1-3}$$

From this controllability result Lionel Rosier, in the same article, deduced that the nonlinear KdV equations (1-1) are locally controllable (around 0 in $L^2(0, L)$) if $L \notin \mathcal{N}$. His work also shows that the $L^2(0, L)$ space can be decomposed as $H \oplus M$, where M is the "uncontrollable" part for the linearized KdV control systems (1-2), and H is the "controllable" part. Moreover, M is of finite dimension, a dimension which strongly depends on some number theory property of the length L. More precisely, the dimension of M is the number of different pairs of positive integers (l_j, k_j) satisfying

$$L = 2\pi \sqrt{\tfrac{1}{3}(l_j^2 + l_j k_j + k_j^2)}. \tag{1-4}$$

For each such pair of (l_j, k_j) with $l_j \geqslant k_j$, we can find two nonzero real-valued functions φ_1^j and φ_2^j such that $\varphi^j := \varphi_1^j + i\varphi_2^j$ is a solution of

$$\begin{cases} -i\omega(l_j, k_j)\varphi^j + (\varphi^j)' + (\varphi^j)''' = 0, \\ \varphi^j(0) = \varphi^j(L) = 0, \\ (\varphi^j)'(0) = (\varphi^j)'(L) = 0, \end{cases} \tag{1-5}$$

where $\varphi_1^j, \varphi_2^j \in C^\infty([0, L])$ and $\omega(l_j, k_j)$ is defined by

$$\omega(l_j, k_j) := \frac{(2l_j + k_j)(l_j - k_j)(2k_j + l_j)}{3\sqrt{3}(l_j^2 + l_j k_j + k_j^2)^{3/2}}. \tag{1-6}$$

When $l_j > k_j$, the functions φ_1^j, φ_2^j are linearly independent, but when $l_j = k_j$, we have $\omega(l_j, k_j) = 0$ and φ_1^j, φ_2^j are linearly dependent. It is also proved in [Rosier 1997] that

$$M = \operatorname{Span}\{\varphi_1^1, \varphi_2^1, \ldots, \varphi_1^n, \varphi_2^n\}. \tag{1-7}$$

Multiplying (1-2) by φ^j, integrating on $(0, L)$, performing integrations by parts and combining with (1-5), we get

$$\frac{d}{dt}\left(\int_0^L y(t, x)\varphi^j(x)\, dx \right) = i\omega(l_j, k_j) \int_0^L y(t, x)\varphi^j(x)\, dx,$$

which shows that M is included in the "uncontrollable" part of (1-2). Let us point out that there exists at most one pair of (l_j, k_j) such that $l_j = k_j$. Hence we can classify $L \in \mathbb{R}^+$ into five different cases and therefore divide $\mathbb{R}^+$ into five disjoint subsets of $(0, +\infty)$, which are defined as follows:

(1) $\mathcal{C} := \mathbb{R}^+ \setminus \mathcal{N}$. Then $M = \{0\}$.

(2) $\mathcal{N}_1 := \{L \in \mathcal{N} :$ there exists exactly one ordered pair (l_j, k_j) satisfying (1-4) and $l_j = k_j\}$. Then the dimension of M is 1.

(3) $\mathcal{N}_2 := \{L \in \mathcal{N} :$ there exists exactly one ordered pair (l_j, k_j) satisfying (1-4) and $l_j > k_j\}$. Then the dimension of M is 2.

(4) $\mathcal{N}_3 := \{L \in \mathcal{N} :$ there exist $n \geqslant 2$ distinct ordered pairs (l_j, k_j) satisfying (1-4) and none satisfy $l_j = k_j\}$. Then the dimension of M is $2n$.

(5) $\mathcal{N}_4 := \{L \in \mathcal{N} :$ there exist $n \geqslant 2$ distinct ordered pairs (l_j, k_j) satisfying (1-4) and one satisfies $l_j = k_j\}$. Then the dimension of M is $2n - 1$.

The five sets $\mathcal{C}$, $\{\mathcal{N}_i\}_{i=1}^4$ are pairwise disjoint and

$$\mathbb{R}^+ = \mathcal{C} \cup \mathcal{N}_1 \cup \mathcal{N}_2 \cup \mathcal{N}_3 \cup \mathcal{N}_4,$$
$$\mathcal{N} = \mathcal{N}_1 \cup \mathcal{N}_2 \cup \mathcal{N}_3 \cup \mathcal{N}_4.$$

Additionally, Eduardo Cerpa [2007, Lemma 2.5] proved that each of these five sets has infinite number of elements; see also [Coron 2007, Proposition 8.3] for the case of $\mathcal{N}_1$.

Let us point out that $L \notin \mathcal{N}$ is equivalent to $M = \{0\}$. Hence, Lionel Rosier solved the (local) controllability problem of nonlinear KdV equations for $L \in \mathcal{C}$. Later on Jean-Michel Coron and Emmanuelle Crépeau [2004] proved the small-time local controllability of nonlinear KdV equations for the second case $L \in \mathcal{N}_1$, by a "power series expansion" method; the nonlinear term yy_x gives this controllability. Later on, Eduardo Cerpa [2007] proved the local controllability in large time for the third case $L \in \mathcal{N}_2$, still by using the "power series expansion" method. In this case, an expansion to the order 2 is sufficient but the local controllability in small time remains open. Finally Eduardo Cerpa and Emmanuelle Crépeau [2009a] concluded the study by proving the local controllability in large time of (1-1) for the two remaining critical cases (for which $\dim M \geqslant 3$). The proofs of all these results rely on the "power series expansion" method, introduced in [Coron and Crépeau 2004]. This method has also been used to prove controllability results for Schrödinger equations [Beauchard 2005; Beauchard and Coron 2006; Beauchard and Morancey 2014; Morancey 2014] and for rapid asymptotic stability of a Navier-Stokes control system in [Chowdhury and Ervedoza 2017]. In this article we use it to get exponential stabilization of (1-1). For studies on the controllability of other KdV control systems problems, let us refer to [Capistrano-Filho et al. 2015; Gagnon 2016; Glass and Guerrero 2010; Goubet and Shen 2007; Rosier 2004; Zhang 1999].

The asymptotic stability of 0 without control (control term equal to 0) has been studied for years; see, in particular, [Cerpa and Coron 2013; Goubet and Shen 2007; Jia and Zhang 2012; Massarolo et al. 2007; Pazoto 2005; Perla Menzala et al. 2002; Rosier and Zhang 2006; Russell and Zhang 1995; 1996]. For example, the local exponential stability for our KdV equation if $L \notin \mathcal{N}$ was proved in [Perla Menzala et al. 2002]. Let also point out here that in [Doronin and Natali 2014], the authors give the existence of (large) stationary solutions, which ensures that the exponential stability result in [Perla Menzala et al. 2002] is only local.

Concerning the stabilization by means of feedback laws, the locally exponential stabilization with arbitrary decay rate (rapid stabilization) with some linear feedback law was obtained by Eduardo Cerpa

and Emmanuelle Crépeau in [2009b] for the linear KdV equation (1-2). For the nonlinear case, the first rapid stabilization for Korteweg–de Vries equations was obtained by Camille Laurent, Lionel Rosier and Bing-Yu Zhang [Laurent et al. 2010] in the case of localized distributed control on a periodic domain. In that case, the linearized control system, let us write it $\dot{y} = Ay + Bu$, is controllable. These authors used an approach due to Marshall Slemrod [1974] to construct linear feedback laws leading to the rapid stabilization of $\dot{y} = Ay + Bu$ and then proved that the same feedback laws give the rapid stabilization of the nonlinear Korteweg de Vries equation. In the case of distributed control, the operator B is bounded. For boundary control the operator B is unbounded. The Slemrod approach has been modified to handle this case by Vilmos Komornik [1997] and by Jose Urquiza [2005], and [Cerpa and Crépeau 2009b] precisely uses the modification presented in [Urquiza 2005]. However, in contrast with the case of distributed control, it leads to unbounded linear feedback laws and one does not know for the moment if these linear feedback laws lead to asymptotic stabilization for the nonlinear Korteweg de Vries equation. One does not even know if the closed system is well posed for this nonlinear equation. The first rapid stabilization result in the nonlinear case and with boundary controls was obtained by Eduardo Cerpa and Jean-Michel Coron [2013]. Their approach relies on the backstepping method/transformation, a method introduced by Miroslav Krstic and his collaborators (see [Krstic and Smyshlyaev 2008] for an excellent starting point to this method). When $L \notin \mathcal{N}$, by using a more general transformation and the controllability of (1-2), Jean-Michel Coron and Qi Lü [2014] proved the rapid stabilization of our KdV control system. Their method can be applied to many other equations, like Schrödinger equations [Coron et al. 2016] and Kuramoto–Sivashinsky equations [Coron and Lü 2015]. When $L \in \mathcal{N}$, as mentioned above, the linearized control system (1-2) is not controllable, but the control system (1-1) is controllable. Let us recall that for the finite-dimensional case, the controllability doesn't imply the existence of a (continuous) stationary feedback law which stabilizes (asymptotically, exponentially, etc.) the control system; see [Brockett 1983; Coron 1990]. However the controllability in general implies the existence of (continuous) *time-varying* feedback laws which asymptotically (and even in finite time) stabilize the control system; see [Coron 1995]. Hence it is natural to look for time-varying feedback laws $u(t, y(t, \cdot))$ such that 0 is (locally) asymptotically stable for the closed-loop system

$$\begin{cases} y_t + y_{xxx} + y_x + yy_x = 0 & \text{for } (t, x) \in (s, +\infty) \times (0, L), \\ y(t, 0) = y(t, L) = 0 & \text{for } t \in (s, +\infty), \\ y_x(t, L) = u(t, y(t, \cdot)) & \text{for } t \in (s, +\infty). \end{cases} \tag{1-8}$$

Let us also point out that in [Laurent et al. 2010], as in [Coron and Rosier 1994] by Jean-Michel Coron and Lionel Rosier, which dealt with finite-dimensional control systems, time-varying feedback laws were used in order to combine two different feedback laws to get rapid *global* asymptotic stability of the closed loop system. Let us emphasize that $u = 0$ leads to (local) asymptotic stability when $L \in \mathcal{N}_1$ [Chu et al. 2015] and $L \in \mathcal{N}_2$ [Tang et al. 2016]. However, in both cases, the convergence is not exponential. It is then natural to ask if we can get exponential convergence to 0 with the help of some suitable time-varying feedback laws $u(t, y(t, \cdot))$. The aim of this paper is to prove that it is indeed possible in the case where

$$L \text{ is in } \mathcal{N}_2 \text{ or in } \mathcal{N}_3. \tag{1-9}$$

Let us denote by

$$P_H : L^2(0, L) \to H \quad \text{and} \quad P_M : L^2(0, L) \to M$$

the orthogonal projections (for the L^2-scalar product) on H and M respectively. Our main result is the following one, where the precise definition of a solution of (1-10) is given in Section 2.

Theorem 1. *Assume that* (1-9) *holds. Then there exists a periodic time-varying feedback law* u, $C > 0$, $\lambda > 0$ *and* $r > 0$ *such that, for every* $s \in \mathbb{R}$ *and for every* $\|y_0\|_{L^2_L} < r$, *the Cauchy problem*

$$\begin{cases} y_t + y_{xxx} + y_x + yy_x = 0 & \text{for } (t, x) \in (s, +\infty) \times (0, L), \\ y(t, 0) = y(t, L) = 0 & \text{for } t \in (s, +\infty), \\ y_x(t, L) = u(t, y(t, \cdot)) & \text{for } t \in (s, +\infty), \\ y(s, \cdot) = y_0 & \text{for } x \in (0, L) \end{cases} \tag{1-10}$$

has at least one solution in $C^0\big([s, +\infty); L^2(0, L)\big) \cap L^2_{\text{loc}}\big([s, +\infty); H^1(0, L)\big)$ *and every solution* y *of* (1-10) *is defined on* $[s, +\infty)$ *and satisfies, for every* $t \in [s, +\infty)$,

$$\|P_H(y(t))\|_{L^2_L} + \|P_M(y(t))\|^{1/2}_{L^2_L} \leqslant Ce^{-\lambda(t-s)}\big(\|P_H(y_0)\|_{L^2_L} + \|P_M(y_0)\|^{1/2}_{L^2_L}\big). \tag{1-11}$$

In order to simplify the notations, in this paper we sometimes simply denote $y(t, \cdot)$ by $y(t)$, if there is no misunderstanding; sometimes we also simply denote $L^2(0, L)$ by L^2_L and $L^2(0, T)$ by L^2_T. Let us explain briefly an important ingredient of our proof of Theorem 1. Taking into account the uncontrollability of the linearized system, it is natural to split the KdV system into a coupled system for $(P_H(y), P_M(y))$. Then the finite-dimensional analogue of our KdV control system is

$$\dot{x} = Ax + R_1(x, y) + Bu, \quad \dot{y} = Ly + Q(x, x) + R_2(x, y), \tag{1-12}$$

where A, B, and L are matrices, Q is a quadratic map, R_1, R_2 are polynomials and u is the control. The state variable x plays the role of $P_H(y)$, while y plays the role of $P_M(y)$. The two polynomials R_1 and R_2 are quadratic and $R_2(x, y)$ vanishes for $y = 0$. For this ODE system, in many cases the Brockett condition [1983] and the Coron condition [2007] for the existence of continuous stationary stabilizing feedback laws do not hold. However, as shown in [Coron and Rivas 2016], many physical systems of form (1-12) can be exponentially stabilized by means of time-varying feedback laws. We follow the construction of these time-varying feedback laws given in this article. However, due to the fact that H is of infinite dimension, many parts of the proof have to be modified compared to those given in [Coron and Rivas 2016]; in particular we do not know how to use a Lyapunov approach, in contrast to what is done in that paper.

This article is organized as follows. In Section 2, we recall some classical results and definitions about (1-1) and (1-2). In Section 3, we study the existence and uniqueness of solutions to the closed-loop system (1-10) with time-varying feedback laws u which are not smooth. In Section 4, we construct our time-varying feedback laws. In Section 5, we prove two estimates for solutions to the closed-loop system (1-10) (Propositions 15 and 16) which imply Theorem 1. The article ends with three appendices where proofs of propositions used in the main parts of the article are given.

2. Preliminaries

We first recall some results on KdV equations and give the definition of a solution to the Cauchy problem (1-10). Let us start with the nonhomogeneous linear Cauchy problem

$$\begin{cases} y_t + y_{xxx} + y_x = \tilde{h} & \text{in } (T_1, T_2) \times (0, L), \\ y(t, 0) = y(t, L) = 0 & \text{on } (T_1, T_2), \\ y_x(t, L) = h(t) & \text{on } (T_1, T_2), \\ y(T_1, x) = y_0(x) & \text{on } (0, L) \end{cases} \tag{2-1}$$

for

$$-\infty < T_1 < T_2 < +\infty, \tag{2-2}$$

$$y_0 \in L^2(0, L), \tag{2-3}$$

$$\tilde{h} \in L^1(T_1, T_2; L^2(0, L)), \tag{2-4}$$

$$h \in L^2(T_1, T_2). \tag{2-5}$$

Let us now give the definition of a solution to (2-1).

Definition 2. A solution to the Cauchy problem (2-1) is a function $y \in L^1(T_1, T_2; L^2(0, L))$ such that, for almost every $\tau \in [T_1, T_2]$, the following holds: for every $\phi \in C^3([T_1, \tau] \times [0, L])$ such that

$$\phi(t, 0) = \phi(t, L) = \phi_x(t, 0) = 0 \quad \forall t \in [T_1, \tau], \tag{2-6}$$

one has

$$-\int_{T_1}^{\tau}\int_0^L (\phi_t + \phi_x + \phi_{xxx}) y \, dx \, dt - \int_{T_1}^{\tau} h(t)\phi_x(t, L) \, dt - \int_{T_1}^{\tau}\int_0^L \phi\tilde{h} \, dx \, dt + \int_0^L y(\tau, x)\phi(\tau, x) \, dx - \int_0^L y_0\phi(T_1, x) \, dx = 0. \tag{2-7}$$

For T_1 and T_2 satisfying (2-2), let us define the linear space $\mathcal{B}_{T_1,T_2}$ by

$$\mathcal{B}_{T_1,T_2} := C^0\big([T_1, T_2]; L^2(0, L)\big) \cap L^2\big(T_1, T_2; H^1(0, L)\big). \tag{2-8}$$

This linear space $\mathcal{B}_{T_1,T_2}$ is equipped with the norm

$$\|y\|_{\mathcal{B}_{T_1,T_2}} := \max\big\{\|y(t)\|_{L^2_L} : t \in [T_1, T_2]\big\} + \left(\int_{T_1}^{T_2} \|y_x(t)\|^2_{L^2_L} \, dt\right)^{1/2}. \tag{2-9}$$

With this norm, $\mathcal{B}_{T_1,T_2}$ is a Banach space.

Let $\mathcal{A} : \mathcal{D}(\mathcal{A}) \subset L^2(0, L) \to L^2(0, L)$ be the linear operator defined by

$$\mathcal{D}(\mathcal{A}) := \big\{\phi \in H^3(0, L) : \phi(0) = \phi(L) = \phi_x(L) = 0\big\}, \tag{2-10}$$

$$\mathcal{A}\phi := -\phi_x - \phi_{xxx} \quad \forall \phi \in \mathcal{D}(\mathcal{A}). \tag{2-11}$$

It is known that both $\mathcal{A}$ and $\mathcal{A}^*$ are closed and dissipative (see, e.g., [Coron 2007, page 39]), and therefore $\mathcal{A}$ generates a strongly continuous semigroup of contractions $S(t)$, $t \in [0, +\infty)$ on $L^2(0, L)$.

Rosier [1997], using the above properties of $\mathcal{A}$ together with multiplier techniques, proved the following existence and uniqueness result for the Cauchy problem (2-1).

Lemma 3. *The Cauchy problem* (2-1) *has one and only one solution. This solution is in* $\mathcal{B}_{T_1,T_2}$ *and there exists a constant* $C_2 > 0$ *depending only on* $T_2 - T_1$ *such that*

$$\|y\|_{\mathcal{B}_{T_1,T_2}} \leqslant C_2\big(\|y_0\|_{L^2_L} + \|h\|_{L^2(T_1,T_2)} + \|\tilde{h}\|_{L^1(T_1,T_2;L^2(0,L))}\big). \tag{2-12}$$

In fact the notion of solution to the Cauchy problem (2-1) considered in [Rosier 1997] is a priori stronger than the one we consider here (it is required to be in $C^0([T_1, T_2]; L^2(0, L))$. However, the uniqueness of the solution in the sense of Definition 2 still follows from classical arguments; see, for example, [Coron 2007, Proof of Theorem 2.37, page 53].

Let us now turn to the nonlinear KdV equation

$$\begin{cases} y_t + y_{xxx} + y_x + yy_x = \widetilde{H} & \text{in } (T_1, T_2) \times (0, L), \\ y(t, 0) = y(t, L) = 0 & \text{on } (T_1, T_2), \\ y_x(t, L) = H(t) & \text{on } (T_1, T_2), \\ y(T_1, x) = y_0(x) & \text{on } (0, L). \end{cases} \tag{2-13}$$

Inspired by Lemma 3, we adopt the following definition.

Definition 4. A solution to (2-13) is a function $y \in \mathcal{B}_{T_1,T_2}$ which is a solution of (2-1) for $\tilde{h} := \widetilde{H} - yy_x \in L^1(T_1, T_2; L^2(0, L))$ and $h := H$.

Throughout this article we will use similar definitions without giving them precisely, as, for example, in the case for system (3-15).

Coron and Crépeau [2004] proved the following lemma on the well-posedness of the Cauchy problem (2-13) for small initial data.

Lemma 5. *There exist* $\eta > 0$ *and* $C_3 > 0$ *depending on* L *and* $T_2 - T_1$ *such that, for every* $y_0 \in L^2(0, L)$, *every* $H \in L^2(T_1, T_2)$ *and every* $\widetilde{H} \in L^1(T_1, T_2; L^2(0, L))$ *satisfying*

$$\|y_0\|_{L^2_L} + \|H\|_{L^2(T_1,T_2)} + \|\widetilde{H}\|_{L^1(T_1,T_2;L^2(0,L))} \leqslant \eta, \tag{2-14}$$

the Cauchy problem (2-13) *has a unique solution and this solution satisfies*

$$\|y\|_{\mathcal{B}_{T_1,T_2}} \leqslant C_3\big(\|y_0\|_{L^2_L} + \|H\|_{L^2(T_1,T_2)} + \|\widetilde{H}\|_{L^1(T_1,T_2;L^2(0,L))}\big). \tag{2-15}$$

3. Time-varying feedback laws and well-posedness of the associated closed-loop system

Throughout this section u denotes a time-varying feedback law; it is a map from $\mathbb{R} \times L^2(0, L)$ with values into $\mathbb{R}$. We assume that this map is a Carathéodory map, i.e., it satisfies the three properties

$$\forall R > 0,\ \exists\, C_B(R) > 0 \text{ such that } \big(\|y\|_{L^2_L} \leqslant R \implies |u(t, y)| \leqslant C_B(R) \quad \forall t \in \mathbb{R}\big), \tag{3-1}$$

$$\forall y \in L^2(0, L), \text{ the function } t \in \mathbb{R} \mapsto u(t, y) \in \mathbb{R} \text{ is measurable}, \tag{3-2}$$

$$\text{for almost every } t \in \mathbb{R}, \text{ the function } y \in L^2(0, L) \mapsto u(t, y) \in \mathbb{R} \text{ is continuous}. \tag{3-3}$$

In this article we always assume that

$$C_B(R) \geq 1 \quad \forall R \in [0, +\infty), \tag{3-4}$$

$$R \in [0, +\infty) \mapsto C_B(R) \in \mathbb{R} \text{ is a nondecreasing function.} \tag{3-5}$$

Let $s \in \mathbb{R}$ and let $y_0 \in L^2(0, L)$. We start by giving the definition of a solution to

$$\begin{cases} y_t + y_{xxx} + y_x + yy_x = 0 & \text{for } t \in \mathbb{R},\ x \in (0, L), \\ y(t, 0) = y(t, L) = 0 & \text{for } t \in \mathbb{R}, \\ y_x(t, L) = u(t, y(t, \cdot)) & \text{for } t \in \mathbb{R}, \end{cases} \tag{3-6}$$

and to the Cauchy problem

$$\begin{cases} y_t + y_{xxx} + y_x + yy_x = 0 & \text{for } t > s,\ x \in (0, L), \\ y(t, 0) = y(t, L) = 0 & \text{for } t > s, \\ y_x(t, L) = u(t, y(t, \cdot)) & \text{for } t > s, \\ y(s, x) = y_0(x) & \text{for } x \in (0, L), \end{cases} \tag{3-7}$$

where y_0 is a given function in $L^2(0, L)$ and s is a given real number.

Definition 6. Let I be an interval of $\mathbb{R}$ with a nonempty interior. A function y is a solution of (3-6) on I if $y \in C^0(I; L^2(0, L))$ is such that, for every $[T_1, T_2] \subset I$ with $-\infty < T_1 < T_2 < +\infty$, the restriction of y to $[T_1, T_2] \times (0, L)$ is a solution of (2-13) with $\widetilde{H} := 0$, $H(t) := u(t, y(t))$ and $y_0 := y(T_1)$. A function y is a solution to the Cauchy problem (3-7) if there exists an interval I with a nonempty interior satisfying $I \cap (-\infty, s] = \{s\}$ such that $y \in C^0(I; L^2(0, L))$ is a solution of (3-6) on I and satisfies the initial condition $y(s) = y_0$ in $L^2(0, L)$. The interval I is denoted by $D(y)$. We say that a solution y to the Cauchy problem (3-7) is maximal if, for every solution z to the Cauchy problem (3-7) such that

$$D(y) \subset D(z), \tag{3-8}$$

$$y(t) = z(t) \quad \text{for every } t \text{ in } D(y), \tag{3-9}$$

one has

$$D(y) = D(z). \tag{3-10}$$

Let us now state our theorems concerning the Cauchy problem (3-7).

Theorem 7. *Assume that u is a Carathéodory function and that, for every $R > 0$, there exists $K(R) > 0$ such that*

$$\big(\|y\|_{L^2_L} \leqslant R \quad \text{and} \quad \|z\|_{L^2_L} \leqslant R\big) \implies \big(|u(t, y) - u(t, z)| \leqslant K(R)\|y - z\|_{L^2_L} \quad \forall t \in \mathbb{R}\big). \tag{3-11}$$

Then, for every $s \in \mathbb{R}$ and for every $y_0 \in L^2(0, L)$, the Cauchy problem (3-7) *has one and only one maximal solution y. If $D(y)$ is not equal to $[s, +\infty)$, there exists $\tau \in \mathbb{R}$ such that $D(y) = [s, \tau)$ and one has*

$$\lim_{t \to \tau^-} \|y(t)\|_{L^2_L} = +\infty. \tag{3-12}$$

Moreover, if $C_B(R)$ *satisfies*

$$\int_0^{+\infty} \frac{R}{(C_B(R))^2}\, dR = +\infty, \tag{3-13}$$

then

$$D(y) = [s, +\infty). \tag{3-14}$$

Theorem 8. *Assume that* u *is a Carathéodory function which satisfies condition* (3-13)*. Then, for every* $s \in \mathbb{R}$ *and for every* $y_0 \in L^2(0, L)$*, the Cauchy problem* (3-7) *has at least one maximal solution* y *such that* $D(y) = [s, +\infty)$*.*

The proofs of Theorems 7 and 8 will be given in Appendix B.

We end this section with the following proposition, which gives the expected connection between the evolution of $P_M(y)$ and $P_H(y)$ and the fact that y is a solution to (3-6).

Proposition 9. *Let* $u : \mathbb{R} \times L^2(0, L) \to \mathbb{R}$ *be a Carathéodory feedback law. Let* $-\infty < s < T < +\infty$*, let* $y \in \mathcal{B}_{s,T}$ *and let* $y_0 \in L^2(0, L)$*. Denote* $P_H(y)$ *by* y_1 *and* $P_M(y)$ *by* y_2*. Then* y *is a solution to the Cauchy problem* (3-7) *if and only if*

$$\begin{cases} y_{1t} + y_{1x} + y_{1xxx} + P_H\big((y_1 + y_2)(y_1 + y_2)_x\big) = 0, \\ y_1(t, 0) = y_1(t, L) = 0, \\ y_{1x}(t, L) = u(t, y_1 + y_2), \\ y_1(0, \cdot) = P_H(y_0), \\ y_{2t} + y_{2x} + y_{2xxx} + P_M\big((y_1 + y_2)(y_1 + y_2)_x\big) = 0, \\ y_2(t, 0) = y_2(t, L) = 0, \\ y_{2x}(t, L) = 0, \\ y_2(0, \cdot) = P_M(y_0). \end{cases} \tag{3-15}$$

The proof of this proposition is given in Appendix A.

4. Construction of time-varying feedback laws

In this section, we construct feedback laws which will lead to the local exponential stability stated in Theorem 1. Let us denote by $M_{\mathbf{1}}$ the set of elements in M having an L^2-norm equal to 1:

$$M_{\mathbf{1}} := \big\{ y \in M : \|y\|_{L^2_L} = 1 \big\}. \tag{4-1}$$

Let M^j be the linear space generated by φ_1^j and φ_2^j for every $j \in \{1, 2, \ldots, n\}$:

$$M^j := \mathrm{Span}\{\varphi_1^j, \varphi_2^j\}. \tag{4-2}$$

The construction of our feedback laws relies on the following proposition.

Proposition 10. *There exist* $T > 0$ *and* $v \in L^\infty([0, T] \times M_{\mathbf{1}}; \mathbb{R})$ *such that the following properties hold*:

($\mathcal{P}_1$) *There exists* $\rho_1 \in (0, 1)$ *such that*

$$\|S(T)y_0\|^2_{L^2(0,L)} \leqslant \rho_1 \|y_0\|^2_{L^2(0,L)} \quad \textit{for every } y_0 \in H.$$

($\mathcal{P}_2$) *For every* $y_0 \in M$,

$$\|S(T)y_0\|^2_{L^2(0,L)} = \|y_0\|^2_{L^2(0,L)}.$$

($\mathcal{P}_3$) *There exists* $C_0 > 0$ *such that*

$$|\, v(t, y) - v(t, z)\,| \leqslant C_0 \|y - z\|_{L^2(0,L)} \quad \forall t \in [0, T],\ \forall y, z \in \boldsymbol{M_1}. \tag{4-3}$$

Moreover, there exists $\delta > 0$ *such that, for every* $z \in \boldsymbol{M_1}$, *the solution* (y_1, y_2) *to the equation*

$$\begin{cases} y_{1t} + y_{1x} + y_{1xxx} = 0, \\ y_1(t, 0) = y_1(t, L) = 0, \\ y_{1x}(t, L) = v(t, z), \\ y_1(0, x) = 0, \\ y_{2t} + y_{2x} + y_{2xxx} + P_M(y_1 y_{1x}) = 0, \\ y_2(t, 0) = y_2(t, L) = 0, \\ y_{2x}(t, L) = 0, \\ y_2(0, x) = 0, \end{cases} \tag{4-4}$$

satisfies

$$y_1(T) = 0 \quad \textit{and} \quad \langle y_2(T), S(T)z\rangle_{L^2(0,L)} < -2\delta. \tag{4-5}$$

Proof of Proposition 10. Property ($\mathcal{P}_2$) is given in [Rosier 1997]; one can also see (4-14) and (4-44). Property ($\mathcal{P}_1$) follows from the dissipativity of $\mathcal{A}$ and the controllability of (1-2) in H (see also [Perla Menzala et al. 2002]). Indeed, integrations by parts (and simple density arguments) show that, in the distribution sense in $(0, +\infty)$,

$$\frac{d}{dt}\|S(t)y_0\|^2_{L^2_L} = -y_x^2(t, 0). \tag{4-6}$$

Moreover, as Rosier [1997] proved for every $T > 0$, there exists $c > 1$ such that, for every $y_0 \in H$,

$$\|y_0\|^2_{L^2_L} \leqslant c\|y_x(t, 0)\|^2_{L^2(0,T)}. \tag{4-7}$$

Integration of identity (4-6) on $(0, T)$ and the use of (4-7) give

$$\|S(T)y_0\|^2_{L^2_L} \leqslant \frac{c-1}{c}\|y_0\|^2_{L^2_L}. \tag{4-8}$$

Hence $\rho_1 := (c - 1)/c \in (0, 1)$ satisfies the required properties.

Our concern now is to deal with ($\mathcal{P}_3$). Let us first recall a result on the controllability of the linear control system

$$\begin{cases} y_t + y_{xxx} + y_x = 0 & \text{in } (0, T) \times (0, L), \\ y(t, 0) = y(t, L) = 0 & \text{on } (0, L), \\ y_x(t, L) = u(t) & \text{on } (0, T), \end{cases} \tag{4-9}$$

where, at time $t \in [0, T]$, the state is $y(t, \cdot) \in L^2(0, L)$. Our goal is to investigate the cases where $L \in \mathcal{N}_2 \cup \mathcal{N}_3$, but in order to explain more clearly our construction of v, we first deal with the case where

$$L = 2\pi\sqrt{\tfrac{1}{3}(1^2 + 1 \times 2 + 2^2)} = 2\pi\sqrt{\tfrac{7}{3}}, \tag{4-10}$$

which corresponds to $l = 1$ and $k = 2$ in (1-3). In that case the uncontrollable subspace M is a two-dimensional vector subspace of $L^2(0, L)$ generated by

$$\varphi_1(x) = C\big(\cos\big(\tfrac{5}{\sqrt{21}}x\big) - 3\cos\big(\tfrac{1}{\sqrt{21}}x\big) + 2\cos\big(\tfrac{4}{\sqrt{21}}x\big)\big),$$
$$\varphi_2(x) = C\big(-\sin\big(\tfrac{5}{\sqrt{21}}x\big) - 3\sin\big(\tfrac{1}{\sqrt{21}}x\big) + 2\sin\big(\tfrac{4}{\sqrt{21}}x\big)\big),$$

where C is a positive constant such that $\|\varphi_1\|_{L^2_L} = \|\varphi_2\|_{L^2_L} = 1$. They satisfy

$$\begin{cases} \varphi_1' + \varphi_1''' = -2\pi\varphi_2/p, \\ \varphi_1(0) = \varphi_1(L) = 0, \\ \varphi_1'(0) = \varphi_1'(L) = 0 \end{cases} \tag{4-11}$$

and

$$\begin{cases} \varphi_2' + \varphi_2''' = 2\pi\varphi_1/p, \\ \varphi_2(0) = \varphi_2(L) = 0, \\ \varphi_2'(0) = \varphi_2'(L) = 0, \end{cases} \tag{4-12}$$

with (see [Cerpa 2007])

$$p := \frac{441\pi}{10\sqrt{21}}. \tag{4-13}$$

For every $t > 0$, one has

$$S(t)M \subset M \quad \text{and} \quad S(t) \text{ restricted to } M \text{ is the rotation of angle } \frac{2\pi t}{p}, \tag{4-14}$$

if the orientation on M is chosen so that (φ_1, φ_2) is a direct basis, a choice which is done from now on. Moreover the control u has no action on M for the linear control system (1-2): for every initial data $y_0 \in M$, whatever $u \in L^2(0, T)$, the solution y of (1-2) with $y(0) = y_0$ satisfies $P_M(y(t)) = S(t)y_0$ for every $t \in [0, +\infty)$. Let us denote by H the orthogonal in $L^2(0, L)$ of M for the L^2-scalar product $H := M^\perp$. This linear space is left invariant by the linear control system (1-2): for every initial data $y_0 \in H$, whatever $u \in L^2(0, T)$, the solution y of (1-2) satisfying $y(0) = y_0$ is such that $y(t) \in H$ for every $t \in [0, +\infty)$. Moreover, as proved by Rosier [1997], the linear control system (1-2) is controllable in H in small time. More precisely, he proved the following lemma.

Lemma 11. *Let $T > 0$. There exists $C > 0$ depending only on T such that, for every y_0, $y_1 \in H$, there exists a control $u \in L^2(0, T)$ satisfying*

$$\|u\|_{L^2_T} \leqslant C(\|y_0\|_{L^2_L} + \|y_1\|_{L^2_L}) \tag{4-15}$$

such that the solution y of the Cauchy problem

$$\begin{cases} y_t + y_{xxx} + y_x = 0 & \text{in } (0, T) \times (0, L), \\ y(t, 0) = y(t, L) = 0 & \text{on } (0, T), \\ y_x(t, L) = u(t) & \text{on } (0, T), \\ y(0, x) = y_0(x) & \text{on } (0, L) \end{cases}$$

satisfies $y(T, \cdot) = y_1$.

A key ingredient of our construction of v is the following proposition.

Proposition 12. *Let $T > 0$. For every $L \in \mathcal{N}_2 \cup \mathcal{N}_3$, for every $j \in \{1, 2, \dots, n\}$, there exists $u^j \in H^1(0, T)$ such that*

$$\alpha(T, \cdot) = 0 \quad \text{and} \quad P_{M^j}(\beta(T, \cdot)) \neq 0,$$

where (α, β) is the solution of

$$\begin{cases} \alpha_t + \alpha_x + \alpha_{xxx} = 0, \\ \alpha(t, 0) = \alpha(t, L) = 0, \\ \alpha_x(t, L) = u^j(t), \\ \alpha(0, x) = 0, \\ \beta_t + \beta_x + \beta_{xxx} + \alpha\alpha_x = 0, \\ \beta(t, 0) = \beta(t, L) = 0, \\ \beta_x(t, L) = 0, \\ \beta(0, x) = 0. \end{cases} \tag{4-16}$$

Proposition 12 is due to Eduardo Cerpa and Emmanuelle Crépeau if one requires only u to be in $L^2(0, T)$ instead of being in $H^1(0, T)$: see [Cerpa 2007, Proposition 3.1] and [Cerpa and Crépeau 2009a, Proposition 3.1]. We explain in Appendix C how to modify the proof of [Cerpa 2007, Proposition 3.1] (as well as [Cerpa and Crépeau 2009a, Proposition 3.1]) in order to get Proposition 12.

We decompose β into $\beta = \beta_1 + \beta_2$, where $\beta_1 := P_H(\beta)$ and $\beta_2 := P_M(\beta)$. Hence, similarly to Proposition 9, we get

$$\begin{cases} \beta_{2t} + \beta_{2x} + \beta_{2xxx} + P_M(\alpha\alpha_x) = 0, \\ \beta_2(t, 0) = \beta_2(t, L) = 0, \\ \beta_{2x}(t, L) = 0, \\ \beta_2(0, x) = 0, \end{cases} \tag{4-17}$$

where $\beta_2(T, \cdot) = P_M(\beta(T, \cdot)) \neq 0$. In particular, $P_{M^j}(\beta_2(T, \cdot)) = P_{M^j}(\beta(T, \cdot)) \neq 0$.

Combining (4-16) and (4-17), we get:

Corollary 13. *For every $L \in \mathcal{N}_2 \cup \mathcal{N}_3$, for every $T_0 > 0$, for every $j \in \{1, 2, \dots, n\}$, there exists $u_0^j \in L^\infty(0, T_0)$ such that the solution (y_1, y_2) to equation* (4-4) *with $v(t, z) := u_0^j(t)$ satisfies*

$$y_1(T_0) = 0 \quad \text{and} \quad P_{M^j}(y_2(T_0)) \neq 0. \tag{4-18}$$

Now we come back to the case when (4-10) holds. Let us fix $T_0 > 0$ such that

$$T_0 < \tfrac{1}{4} p. \tag{4-19}$$

Let

$$q := \tfrac{1}{4} p. \tag{4-20}$$

Let u_0 be as in Corollary 13. We define

$$Y_1(t) := y_1(t), \quad Y_2(t) := y_2(t) \quad \text{for } t \in [0, T_0] \tag{4-21}$$

and

$$\psi_1 := Y_2(T_0) \in M \setminus \{0\}. \tag{4-22}$$

Let

$$\psi_2 = S(q)\psi_1 \in M, \quad \psi_3 = S(2q)\psi_1 \in M, \quad \psi_4 = S(3q)\psi_1 \in M, \tag{4-23}$$

$$T := 3q + T_0, \tag{4-24}$$

$$K_1 := [3q, 3q + T_0], \tag{4-25}$$

$$K_2 := [2q, 2q + T_0], \tag{4-26}$$

$$K_3 := [q, q + T_0], \tag{4-27}$$

$$K_4 := [0, T_0]. \tag{4-28}$$

Note that (4-19) implies

$$K_1,\ K_2,\ K_3 \text{ and } K_4 \text{ are pairwise disjoint.} \tag{4-29}$$

Let us define four functions $[0, T] \to \mathbb{R}$: u_1, u_2, u_3 and u_4 by requiring that, for every $i \in \{1, 2, 3, 4\}$,

$$u_i := \begin{cases} 0 & \text{on } [0, T] \setminus K_i, \\ u_0(\cdot - \tau_i) & \text{on } K_i, \end{cases} \tag{4-30}$$

with

$$\tau_1 = 3q, \quad \tau_2 = 2q, \quad \tau_3 = q, \quad \tau_4 = 0. \tag{4-31}$$

One can easily verify that, for every $i \in \{1, 2, 3, 4\}$, the solution of (4-4) for $v = u_i$ is given explicitly by

$$y_{i,1}(t) = \begin{cases} 0 & \text{on } [0, T] \setminus K_i, \\ Y_1(\cdot - \tau_i) & \text{on } K_i \end{cases} \tag{4-32}$$

and

$$y_{i,2}(t) = \begin{cases} 0 & \text{on } [0, \tau_i], \\ Y_2(\cdot - \tau_i) & \text{on } K_i, \\ S(\cdot - \tau_i - T_0)\psi_1 & \text{on } [\tau_i + T_0, T]. \end{cases} \tag{4-33}$$

For $z \in M_1$, let α_1, α_2, α_3 and α_4 in $[0, +\infty)$ be such that

$$-S(T)z = \alpha_1\psi_1 + \alpha_2\psi_2 + \alpha_3\psi_3 + \alpha_4\psi_4, \tag{4-34}$$

$$\alpha_1\alpha_3 = 0, \quad \alpha_2\alpha_4 = 0. \tag{4-35}$$

Let us define

$$v(t, z) := \alpha_1 u_1(t) + \alpha_2 u_2(t) + \alpha_3 u_3(t) + \alpha_4 u_4(t). \tag{4-36}$$

We notice that

$$(\alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2)\|\psi_1\|_{L_L^2}^2 = 1, \tag{4-37}$$

which, together with (4-36), implies that

$$v \in L^\infty([0, T] \times M_1; \mathbb{R}). \tag{4-38}$$

Moreover, using the above construction (and in particular (4-29)), one easily checks that the solution of (4-4) satisfies

$$y_1(t) = \alpha_1 y_{1,1}(t) + \alpha_2 y_{2,1}(t) + \alpha_3 y_{3,1}(t) + \alpha_4 y_{4,1}(t) \quad \text{for } t \in [0, T], \tag{4-39}$$

$$y_2(t) = \alpha_1^2 y_{1,2}(t) + \alpha_2^2 y_{2,2}(t) + \alpha_3^2 y_{3,2}(t) + \alpha_4^2 y_{4,2}(t) \quad \text{for } t \in [0, T]. \tag{4-40}$$

In particular

$$y_1(T) = 0, \tag{4-41}$$

$$y_2(T) = \alpha_1^2 \psi_1 + \alpha_2^2 \psi_2 + \alpha_3^2 \psi_3 + \alpha_4^2 \psi_4. \tag{4-42}$$

From (4-34), (4-37) and (4-42), we can find that (4-5) holds if $\delta > 0$ is small enough. It is easy to check that the Lipschitz condition (4-3) is also satisfied. This completes the construction of $v(t, z)$ such that $(\mathcal{P}_3)$ holds and also the proof of Proposition 10 if (4-10) holds.

For other values of $L \in \mathcal{N}_2$, only the values of φ_1, φ_2 and p have to be modified. For $L \in \mathcal{N}_3$, as mentioned in the Introduction, M is now of dimension $2n$, where n is the number of ordered pairs. It is proved in [Cerpa and Crépeau 2009a] that (compare with (4-11)–(4-14)), by a good choice of order on $\{\varphi^j\}$, one can assume

$$0 < p^1 < p^2 < \cdots < p^n, \tag{4-43}$$

where $p^j := 2\pi/\omega^j$. For every $t > 0$, one has

$$S(t)M^j \subset M^j \quad \text{and} \quad S(t) \text{ restricted to } M^j \text{ is the rotation of angle } \frac{2\pi t}{p^j}. \tag{4-44}$$

From (4-43), (4-44) and Corollary 13, one can get the following corollary (see also [Cerpa and Crépeau 2009a, Proposition 3.3]):

Corollary 14. *For every $L \in \mathcal{N}_3$, there exists $T_L > 0$ such that, for every $j \in \{1, 2, \ldots, n\}$, there exists $u_0^j \in L^\infty(0, T_L)$ such that the solution (y_1, y_2) to equation* (4-4) *with $v(t, z) := u_0^j(t)$ satisfies*

$$y_1(T_L) = 0 \quad \textit{and} \quad y_2(T_L) = \varphi_1^j. \tag{4-45}$$

Let us define

$$\psi_1^j := \varphi_1^j, \quad \psi_2^j := S(q^j)\varphi_1^j, \quad \psi_3^j := S(2q^j)\varphi_1^j, \quad \psi_4^j := S(3q^j)\varphi_1^j, \tag{4-46}$$

where $q^j := p^j/4$.

Comparing with (4-22)–(4-33), we can find $T > T_L$ and closed interval sets $\{K_i^j\}$, where $i \in \{1, 2, 3, 4\}$ and $j \in \{1, 2, \ldots, n\}$, such that

$$K_i^j \subset [0, T], \tag{4-47}$$

$$\{K_i^j\} \text{ are pairwise disjoint.} \tag{4-48}$$

We can also find functions $\{u_i^j\} \in L^\infty([0, T]; \mathbb{R})$, with

$$u_i^j(t) \text{ supports on } K_i^j, \tag{4-49}$$

such that when we define the control as u_i^j, we get the solution of (4-4) satisfies

$$y_{i,1}^j(t) \text{ supports on } K_i^j, \tag{4-50}$$

$$y_{i,1}^j(T) = 0, \tag{4-51}$$

$$y_{i,2}^j(T) = \psi_i^j. \tag{4-52}$$

Then for $z \in M_1$, let α_i^j in $[0, +\infty)$ be such that

$$-S(T)z = \sum_{i,j} \alpha_i^j \psi_i^j, \tag{4-53}$$

$$\alpha_1^j \alpha_3^j = 0, \quad \alpha_2^j \alpha_4^j = 0, \quad \sum_{i,j} (\alpha_i^j)^2 = 1, \tag{4-54}$$

where $i \in \{1, 2, 3, 4\}$ and $j \in \{1, 2, \ldots, n\}$. Let us define

$$v(t, z) := \sum_{i,j} \alpha_i^j u_i^j(t). \tag{4-55}$$

Then the solution of (4-4) with control defined as $v(t, z)$ satisfies

$$y_1(T) = 0, \tag{4-56}$$

$$y_2(T) = \sum_{i,j} (\alpha_i^j)^2 \psi_i^j. \tag{4-57}$$

One can easily verify that condition (4-5) holds when $\delta > 0$ is small enough, and that Lipschitz condition (4-3) also holds. This completes the construction of $v(t, z)$ and the proof of Proposition 10. □

We are now able to define the periodic time-varying feedback laws $u_\varepsilon : \mathbb{R} \times L^2(0, L) \to \mathbb{R}$, which will lead to the exponential stabilization of (1-1). For $\varepsilon > 0$, we define u_ε by

$$u_\varepsilon|_{[0,T)\times L_L^2}(t, y) := \begin{cases} 0 & \text{if } \|y^M\|_{L_L^2} = 0, \\ \varepsilon\sqrt{\|y^M\|_{L_L^2}}\, v\big(t, S(-t)y^M/\|y^M\|_{L_L^2}\big) & \text{if } 0 < \|y^M\|_{L_L^2} \leqslant 1, \\ \varepsilon v\big(t, S(-t)y^M/\|y^M\|_{L_L^2}\big) & \text{if } \|y^M\|_{L_L^2} > 1, \end{cases} \tag{4-58}$$

with $y^M := P_M(y)$, and

$$u_\varepsilon(t, y) := u_\varepsilon|_{[0,T)\times L_L^2}(t - [t/T]T, y) \quad \forall t \in \mathbb{R},\ \forall y \in L^2(0, L). \tag{4-59}$$

5. Proof of Theorem 1

Let us first point out that Theorem 1 is a consequence of the following two propositions.

Proposition 15. *There exist $\varepsilon_1 > 0$, $r_1 > 0$ and C_1 such that, for every Carathéodory feedback law u satisfying*

$$|u(t, z)| \leqslant \varepsilon_1 \min\big\{1, \sqrt{\|P_M(z)\|_{L_L^2}}\big\} \quad \forall t \in \mathbb{R},\ \forall z \in L^2(0, L), \tag{5-1}$$

for every $s \in \mathbb{R}$ and for every maximal solution y of (3-6) *defined at time s and satisfying $\|y(s)\|_{L^2_L} < r_1$, y is well-defined on $[s, s+T]$ and one has*

$$\|P_H(y)\|^2_{\mathcal{B}_{s,s+T}} + \|P_M(y)\|_{\mathcal{B}_{s,s+T}} \leqslant C_1\big(\|P_H(y(s))\|^2_{L^2_L} + \|P_M(y(s))\|_{L^2_L}\big). \tag{5-2}$$

Proposition 16. *For ρ_1 as in Proposition 10, let $\rho_2 > \rho_1$. There exists $\varepsilon_0 \in (0, 1)$ such that, for every $\varepsilon \in (0, \varepsilon_0)$, there exists $r_\varepsilon > 0$ such that, for every solution y to* (3-6) *on* $[0, T]$, *for the feedback law $u := u_\varepsilon$ defined in* (4-58) *and* (4-59), *and satisfying $\|y(0)\|_{L^2_L} < r_\varepsilon$, one has*

$$\|P_H(y(T))\|^2_{L^2_L} + \varepsilon\|P_M(y(T))\|_{L^2_L} \leqslant \rho_2\|P_H(y(0))\|^2_{L^2_L} + \varepsilon(1 - \delta\varepsilon^2)\|P_M(y(0))\|_{L^2_L}. \tag{5-3}$$

Indeed, it suffices to choose $\rho_2 \in (\rho_1, 1)$, $\varepsilon \in (0, \varepsilon_0)$ and $u := u_\varepsilon$ defined in (4-58) and (4-59). Then, using the T-periodicity of u with respect to time, Proposition 15 and Proposition 16, one checks that inequality (1-11) holds with

$$\lambda := \min\left\{-\frac{\ln(\rho_2)}{2T}, -\frac{\ln(1-\delta\varepsilon^2)}{2T}\right\}$$

provided that C is large enough and that r is small enough. We now prove Propositions 15 and 16 successively.

Proof of Proposition 15. Performing a time translation if necessary, we may assume without loss of generality that $s = 0$. The fact that the maximal solution y is at least defined on $[0, T]$ follows from Theorem 8 and (5-1). We choose ε_1 and r_1 small enough so that

$$r_1 + \varepsilon_1 T^{1/2} \leqslant \eta, \tag{5-4}$$

where $\eta > 0$ is as in Lemma 5. From (5-1) and (5-4), we have

$$\|y(0)\|_{L^2_L} + \|u(t, y(t))\|_{L^2_T} \leqslant \eta, \tag{5-5}$$

which allows us to apply Lemma 5 with $H(t) := u(t, y(t))$ and $\widetilde{H} := 0$. Then, using (5-1) once more, we get

$$\begin{aligned}\|y\|_{\mathcal{B}} &\leqslant C_3\big(\|y_0\|_{L^2_L} + \|u(t, y(t))\|_{L^2_T}\big)\\ &\leqslant C_3\big(r_1 + \varepsilon_1\sqrt{T\|P_M(y)\|_{C^0L^2_L}}\big) \leqslant C_3\Big(r_1 + \varepsilon_1^2 T C_3 + \frac{1}{4C_3}\|y\|_{\mathcal{B}}\Big),\end{aligned}$$

which implies that

$$\|y\|_{\mathcal{B}} \leqslant 2C_3(r_1 + \varepsilon_1^2 T C_3). \tag{5-6}$$

In the above inequalities and until the end of the proof of Proposition 16, $\mathcal{B} := \mathcal{B}_{0,T}$.

We have the following lemma; see the proof of [Rosier 1997, Proposition 4.1 and (4.14)] or [Perla Menzala et al. 2002, page 121].

Lemma 17. *If $y \in L^2(0, T; H^1(0, L))$, then $yy_x \in L^1(0, T; L^2(0, L))$. Moreover, there exists $c_4 > 0$, which is independent of T, such that, for every $T > 0$ and for every $y, z \in L^2(0, T; H^1(0, L))$, we have*

$$\|yy_x - zz_x\|_{L^1_T L^2_L} \leqslant c_4 T^{1/4}\big(\|y\|_{\mathcal{B}} + \|z\|_{\mathcal{B}}\big)\|y - z\|_{\mathcal{B}}. \tag{5-7}$$

Let us define $C_4 := c_4 T^{1/4}$. To simplify the notation, until the end of this section, we write y_1 and y_2 for $P_H(y)$ and $P_M(y)$ respectively. From (5-1), (5-6), Lemma 3, Lemma 17 and Proposition 9, we get

$$\begin{aligned}\|y_1\|_{\mathcal{B}} &\leqslant C_2\big(\|y_0^H\|_{L_L^2} + \|u(t, y_1+y_2)\|_{L_T^2} + \big\|P_H\big((y_1+y_2)(y_1+y_2)_x\big)\big\|_{L_T^1 L_L^2}\big)\\ &\leqslant C_2\big(\|y_0^H\|_{L_L^2} + \varepsilon_1\big\|\sqrt{\|y_2\|_{L_L^2}}\big\|_{L_T^2} + \big\|(y_1+y_2)(y_1+y_2)_x\big\|_{L_T^1 L_L^2}\big)\\ &\leqslant C_2\big(\|y_0^H\|_{L_L^2} + \varepsilon_1\|y_2\|^{1/2}_{L_T^1 L_L^2} + C_4\|y_1+y_2\|^2_{L_T^2 H_L^1}\big) \qquad (5\text{-}8)\end{aligned}$$

and

$$\begin{aligned}\|y_2\|_{\mathcal{B}} &\leqslant C_2\big(\|y_0^M\|_{L_L^2} + \big\|P_M\big((y_1+y_2)(y_1+y_2)_x\big)\big\|_{L_T^1 L_L^2}\big)\\ &\leqslant C_2\big(\|y_0^M\|_{L_L^2} + \big\|(y_1+y_2)(y_1+y_2)_x\big\|_{L_T^1 L_L^2}\big)\\ &\leqslant C_2\big(\|y_0^M\|_{L_L^2} + C_4\|y_1+y_2\|^2_{L_T^2 H_L^1}\big)\\ &\leqslant 2C_2\big(\|y_0^M\|_{L_L^2} + C_4\|y_1\|^2_{\mathcal{B}} + C_4\|y_2\|^2_{\mathcal{B}}\big). \qquad (5\text{-}9)\end{aligned}$$

Since M is a finite-dimensional subspace of $H^1(0, L)$, there exists $C_5 > 0$ such that

$$\|f\|_{H^1(0,L)} \leqslant C_5\|f\|_{L_L^2} \quad \text{for every } f \in M. \qquad (5\text{-}10)$$

Hence

$$\|y_2\|_{\mathcal{B}} = \|y_2\|_{L_T^\infty L_L^2} + \|y_2\|_{L_T^2 H_L^1} \leqslant \|y_2\|_{L_T^\infty L_L^2} + C_5\sqrt{T}\|y_2\|_{L_T^\infty L_L^2}. \qquad (5\text{-}11)$$

Since $y_2(t)$ is the L^2-orthogonal projection on M of $y(t)$, we have

$$\|y_2\|_{L_T^\infty L_L^2} \leqslant \|y\|_{L_T^\infty L_L^2} \leqslant \|y\|_{\mathcal{B}},$$

which, together with (5-6) and (5-11), implies

$$\|y_2\|_{\mathcal{B}} \leqslant (1 + C_5\sqrt{T})\|y\|_{\mathcal{B}} \leqslant 2(1 + C_5\sqrt{T})C_3(r_1 + \varepsilon_1^2 T C_3). \qquad (5\text{-}12)$$

Decreasing if necessary r_1 and ε_1, we may assume

$$4C_2C_4(1 + C_5\sqrt{T})C_3(r_1 + \varepsilon_1^2 T C_3) < \tfrac{1}{2}. \qquad (5\text{-}13)$$

From estimation (5-9) and condition (5-13), we get

$$\|y_2\|_{\mathcal{B}} \leqslant 4C_2\big(\|y_0^M\|_{L_L^2} + C_4\|y_1\|^2_{\mathcal{B}}\big). \qquad (5\text{-}14)$$

From (5-6), (5-8), (5-12) and (5-14), we deduce that

$$\begin{aligned}\|y_1\|^2_{\mathcal{B}} &\leqslant 3C_2^2\big(\|y_0^H\|^2_{L_L^2} + \varepsilon_1^2\|y_2\|_{L_T^1 L_L^2} + C_4^2\|y_1+y_2\|^4_{L_T^2 H_L^1}\big)\\ &\leqslant 3C_2^2\big(\|y_0^H\|^2_{L_L^2} + \varepsilon_1^2 T\|y_2\|_{L_T^\infty L_L^2} + 2C_4^2\|y\|^2_{\mathcal{B}}\big(\|y_1\|^2_{\mathcal{B}} + \|y_2\|^2_{\mathcal{B}}\big)\big)\\ &\leqslant 3C_2^2\|y_0^H\|^2_{L_L^2} + 3C_2^2\big(\varepsilon_1^2 T + 16C_4^2(1+C_5\sqrt{T})C_3^3(r_1+\varepsilon_1^2 T C_3)^3\big)\|y_2\|_{\mathcal{B}}\\ &\qquad + 24C_2^2C_4^2C_3^2(r_1+\varepsilon_1^2 T C_3)^2\|y_1\|^2_{\mathcal{B}}\\ &\leqslant 3C_2^2\|y_0^H\|^2_{L_L^2} + 12C_2^3\big(\varepsilon_1^2 T + 16C_4^2(1+C_5\sqrt{T})C_3^3(r_1+\varepsilon_1^2 T C_3)^3\big)\|y_0^M\|_{L_L^2}\\ &\quad + \Big(12C_2^3C_4\big(\varepsilon_1^2 T + 16C_4^2(1+C_5\sqrt{T})C_3^3(r_1+\varepsilon_1^2 T C_3)^3\big) + 24C_2^2C_4^2C_3^2(r_1+\varepsilon_1^2 T C_3)^2\Big)\|y_1\|^2_{\mathcal{B}}. \qquad (5\text{-}15)\end{aligned}$$

Again, decreasing if necessary r_1 and ε_1, we may assume

$$12C_2^3C_4\big(\varepsilon_1^2T+16C_4^2(1+C_5\sqrt{T})C_3^3(r_1+\varepsilon_1^2TC_3)^3\big)+24C_2^2C_4^2C_3^2(r_1+\varepsilon_1^2TC_3)^2<\tfrac{1}{2}. \tag{5-16}$$

From (5-15) and (5-16), we get

$$\begin{aligned}\|y_1\|_{\mathcal{B}}^2&\leqslant 6C_2^2\|y_0^H\|_{L_L^2}^2+24C_2^3\big(\varepsilon_1^2T+16C_4^2(1+C_5\sqrt{T})C_3^3(r_1+\varepsilon_1^2TC_3)^3\big)\|y_0^M\|_{L_L^2}\\&\leqslant 6C_2^2\|y_0^H\|_{L_L^2}^2+C_4^{-1}\|y_0^M\|_{L_L^2},\end{aligned}$$

which, combined with (5-14), gives the existence of $C_1>0$ independent of y such that

$$\|y_1\|_{\mathcal{B}}^2+\|y_2\|_{\mathcal{B}}\leqslant C_1\big(\|y_0^H\|_{L_L^2}^2+\|y_0^M\|_{L_L^2}\big). \tag{5-17}$$

This completes the proof of Proposition 15. □

Proof of Proposition 16. To simplify the notation, from now on we denote by C various constants which vary from place to place but do not depend on ε and r.

By Lemma 3 applied with $y:=y_1(t)-S(t)y_0^H$, $h(t):=u_\varepsilon(t,y(t))$ and $\tilde{h}:=(y_1+y_2)(y_1+y_2)_x$ and by Proposition 15, we have

$$\begin{aligned}\|y_1(t)-S(t)y_0^H\|_{\mathcal{B}}&\leqslant C\big(\|u_\varepsilon\|_{L_T^2}+\big\|P_H((y_1+y_2)(y_1+y_2)_x)\big\|_{L_T^1L_L^2}\big)\\&\leqslant C\big(\varepsilon\|y_2\|_{L_T^1L_L^2}^{1/2}+\|y_1+y_2\|_{\mathcal{B}}^2\big)\\&\leqslant C\big(\varepsilon\|y_2\|_{\mathcal{B}}^{1/2}+\|y_1\|_{\mathcal{B}}^2+\|y_2\|_{\mathcal{B}}^2\big)\\&\leqslant C(\varepsilon+\sqrt{r})\big(\|y_0^H\|_{L_L^2}^2+\|y_0^M\|_{L_L^2}\big)^{1/2},\end{aligned} \tag{5-18}$$

where $r:=\|y_0\|_{L_L^2}<r_\varepsilon<1$. On r_ε, we impose that

$$r_\varepsilon<\varepsilon^{12}. \tag{5-19}$$

From (5-18) and (5-19), we have

$$\|y_1(t)-S(t)y_0^H\|_{\mathcal{B}}\leqslant C\varepsilon\big(\|y_0^H\|_{L_L^2}^2+\|y_0^M\|_{L_L^2}\big)^{1/2}. \tag{5-20}$$

Notice that, by Lemma 3, we have

$$\|S(t)y_0^M\|_{\mathcal{B}}\leqslant C\|y_0^M\|_{L_L^2}, \tag{5-21}$$

$$\|S(t)y_0^H\|_{\mathcal{B}}\leqslant C\|y_0^H\|_{L_L^2}. \tag{5-22}$$

Proceeding as in the proof of (5-20), we have

$$\begin{aligned}\|y_2(t)-S(t)y_0^M\|_{\mathcal{B}}&\leqslant C\big\|P_M((y_1+y_2)(y_1+y_2)_x)\big\|_{L_T^1L_L^2}\\&\leqslant C\|y_1+y_2\|_{\mathcal{B}}^2\\&\leqslant C\big(\|y_2\|_{\mathcal{B}}+\|S(t)y_0^H\|_{\mathcal{B}}+\varepsilon\big(\|y_0^H\|_{L_L^2}^2+\|y_0^M\|_{L_L^2}\big)^{1/2}\big)^2\\&\leqslant C\big((r+\varepsilon^2)\big(\|y_0^H\|_{L_L^2}^2+\|y_0^M\|_{L_L^2}\big)+\|y_0^H\|_{L_L^2}^2\big)\\&\leqslant C\big(\varepsilon^2\|y_0^M\|_{L_L^2}+\|y_0^H\|_{L_L^2}^2\big).\end{aligned} \tag{5-23}$$

Let us now study successively the two cases

$$\|y_0^H\|_{L_L^2} \geqslant \varepsilon^{2/3}\sqrt{\|y_0^M\|_{L_L^2}}, \tag{5-24}$$

$$\|y_0^H\|_{L_L^2} < \varepsilon^{2/3}\sqrt{\|y_0^M\|_{L_L^2}}. \tag{5-25}$$

We start with the case where (5-24) holds. From $(\mathcal{P}_1)$, $(\mathcal{P}_2)$, (5-20), (5-23) and (5-24), we get the existence of $\varepsilon_2 \in (0, \varepsilon_1)$ such that, for every $\varepsilon \in (0, \varepsilon_2)$,

$$\begin{aligned}
&\|y_1(T)\|_{L_L^2}^2 + \varepsilon\|y_2(T)\|_{L_L^2} \\
&\quad\leqslant \big(C\varepsilon\big(\|y_0^H\|_{L_L^2}^2 + \|y_0^M\|_{L_L^2}\big)^{1/2} + \|S(T)y_0^H\|_{L_L^2}\big)^2 + \varepsilon\big(C\big(\varepsilon^2\|y_0^M\|_{L_L^2} + \|y_0^H\|_{L_L^2}^2\big) + \|S(T)y_0^M\|_{L_L^2}\big) \\
&\quad\leqslant (\rho_1\rho_2)^{1/2}\|y_0^H\|_{L_L^2}^2 + C\varepsilon^2\big(\|y_0^H\|_{L_L^2}^2 + \|y_0^M\|_{L_L^2}\big) + C\varepsilon\|y_0^H\|_{L_L^2}^2 + (\varepsilon + C\varepsilon^3)\|y_0^M\|_{L_L^2} \\
&\quad\leqslant \rho_2\|y_0^H\|_{L_L^2}^2 + \varepsilon(1 - \delta\varepsilon^2)\|y_0^M\|_{L_L^2}.
\end{aligned} \tag{5-26}$$

Let us now study the case where (5-25) holds. Let us define

$$b := y_0^M. \tag{5-27}$$

Then, from (5-20), (5-22), (5-23) and (5-25), we get

$$\|y_1(t)\|_{\mathcal{B}} \leqslant \|S(t)y_0^H\|_{\mathcal{B}} + C\varepsilon\big(\|y_0^H\|_{L_L^2}^2 + \|y_0^M\|_{L_L^2}\big)^{1/2} \leqslant C\varepsilon\sqrt{\|b\|_{L_L^2}} + C\|y_0^H\|_{L_L^2} \leqslant C\varepsilon^{2/3}\sqrt{\|b\|_{L_L^2}} \tag{5-28}$$

and

$$\|y_2(t) - S(t)y_0^M\|_{\mathcal{B}} \leqslant \varepsilon^{4/3}\|b\|_{L_L^2}, \tag{5-29}$$

which shows that $y_2(\cdot)$ is close to $S(\cdot)y_0^M$. Let $z : [0, T] \to L^2(0, L)$ be the solution to the Cauchy problem

$$\begin{cases}
z_{1t} + z_{1xxx} + z_{1x} = 0 & \text{in } (0, T) \times (0, L), \\
z_1(t, 0) = z_1(t, L) = 0 & \text{on } (0, T), \\
z_{1x}(t, L) = v(t, b/\|b\|_{L_L^2}) & \text{on } (0, T), \\
z_1(0, x) = 0 & \text{on } (0, L).
\end{cases} \tag{5-30}$$

From $(\mathcal{P}_3)$, we know that $z_1(T) = 0$. Moreover, Lemma 3 tells us that

$$\|z_1(t)\|_{\mathcal{B}} \leqslant C\left\|v\left(t, \frac{b}{\|b\|_{L_L^2}}\right)\right\|_{L_T^2} \leqslant C. \tag{5-31}$$

Let us define w_1 by

$$w_1 := y_1 - S(t)y_0^H - \varepsilon\|b\|_{L_L^2}^{1/2} z_1. \tag{5-32}$$

Then w_1 is the solution to the Cauchy problem

$$\begin{cases}
w_{1t} + w_{1xxx} + w_{1x} + P_H\big((y_1 + y_2)(y_1 + y_2)_x\big) = 0, \\
w_1(t, 0) = w_1(t, L) = 0, \\
w_{1x}(t, L) = \varepsilon\big(\|y_2(t)\|_{L_L^2}^{1/2} v\big(t, S(-t)y_2(t)/\|y_2(t)\|_{L_L^2}\big) - \|b\|_{L_L^2}^{1/2} v(t, b/\|b\|_{L_L^2})\big), \\
w_1(0, x) = 0.
\end{cases} \tag{5-33}$$

By Lemma 3, we get

$$\|w_1\|_{\mathcal{B}} \leqslant C\big\| P_H\big((y_1+y_2)(y_1+y_2)_x\big)\big\|_{L^1_T L^2_L} + \varepsilon C\left\| \left(\|y_2(t)\|^{1/2}_{L^2_L} v\left(t, \frac{S(-t)y_2(t)}{\|y_2(t)\|_{L^2_L}}\right) - \|b\|^{1/2}_{L^2_L} v\left(t, \frac{b}{\|b\|_{L^2_L}}\right)\right)\right\|_{L^2_T}. \quad (5\text{-}34)$$

Note that (5-29) ensures that the right-hand side of (5-34) is of order ε^2. Indeed, for the first term of the right-hand side of inequality (5-34), we have, using (5-19), (5-28) and (5-29),

$$\begin{aligned} C\big\| P_H\big((y_1+y_2)(y_1+y_2)_x\big)\big\|_{L^1_T L^2_L} &\leqslant C\|y_1+y_2\|^2_{\mathcal{B}} \\ &\leqslant C\varepsilon^{4/3}\|b\|_{L^2_L} + C\|b\|_{L^2_L} \leqslant C\|b\|^{1/2}_{L^2_L}\|b\|^{1/2}_{L^2_L} \leqslant C\varepsilon^6\|b\|^{1/2}_{L^2_L}. \end{aligned} \quad (5\text{-}35)$$

For the second term of the right-hand side of inequality (5-34), by (4-14), the Lipschitz condition (4-3) on v and (5-29), we get, for every $t \in [0, T]$,

$$\begin{aligned} &\left| \|b\|^{1/2}_{L^2_L}\left(v\left(t, \frac{b}{\|b\|_{L^2_L}}\right) - v\left(t, \frac{S(-t)y_2(t)}{\|y_2(t)\|_{L^2_L}}\right)\right)\right| \\ &\quad\leqslant C\|b\|^{1/2}_{L^2_L}\left\| \left(\frac{b}{\|b\|_{L^2_L}} - \frac{S(-t)y_2(t)}{\|y_2(t)\|_{L^2_L}}\right)\right\|_{L^2_L} \\ &\quad\leqslant C\|b\|^{-1/2}_{L^2_L}\|y_2(t)\|^{-1}_{L^2_L}\big(\|y_2(t)\|_{L^2_L}\|b - S(-t)y_2(t)\|_{L^2_L} + \|S(-t)y_2(t)\|_{L^2_L}\big|\|y_2(t)\|_{L^2_L} - \|b\|_{L^2_L}\big|\big) \\ &\quad\leqslant C\varepsilon^{4/3}\|b\|^{1/2}_{L^2_L} \end{aligned} \quad (5\text{-}36)$$

and

$$\left| \big(\|y_2(t)\|^{1/2}_{L^2_L} - \|b\|^{1/2}_{L^2_L}\big) v\left(t, \frac{S(-t)y_2(t)}{\|y_2(t)\|_{L^2_L}}\right)\right| \leqslant C\varepsilon^{4/3}\|b\|^{1/2}_{L^2_L}. \quad (5\text{-}37)$$

Combining (5-35)–(5-37), we obtain the following estimate on w_1:

$$\|w_1\|_{\mathcal{B}} \leqslant C\varepsilon^2\|b\|^{1/2}_{L^2_L}. \quad (5\text{-}38)$$

We fix

$$\rho_3 \in (\rho_1, \rho_2). \quad (5\text{-}39)$$

Then, by (5-32), $(\mathcal{P}_1)$ and the fact that $z_1(T) = 0$, we get

$$\|y_1(T)\|^2_{L^2_L} \leqslant \rho_3\|y_0^H\|^2_{L^2_L} + C\varepsilon^4\|b\|_{L^2_L}. \quad (5\text{-}40)$$

We then come to the estimate of y_2. Let $\tau_1(t) := S(t)y_0^H$ and let $\tau_2 : [0, T] \to L^2(0, L)$ and $z_2 : [0, T] \to L^2(0, L)$ be the solutions to the Cauchy problems

$$\begin{cases} \tau_{2t} + \tau_{2xxx} + \tau_{2x} + P_M(\tau_1 y_{1x} + \tau_{1x} y_1) - P_M(\tau_1\tau_{1x}) = 0, \\ \tau_2(t, 0) = \tau_2(t, L) = 0, \\ \tau_{2x}(t, L) = 0, \\ \tau_2(0, x) = 0 \end{cases} \quad (5\text{-}41)$$

and

$$\begin{cases} z_{2t} + z_{2xxx} + z_{2x} + P_M(z_1 z_{1x}) = 0, \\ z_2(t,0) = z_2(t,L) = 0, \\ z_{2x}(t,L) = 0, \\ z_2(0,x) = 0. \end{cases} \tag{5-42}$$

Lemmas 3 and 17, (5-25) and (5-28) show us that

$$\begin{aligned} \|\tau_2\|_{\mathcal{B}} &\leqslant C \left\| P_M\big(\tau_1 y_{1x} + \tau_{1x} y_1 - \tau_1 \tau_{1x}\big) \right\|_{L_T^1 L_L^2} \\ &\leqslant C \|\tau_1\|_{\mathcal{B}} \big(\|y_1\|_{\mathcal{B}} + \|\tau_1\|_{\mathcal{B}} \big) \\ &\leqslant C \varepsilon^{2/3} \|b\|_{L_L^2}^{1/2} \|y_0^H\|_{L_L^2} \end{aligned} \tag{5-43}$$

and

$$\|z_2\|_{\mathcal{B}} \leqslant \|z_1\|_{\mathcal{B}}^2 \leqslant C. \tag{5-44}$$

From $(\mathcal{P}_3)$, (5-30) and (5-42), we get

$$\langle z_2(T), S(T)b \rangle_{(L_L^2, L_L^2)} \ < \ -2\delta \|b\|_{L_L^2}. \tag{5-45}$$

Hence

$$\begin{aligned} \left\| S(T)b + \varepsilon^2 \|b\|_{L_L^2} z_2(T) \right\|_{L_L^2} &= \left(\left\langle S(T)b + \varepsilon^2 \|b\|_{L_L^2} z_2(T),\, S(T)b + \varepsilon^2 \|b\|_{L_L^2} z_2(T) \right\rangle_{(L_L^2, L_L^2)} \right)^{1/2} \\ &\leqslant \left(\|b\|_{L_L^2}^2 + \varepsilon^4 \|b\|_{L_L^2}^2 C - 4\delta\varepsilon^2 \|b\|_{L_L^2}^2 \right)^{1/2} \\ &\leqslant \|b\|_{L_L^2} (1 - 2\delta\varepsilon^2 + C\varepsilon^4). \end{aligned} \tag{5-46}$$

Let us define $w_2 : [0, T] \to L^2(0, L)$ by

$$w_2 := y_2 - \tau_2 - \varepsilon^2 \|b\|_{L_L^2} z_2 - S(t)b. \tag{5-47}$$

Then, from (3-15), (5-41) and (5-42), we get that

$$\begin{aligned} w_{2t} &= y_{2t} - \tau_{2t} - \varepsilon^2 \|b\|_{L_L^2} z_{2t} - (S(t)b)_t \\ &= -w_{2x} - w_{2xxx} - P_M((y_1+y_2)(y_1+y_2)_x) + P_M(\tau_1 y_{1x} + \tau_{1x} y_1) - P_M(\tau_1 \tau_{1x}) + \varepsilon^2 \|b\|_{L_L^2} P_M(z_1 z_{1x}) \\ &= -w_{2x} - w_{2xxx} - \varepsilon \|b\|_{L_L^2}^{1/2} P_M(w_1 z_{1x} + w_{1x} z_1) - P_M(w_1 w_{1x}) - P_M(y_1 y_{2x} + y_2 y_{1x} + y_2 y_{2x}). \end{aligned}$$

Hence, w_2 is the solution to the Cauchy problem

$$\begin{cases} w_{2t} + w_{2xxx} + w_{2x} + \varepsilon \|b\|_{L_L^2}^{1/2} P_M(w_1 z_{1x} + w_{1x} z_1) + P_M(w_1 w_{1x}) + P_M(y_1 y_{2x} + y_2 y_{1x} + y_2 y_{2x}) = 0, \\ w_2(t,0) = w_2(t,L) = 0, \\ w_{2x}(t,L) = 0, \\ w_2(0,x) = 0. \end{cases} \tag{5-48}$$

From Lemmas 3 and 17, Proposition 15, (5-19), (5-25) and (5-38), we get

$$\begin{aligned}\|w_2\|_{\mathcal{B}} &\leqslant C\varepsilon\|b\|_{L^2_L}^{1/2}\big\|P_M(w_1z_{1x}+w_{1x}z_1)\big\|_{L^1_TL^2_L}+C\|P_M(w_1w_{1x})\|_{L^1_TL^2_L}\\ &\qquad +C\big\|P_M(y_1y_{2x}+y_2y_{1x}+y_2y_{2x})\big\|_{L^1_TL^2_L}\\ &\leqslant C\varepsilon\|b\|_{L^2_L}^{1/2}\varepsilon^2\|b\|_{L^2_L}^{1/2}+C\varepsilon^4\|b\|_{L^2_L}+C\big(\|y_0^H\|_{L^2_L}^2+\|y_0^M\|_{L^2_L}\big)^{3/2}\\ &\leqslant C\varepsilon^3\|b\|_{L^2_L}. \end{aligned}\tag{5-49}$$

We can now estimate $y_2(T)$ from (5-43), (5-46), (5-47) and (5-49):

$$\begin{aligned}\|y_2(T)\|_{L^2_L}&=\big\|w_2(T)+\tau_2(T)+\varepsilon^2\|b\|_{L^2_L}z_2(T)+S(T)b\big\|_{L^2_L}\\ &\leqslant\|b\|_{L^2_L}\big(C\varepsilon^3+1-2\delta\varepsilon^2+C\varepsilon^4\big)+C\varepsilon^{2/3}\|b\|_{L^2_L}^{1/2}\|y_0^H\|_{L^2_L}. \end{aligned}\tag{5-50}$$

Combining (5-27), (5-39), (5-40) and (5-50), we get the existence of $\varepsilon_3>0$ such that, for every $\varepsilon\in(0,\varepsilon_3]$,

$$\begin{aligned}&\|y_1(T)\|_{L^2_L}^2+\varepsilon\|y_2(T)\|_{L^2_L}\\ &\quad\leqslant\rho_3\|y_0^H\|_{L^2_L}^2+C\varepsilon^4\|b\|_{L^2_L}+\varepsilon\big(\|b\|_{L^2_L}\big(C\varepsilon^3+1-2\delta\varepsilon^2+C\varepsilon^4\big)+C\varepsilon^{2/3}\|b\|_{L^2_L}^{1/2}\|y_0^H\|_{L^2_L}\big)\\ &\quad\leqslant\rho_2\|y_0^H\|_{L^2_L}^2+\varepsilon(1-\delta\varepsilon^2)\|y_0^M\|_{L^2_L}. \end{aligned}\tag{5-51}$$

This concludes the proof of Proposition 16. □

Appendix A: Proof of Proposition 9

Proof of Proposition 9. It is clear that, if (y_1, y_2) is a solution to (3-15), then y is solution to (3-7). Let us assume that y is a solution to the Cauchy problem (3-7). Then, by Definition 4, for every $\tau\in[s,T]$ and for every $\phi\in C^3([s,\tau]\times[0,L])$ satisfying

$$\phi(t,0)=\phi(t,L)=\phi_x(t,0)=0\quad\forall t\in[s,\tau],\tag{A-1}$$

we have

$$\begin{aligned}-\int_s^\tau\int_0^L(\phi_t+\phi_x+\phi_{xxx})y\,dx\,dt-\int_s^\tau u(t,y(t,\cdot))\phi_x(t,L)\,dt+\int_s^\tau\int_0^L\phi yy_x\,dx\,dt\\ +\int_0^L y(\tau,x)\phi(\tau,x)\,dx-\int_0^L y_0\phi(s,x)\,dx=0.\end{aligned}\tag{A-2}$$

Let us denote by ϕ_1 and ϕ_2 the projections of ϕ on H and M respectively: $\phi_1:=P_H(\phi)$, $\phi_2:=P_M(\phi)$. Because M is spanned by φ_1^j and φ_2^j, $j\in\{1,\dots,n\}$, which are of class C^∞ and satisfy

$$\begin{aligned}\varphi_1^j(0)=\varphi_1^j(L)=\varphi_{1x}^j(0)=\varphi_{1x}^j(L)=0,\\ \varphi_2^j(0)=\varphi_2^j(L)=\varphi_{2x}^j(0)=\varphi_{2x}^j(L)=0,\end{aligned}$$

the functions $\phi_1,\phi_2\in C^3([s,\tau]\times[0,L])$ and satisfy

$$\phi_1(t,0)=\phi_1(t,L)=\phi_{1x}(t,0)=0\qquad\forall t\in[s,\tau],\tag{A-3}$$

$$\phi_2(t,0)=\phi_2(t,L)=\phi_{2x}(t,0)=\phi_{2x}(t,L)=0\quad\forall t\in[s,\tau].\tag{A-4}$$

Using (A-2) for $\phi = \phi_2$ in (A-2) together with (A-4), we get

$$-\int_s^\tau\int_0^L (\phi_{2t}+\phi_{2x}+\phi_{2xxx})y\,dx\,dt+\int_s^\tau\int_0^L \phi_2 yy_x\,dx\,dt + \int_0^L y(\tau,x)\phi_2(\tau,x)\,dx-\int_0^L y_0\phi_2(s,x)\,dx=0, \quad \text{(A-5)}$$

which, combined with the fact that $\phi_{2t}+\phi_{2x}+\phi_{2xxx}\in M$, gives

$$-\int_s^\tau\int_0^L (\phi_{2t}+\phi_{2x}+\phi_{2xxx})y_2\,dx\,dt+\int_s^\tau\int_0^L \phi_2 P_M(yy_x)\,dx\,dt + \int_0^L y_2(\tau,x)\phi_2(\tau,x)\,dx-\int_0^L P_M(y_0)\phi_2(s,x)\,dx=0. \quad \text{(A-6)}$$

Simple integrations by parts show that $\phi_{1x}+\phi_{1xxx}\in M^\perp = H$. Since, ϕ_1 and ϕ_{1t} are also in H, we get from (A-6) that

$$-\int_s^\tau\int_0^L (\phi_{t}+\phi_{x}+\phi_{xxx})y_2\,dx\,dt+\int_s^\tau\int_0^L \phi P_M(yy_x)\,dx\,dt + \int_0^L y_2(\tau,x)\phi(\tau,x)\,dx-\int_0^L P_M(y_0)\phi(s,x)\,dx=0, \quad \text{(A-7)}$$

which is exactly the definition of a solution of the y_2-part of the linear KdV system (3-15). We then combine (A-2) and (A-7) to get

$$-\int_s^\tau\int_0^L (\phi_{t}+\phi_{x}+\phi_{xxx})y_1\,dx\,dt-\int_s^\tau u(t,y(t,\cdot))\phi_x(t,L)\,dt+\int_s^\tau\int_0^L \phi P_H(yy_x)\,dx\,dt + \int_0^L y_1(\tau,x)\phi(\tau,x)\,dx-\int_0^L P_H(y_0)\phi(0,x)\,dx=0, \quad \text{(A-8)}$$

and we get the definition of a solution to the y_1-part of the linear KdV system (3-15). This concludes the proof of Proposition 9. □

Appendix B: Proofs of Theorems 7 and 8

Our strategy to prove Theorem 7 is to prove first the existence of a solution for small times and then to use some a priori estimates to control the L^2_L-norm of the solution with which we can extend the solution to a longer time, and to continue until the solution blows up. We start by proving the following lemma.

Lemma 18. *Let $C_2>0$ be as in Lemma 3 for $T_2-T_1=1$. Assume that u is a Carathéodory function and that, for every $R>0$, there exists $K(R)>0$ such that*

$$\big(\|y\|_{L^2_L}\leqslant R \quad \text{and} \quad \|z\|_{L^2_L}\leqslant R\big) \implies \big(|u(t,y)-u(t,z)|\leqslant K(R)\|y-z\|_{L^2_L} \quad \forall t\in\mathbb{R}\big). \quad \text{(B-1)}$$

Then, for every $R\in(0,+\infty)$, there exists a time $T(R)>0$ such that, for every $s\in\mathbb{R}$ and for every $y_0\in L^2(0,L)$ with $\|y_0\|_{L^2_L}\leqslant R$, the Cauchy problem (3-7) *has one and only one solution y on $[s,s+T(R)]$. Moreover, this solution satisfies*

$$\|y\|_{\mathcal{B}_{s,s+T(R)}}\leqslant C_R := 3C_2R. \quad \text{(B-2)}$$

Proof of Lemma 18. Let us first point out that it follows from our choice of C_2 and Lemma 3 that, for every $-\infty < T_1 < T_2 < +\infty$ such that $T_2 - T_1 \leqslant 1$, for every solution y of problem (2-1), estimation (2-12) holds.

Let $y_0 \in L^2(0, L)$ be such that

$$\|y_0\|_{L^2_L} \leqslant R. \tag{B-3}$$

Let us define $\mathcal{B}_1$ by

$$\mathcal{B}_1 := \{y \in \mathcal{B}_{s,s+T(R)} : \|y\|_{\mathcal{B}_{s,s+T(R)}} \leqslant C_R\}.$$

The set $\mathcal{B}_1$ is a closed subset of $\mathcal{B}_{s,s+T(R)}$. For every $y \in \mathcal{B}_1$, we define $\Psi(y)$ as the solution of (2-1) with $\tilde{h} := -yy_x,\ h(t) := u(t, y(t, \cdot))$ and $y_0 := y_0$. Let us prove that, for $T(R)$ small enough, the smallness being independent of y_0 provided that it satisfies (B-3), we have

$$\Psi(\mathcal{B}_1) \subset \mathcal{B}_1. \tag{B-4}$$

Indeed for $y \in \mathcal{B}_1$, by Lemmas 3 and 17, we have, if $T(R) \leqslant 1$,

$$\begin{aligned}\|\Psi(y)\|_{\mathcal{B}} &\leqslant C_2\big(\|y_0\|_{L^2_L} + \|h\|_{L^2_T} + \|\tilde{h}\|_{L^1(0,T;L^2(0,L))}\big)\\ &\leqslant C_2\big(\|y_0\|_{L^2_L} + \|u(t, y(t, \cdot))\|_{L^2_T} + \|-yy_x\|_{L^1(s,s+T(R);L^2(0,L))}\big)\\ &\leqslant C_2\big(R + C_B(C_R)T(R)^{1/2} + c_4T(R)^{1/4}\|y\|^2_{\mathcal{B}}\big).\end{aligned} \tag{B-5}$$

In (B-5) and until the end of the proof of Lemma 18, for ease of notation, we simply write $\|\cdot\|_{\mathcal{B}}$ for $\|\cdot\|_{\mathcal{B}_{s,s+T(R)}}$. From (B-5), we get that, if

$$T(R) \leqslant \min\left\{\left(\frac{R}{C_B(C_R)}\right)^2, \left(\frac{1}{9c_4C_2^2R}\right)^4, 1\right\}, \tag{B-6}$$

then (B-4) holds. From now on, we assume that (B-6) holds.

Note that every $y \in \mathcal{B}_1$ such that $\Psi(y) = y$ is a solution of (3-7). In order to use the Banach fixed point theorem, it remains to estimate $\|\Psi(y) - \Psi(z)\|_{\mathcal{B}}$. We know that $\Psi(y) - \Psi(z)$ is the solution of equation (2-1) with $T_1 := s,\ T_2 = s + T(R),\ \tilde{h} := -yy_x + zz_x,\ h(t) := u(t, y(t, \cdot)) - u(t, z(t, \cdot))$ and $y_0 := 0$. Hence, from Lemmas 3 and 17 and (B-1), we get

$$\begin{aligned}\|\Psi(y) - \Psi(z)\|_{\mathcal{B}} &\leqslant C_2\big(\|y_0\|_{L^2_L} + \|h\|_{L^2_T} + \|\tilde{h}\|_{L^1(0,T;L^2(0,L))}\big)\\ &\leqslant C_2\big(0 + T(R)^{1/2}K(C_R)\|y - z\|_{\mathcal{B}} + c_4T(R)^{1/4}\|y - z\|_{\mathcal{B}}\big(\|y\|_{\mathcal{B}} + \|z\|_{\mathcal{B}}\big)\big)\\ &\leqslant C_2\|y - z\|_{\mathcal{B}}\big(T(R)^{1/2}K(C_R) + 2c_4T(R)^{1/4}C_R\big),\end{aligned}$$

which shows that, if

$$T(R) \leqslant \min\left\{\left(\frac{1}{12c_4C_2^2R}\right)^4, \left(\frac{1}{4C_2K(3C_2R)}\right)^2\right\}, \tag{B-7}$$

then,

$$\|\Psi(y) - \Psi(z)\|_{\mathcal{B}} \leqslant \tfrac{3}{4}\|y - z\|_{\mathcal{B}}.$$

Hence, by the Banach fixed point theorem, there exists $y \in \mathcal{B}_1$ such that $\Psi(y) = y$, which is the solution that we are looking for. We define $T(R)$ as

$$T(R) := \min\left\{\left(\frac{R}{C_B(3C_2R)}\right)^2, \left(\frac{1}{12c_4C_2^2R}\right)^4, \left(\frac{1}{4C_2K(3C_2R)}\right)^2, 1\right\}. \tag{B-8}$$

It only remains to prove the uniqueness of the solution to the Cauchy problem (3-7) (the above proof gives only the uniqueness in the set $\mathcal{B}_1$). Clearly it suffices to prove that two solutions to (3-6) which are equal at a time τ are equal in a neighborhood of τ in $[\tau, +\infty)$. This property follows from the above proof and from the fact that, for every solution $y : [\tau, \tau_1] \to L^2(0, L)$ of (3-7), if $T > 0$ is small enough (the smallness depending on y),

$$\|y\|_{\mathcal{B}_{\tau,\tau+T}} \leqslant 3C_2\|y(\tau)\|_{L^2_L}. \tag{B-9}$$

This concludes the proof of Lemma 18. □

Proceeding similarly to the proof of Lemma 18, one can get the following lemma concerning the Cauchy problem (2-13).

Lemma 19. *Let $C_2 > 0$ be as in Lemma 3 for $T_2 - T_1 = 1$. Given $R, M > 0$, there exists $T(R, M) > 0$ such that, for every $s \in \mathbb{R}$, for every $y_0 \in L^2(0, L)$ with $\|y_0\|_{L^2_L} \leqslant R$, and for every measurable $H : (s, s + T(R, M)) \to \mathbb{R}$ such that $|H(t)| \leqslant M$ for every $t \in (s, s + T(R, M))$, the Cauchy problem*

$$\begin{cases} y_t + y_{xxx} + y_x + yy_x = 0 & \text{in } (s, s + T(R, M)) \times (0, L), \\ y(t, 0) = y(t, L) = 0 & \text{on } (s, s + T(R, M)), \\ y_x(t, L) = H(t) & \text{on } (s, s + T(R, M)), \\ y(s, x) = y_0(x) & \text{on } (0, L) \end{cases} \tag{B-10}$$

has one and only one solution y on $[s, s + T(R, M)]$. Moreover, this solution satisfies

$$\|y\|_{\mathcal{B}_{s,s+T(R,M)}} \leqslant 3C_2R. \tag{B-11}$$

We are now in position to prove Theorem 7.

Proof of Theorem 7. The uniqueness follows from the proof of the uniqueness part of Lemma 18. Let us give the proof of the existence. Let $y_0 \in L^2(0, L)$, let $s \in \mathbb{R}$ and let $T_0 := T(\|y_0\|_{L^2_L})$. By Lemma 18, there exists a solution $y \in \mathcal{B}_{s,s+T_0}$ to the Cauchy problem (3-7). Hence, together with the uniqueness of the solution, we can find a maximal solution $y : D(y) \to L^2(0, L)$ with $[s, s + T_0] \subset D(y)$. By the maximality of the solution y and Lemma 18, there exists $\tau \in [s + T_0, +\infty)$ such that $D(y) = [s, \tau)$. Let us assume that $\tau < +\infty$ and that (3-12) does not hold. Then there exist an increasing sequence $(t_n)_{n\in\mathbb{N}}$ of real numbers in (s, τ) and $R \in (0, +\infty)$ such that

$$\lim_{n\to+\infty} t_n = \tau, \tag{B-12}$$

$$\|y(t_n)\|_{L^2_L} \leq R \quad \forall n \in \mathbb{N}. \tag{B-13}$$

By (B-12), there exists $n_0 \in \mathbb{N}$ such that

$$t_{n_0} \geq \tau - \tfrac{1}{2}T(R). \tag{B-14}$$

From Lemma 18, there is a solution $z : [t_{n_0}, t_{n_0} + T(R)] \to L^2(0, L)$ of (3-7) for the initial time $s := t_{n_0}$ and the initial data $z(t_{n_0}) := y(t_{n_0})$. Let us then define $\tilde{y} : [s, t_{n_0} + T(R)] \to L^2(0, L)$ by

$$\tilde{y}(t) := y(t) \quad \forall t \in [s, t_{n_0}], \tag{B-15}$$

$$\tilde{y}(t) := z(t) \quad \forall t \in [t_{n_0}, t_{n_0} + T(R)]. \tag{B-16}$$

Then $\tilde{y}$ is also a solution to the Cauchy problem (3-7). By the uniqueness of this solution, we have $y = \tilde{y}$ on $D(y) \cap D(\tilde{y})$. However, from (B-14), we have that $D(y) \subsetneqq D(\tilde{y})$, in contradiction with the maximality of y.

Finally, we prove that, if $C(R)$ satisfies (3-13), then, for the maximal solution y to (3-7), we have $D(y) = [s, +\infty)$. We argue by contradiction and therefore assume that the maximal solution y is such that $D(y) = [s, \tau)$ with $\tau < +\infty$. Then (3-12) holds. Let us estimate $\|y(t)\|_{L^2_L}$ when t tends to τ^-. We define the energy $E : [s, \tau) \to [0, +\infty)$ by

$$E(t) := \int_0^L |y(t, x)|^2 \, dx. \tag{B-17}$$

Then $E \in C^0([s, \tau))$ and, in the distribution sense, it satisfies

$$\frac{dE}{dt} \leqslant |u(t, y(t, \cdot))|^2 \leqslant C_B^2(\sqrt{E}). \tag{B-18}$$

(We get such an estimate first in the classical sense for regular initial data and regular boundary conditions $y_x(t, L) = \varphi(t)$ with the related compatibility conditions; the general case then follows from this special case by smoothing the initial data and the boundary conditions, by passing to the limit, and by using the uniqueness of the solution.) From (3-12) and (B-18), we get

$$\frac{1}{2}\int_0^{+\infty} \frac{1}{C_B^2(\sqrt{E})} \, dE < +\infty. \tag{B-19}$$

However the left-hand side of (B-19) is equal to the left-hand side of (3-13). Hence (3-13) and (B-19) are in contradiction. This completes the proof of Theorem 7. □

The proof of Theorem 8 is more difficult. For this proof, we adapt a strategy introduced by Carathéodory to solve ordinary differential equations $\dot{y} = f(t, y)$ when f is not smooth. Roughly speaking it consists in solving $\dot{y} = f(t, y(t - h))$, where h is a positive time-delay, and then letting h tend to 0. Here we do not put the time-delay on y (it does not seem to be possible) but only on the feedback law: $u(t, y(t))$ is replaced by $u(t, y(t - h))$.

Proof of Theorem 8. Let us define $H : [0, +\infty) \to [0, +\infty)$ by

$$H(a) := \int_0^a \frac{1}{(C_B(\sqrt{E}))^2} \, dE = 2\int_0^{\sqrt{a}} \frac{R}{(C_B(R))^2} \, dR. \tag{B-20}$$

From (3-13), we know that H is a bijection from $[0, +\infty)$ into $[0, +\infty)$. We denote by $H^{-1} : [0, +\infty) \to [0, +\infty)$ the inverse of this map.

For a given $y_0 \in L^2(0, L)$ and $s \in \mathbb{R}$, let us prove that there exists a solution y defined on $[s, +\infty)$ to the Cauchy problem (3-7), which also satisfies

$$\|y(t)\|^2_{L^2(0,L)} \leqslant H^{-1}\big(H(\|y(s)\|^2_{L^2_L}) + (t-s)\big) < +\infty \quad \forall t \in [s, +\infty). \tag{B-21}$$

Let $n \in \mathbb{N}^*$. Let us consider the Cauchy system on $[s, s + 1/n]$

$$\begin{cases} y_t + y_{xxx} + y_x + yy_x = 0 & \text{in } (s, s + (1/n)) \times (0, L), \\ y(t, 0) = y(t, L) = 0 & \text{on } (s, s + (1/n)), \\ y_x(t, L) = u(t, y_0) & \text{on } (s, s + (1/n)), \\ y(s, x) = y_0(x) & \text{on } (0, L). \end{cases} \tag{B-22}$$

By Theorem 7 applied with the feedback law $(t, y) \mapsto u(t, y_0)$ (a measurable bounded feedback law which now does not depend on y and therefore satisfies (3-11)), the Cauchy problem (B-22) has one and only one solution y. Let us now consider the Cauchy problem on $[s + (1/n), s + (2/n)]$

$$\begin{cases} y_t + y_{xxx} + y_x + yy_x = 0 & \text{in } (s + (1/n), s + (2/n)) \times (0, L), \\ y(t, 0) = y(t, L) = 0 & \text{on } (s + (1/n), s + (2/n)), \\ y_x(t, L) = u(t, y(t - (1/n))) & \text{on } (s + (1/n), s + (2/n)), \\ y(s, x) = y_0(x) & \text{on } (0, L). \end{cases} \tag{B-23}$$

As for (B-22), this Cauchy problem has one and only one solution, which we still denote by y. We keep going and, by induction on the integer i, define $y \in C^0([s, +\infty); L^2(0, L))$ so that, on $[s + (i/n), s + ((i + 1)/n)]$, $i \in \mathbb{N} \setminus \{0\}$, we have y is the solution to the Cauchy problem

$$\begin{cases} y_t + y_{xxx} + y_x + yy_x = 0 & \text{in } (s + (i/n), s + ((i + 1)/n)) \times (0, L), \\ y(t, 0) = y(t, L) = 0 & \text{on } (s + (i/n), s + ((i + 1)/n)), \\ y_x(t, L) = u(t, y(t - (1/n))) & \text{on } (s + (i/n), s + ((i + 1)/n)), \\ y(s + (i/n)) = y(s + (i/n) - 0) & \text{on } (0, L), \end{cases} \tag{B-24}$$

where, in the last equation, we mean that the initial value, i.e., the value at time $(s + (i/n))$, is the value at time $(s + (i/n))$ of the y defined previously on $[(s + ((i - 1)/n)), s + (i/n)]$.

Again, we let, for $t \in [s, +\infty)$,

$$E(t) := \int_0^L |y(t, x)|^2 \, dx. \tag{B-25}$$

Then $E \in C^0([s, +\infty))$ and, in the distribution sense, it satisfies (compare with (B-18))

$$\frac{dE}{dt} \leqslant |u(t, y_0)|^2 \leqslant C_B^2(\sqrt{E(s)}), \qquad t \in (s, s + (1/n)), \tag{B-26}$$

$$\frac{dE}{dt} \leqslant |u(t, y(t - (1/n))|^2 \leqslant C_B^2(\sqrt{E(t - (1/n))}), \quad t \in \big(s + (i/n), s + ((i + 1)/n)\big), \ i > 0. \tag{B-27}$$

Let $\varphi : [0, +\infty) \to [0, +\infty)$ be the solution of

$$\frac{d\varphi}{dt} = C_B^2\big(\sqrt{\varphi(t)}\big), \quad \varphi(s) = E(s). \tag{B-28}$$

Using (B-26)–(B-28) and simple comparison arguments, one gets

$$E(t) \leqslant \varphi(t) \quad \forall t \in [s, +\infty), \tag{B-29}$$

that is,

$$E(t) \leqslant H^{-1}\big(H(E(s)) + (t-s)\big) \quad \forall t \in [s, +\infty). \tag{B-30}$$

We now want to let $n \to +\infty$. In order to show the dependence on n, we write y^n instead of y. In particular (B-30) becomes

$$\|y^n(t)\|^2_{L^2(0,L)} \leqslant H^{-1}\big(H(\|y_0(s)\|^2_{L^2_L}) + (t-s)\big) \quad \forall t \in [s, +\infty). \tag{B-31}$$

From Lemma 19, (B-31) and the construction of y^n, we get that, for every $T > s$, there exists $M(T) > 0$ such that

$$\|y^n\|_{\mathcal{B}_{s,T}} \leqslant M(T) \quad \forall n \in \mathbb{N}. \tag{B-32}$$

Hence, upon extracting a subsequence of $(y^n)_n$, which we still denote by $(y^n)_n$, there exists

$$y \in L^\infty_{\text{loc}}\big([s, +\infty); L^2(0, L)\big) \cap L^2_{\text{loc}}\big([s, +\infty); H^1(0, L)\big) \tag{B-33}$$

such that, for every $T > s$,

$$y^n \rightharpoonup y \quad \text{in } L^\infty(s, T; L^2(0, L)) \text{ weak} * \text{ as } n \to +\infty, \tag{B-34}$$

$$y^n \rightharpoonup y \quad \text{in } L^2(s, T; H^1(0, L)) \text{ weak as } n \to +\infty. \tag{B-35}$$

Let us define $z^n : [s, s + \infty) \times (0, L) \to \mathbb{R}$ and $\gamma^n : [s, +\infty) \to \mathbb{R}$ by

$$z^n(t) := y_0 \qquad \forall t \in [s, s + (1/n)], \tag{B-36}$$

$$z^n(t) := y^n(t - (1/n)) \quad \forall t \in (s + (1/n), +\infty), \tag{B-37}$$

$$\gamma^n(t) := u(t, z^n) \qquad \forall t \in [s, +\infty). \tag{B-38}$$

Note that y^n is the solution to the Cauchy problem

$$\begin{cases} y^n_t + y^n_{xxx} + y^n_x + y^n y^n_x = 0 & \text{in } (s, +\infty) \times (0, L), \\ y^n(t, 0) = y^n(t, L) = 0 & \text{on } (s, +\infty), \\ y^n_x(t, L) = \gamma^n(t) & \text{on } (s, +\infty), \\ y^n(s, x) = y_0(x) & \text{on } (0, L). \end{cases} \tag{B-39}$$

From (B-32) and the first line of (B-39), we get that

$$\forall T > 0, \quad \left(\frac{d}{dt} y^n\right)_{n \in \mathbb{N}} \text{ is bounded in } L^2(s, s + T; H^{-2}(0, L)). \tag{B-40}$$

From (B-34), (B-35), (B-40) and the Aubin-Lions lemma [Aubin 1963], we get

$$y^n \to y \quad \text{in } L^2(s,T;L^2(0,L)) \text{ as } n \to +\infty \quad \forall T > s. \tag{B-41}$$

From (B-41) we know that, upon extracting a subsequence if necessary, still denoted by $(y^n)_n$,

$$\lim_{n\to+\infty} \|y^n(t) - y(t)\|_{L^2_L} = 0 \quad \text{for almost every } t \in (s,+\infty). \tag{B-42}$$

Letting $n \to +\infty$ in inequality (B-30) for y^n and using (B-42), we get

$$\|y(t)\|^2_{L^2(0,L)} \leqslant H^{-1}\big(H(\|y_0\|^2_{L^2_L}) + (t-s)\big) \quad \text{for almost every } t \in (0,+\infty). \tag{B-43}$$

Note that, for every $T > s$,

$$\begin{aligned}
\|z^n - y\|_{L^2((s,T);L^2_L)} &\leq (1/\sqrt{n})\|y_0\|_{L^2_L} + \big\|y^n(\cdot-(1/n)) - y(\cdot-(1/n))\big\|_{L^2(s+(1/n),T;L^2(0,L))} \\
&\quad + \big\|y(\cdot-(1/n)) - y(\cdot)\big\|_{L^2(s+(1/n),T;L^2(0,L))} + \|y\|_{L^2(s,s+(1/n);L^2(0,L))} \\
&\leq (1/\sqrt{n})\|y_0\|_{L^2_L} + \|y^n - y\|_{L^2(s,T;L^2(0,L))} \\
&\quad + \big\|y(\cdot-(1/n)) - y(\cdot)\big\|_{L^2(s+(1/n),T;L^2(0,L))} + \|y(\cdot)\|_{L^2(s,s+(1/n);L^2(0,L))}.
\end{aligned} \tag{B-44}$$

From (B-36), (B-37), (B-41) and (B-44), we get

$$z^n \to y \quad \text{in } L^2(s,T;L^2(0,L)) \text{ as } n \to +\infty \quad \forall T > s. \tag{B-45}$$

Extracting, if necessary, from the sequence $(z^n)_n$ a subsequence, still denoted by $(z^n)_n$, and using (B-45), we have

$$\lim_{n\to+\infty} \|z^n(t) - y(t)\|_{L^2_L} = 0 \quad \text{for almost every } t \in (s,+\infty). \tag{B-46}$$

From (3-1)–(3-3), (B-32), (B-36), (B-37) and (B-46), extracting a subsequence from the sequence $(\gamma^n)_n$ if necessary, still denoted by $(\gamma^n)_n$, we may assume that

$$\gamma^n \rightharpoonup \gamma(t) := u(t, y(t)) \text{ in } L^\infty(s,T) \text{ weak} * \text{ as } n \to +\infty \quad \forall T > s. \tag{B-47}$$

Let us now check that

$$y \text{ is a solution to the Cauchy problem (3-7).} \tag{B-48}$$

Let $\tau \in [s,+\infty)$ and let $\phi \in C^3([s,\tau]\times[0,L])$ be such that

$$\phi(t,0) = \phi(t,L) = \phi_x(t,0) = 0 \quad \forall t \in [T_1,\tau]. \tag{B-49}$$

From (B-39), one has, for every $n \in \mathbb{N}$,

$$\begin{aligned}
-\int_{T_1}^{\tau}\int_0^L (\phi_t + \phi_x + \phi_{xxx}) y^n \,dx\,dt - \int_{T_1}^{\tau} \gamma^n \phi_x(t,L)\,dt + \int_{T_1}^{\tau}\int_0^L \phi y^n y^n_x \,dx\,dt \\
+ \int_0^L y(\tau,x)\phi(\tau,x)\,dx - \int_0^L y_0\phi(s,x)\,dx = 0.
\end{aligned} \tag{B-50}$$

Let τ be such that

$$\lim_{n\to+\infty} \|y^n(\tau) - y(\tau)\|_{L^2_L} = 0. \tag{B-51}$$

Let us recall that, by (B-42), (B-51) holds for almost every $\tau \in [s, +\infty)$. Using (B-35), (B-41), (B-47), (B-51) and letting $n \to +\infty$ in (B-50), we get

$$-\int_{T_1}^{\tau}\int_0^L (\phi_t + \phi_x + \phi_{xxx}) y \, dx \, dt - \int_{T_1}^{\tau} u(t, y(t))\phi_x(t, L) \, dt + \int_{T_1}^{\tau}\int_0^L \phi y y_x \, dx \, dt + \int_0^L y(\tau, x)\phi(\tau, x) \, dx - \int_0^L y_0 \phi(s, x) \, dx = 0. \tag{B-52}$$

Thus y is a solution to (2-1), with $T_1 := s$, T_2 arbitrary in $(s, +\infty)$, $\tilde{h} := -yy_x \in L^1([s, T_2]; L^2(0, L))$ and $h = u(\cdot, y(\cdot)) \in L^2(s, T_2)$. Let us emphasize that, by Lemma 3, it also implies that $y \in \mathcal{B}_{s,T}$ for every $T \in (s, +\infty)$. This concludes the proof of (B-48) and of Theorem 8. □

Appendix C: Proof of Proposition 12

Let us first recall that Proposition 12 is due to Eduardo Cerpa if one requires only u to be in $L^2(0, T)$ instead of being in $H^1(0, T)$; see [Cerpa 2007, Proposition 3.1] and [Cerpa and Crépeau 2009a, Proposition 3.1]. In his proof, he uses Lemma 11, the controllability in H with controls $u \in L^2$. Actually, the only place in his proof where the controllability in H is used is on page 887 of [Cerpa 2007] for the construction of α_1, where, with the notations of that paper $\Re(y_\lambda)$, $\Im(y_\lambda) \in H$. We notice that $\Re(y_\lambda)$, $\Im(y_\lambda)$ share more regularity and better boundary conditions. Indeed, one has

$$\begin{cases} \lambda y_\lambda + y'_\lambda + y'''_\lambda = 0, \\ y_\lambda(0) = y_\lambda(L) = 0, \end{cases}$$

which implies that

$$\Re(y_\lambda), \Im(y_\lambda) \in \mathcal{H}^3,$$

where

$$\mathcal{H}^3 := H \cap \big\{\omega \in H^3(0, L) : \omega(0) = \omega(L) = 0\big\}. \tag{C-1}$$

In order to adapt Cerpa's proof in the framework of $u \in H^1(0, T)$, it is sufficient to prove the following controllability result in $\mathcal{H}^3$ with control $u \in H^1(0, T)$.

Proposition 20. *For every y_0, $y_1 \in \mathcal{H}^3$ and for every $T > 0$, there exists a control $u \in H^1(0, T)$ such that the solution $y \in \mathcal{B}$ to the Cauchy problem*

$$\begin{cases} y_t + y_{xxx} + y_x = 0, \\ y(t, 0) = y(t, L) = 0, \\ y_x(t, L) = u(t), \\ y(0, \cdot) = y_0 \end{cases}$$

satisfies $y(T, \cdot) = y_1$.

The proof of Proposition 12 is the same as the one of [Cerpa 2007, Proposition 3.1], with the only difference that one uses Proposition 20 instead of Lemma 11.

Proof of Proposition 20. Let us first point out that 0 is not an eigenvalue of the operator $\mathcal{A}$. Indeed this follows from property $(\mathcal{P}_2)$, (1-5) and (1-6). Using Lemma 11 and [Tucsnak and Weiss 2009, Proposition 10.3.4] with $\beta = 0$, it suffices to check that

$$\text{for every } f \in H, \text{ there exists } y \in \mathcal{H}^3 \text{ such that } -y_{xxx} - y_x = f. \tag{C-2}$$

Let $f \in H$. We know that there exists $y \in H^3(0, L)$ such that

$$-y_{xxx} - y_x = f, \tag{C-3}$$

$$y(0) = y(L) = y_x(L) = 0. \tag{C-4}$$

Simple integrations by parts, together with (4-11), (4-12), (C-3) and (C-4), show that, with $\varphi := \varphi_1 + i\varphi_2$,

$$0 = \int_0^L f\varphi \, dx = \int_0^L (-y_{xxx} - y_x)\varphi \, dx = \int_0^L y(\varphi_{xxx} + \varphi_x) \, dx = i\frac{2\pi}{p} \int_0^L y\varphi \, dx, \tag{C-5}$$

which, together with (C-4), implies that $y \in \mathcal{H}^3$. This concludes the proof of (C-2) as well as the proof of Proposition 20 and of Proposition 12. □

Acknowledgments

We thank Jixun Chu, Ludovick Gagnon, Peipei Shang and Shuxia Tang for useful comments on a preliminary version of this article.

References

[Aubin 1963] J.-P. Aubin, "Un théorème de compacité", *C. R. Acad. Sci. Paris* **256** (1963), 5042–5044. MR Zbl

[Beauchard 2005] K. Beauchard, "Local controllability of a 1-D Schrödinger equation", *J. Math. Pures Appl.* (9) **84**:7 (2005), 851–956. MR Zbl

[Beauchard and Coron 2006] K. Beauchard and J.-M. Coron, "Controllability of a quantum particle in a moving potential well", *J. Funct. Anal.* **232**:2 (2006), 328–389. MR Zbl

[Beauchard and Morancey 2014] K. Beauchard and M. Morancey, "Local controllability of 1D Schrödinger equations with bilinear control and minimal time", *Math. Control Relat. Fields* **4**:2 (2014), 125–160. MR Zbl

[Bona and Smith 1975] J. L. Bona and R. Smith, "The initial-value problem for the Korteweg–de Vries equation", *Philos. Trans. Roy. Soc. London Ser. A* **278**:1287 (1975), 555–601. MR Zbl

[Bona et al. 2003] J. L. Bona, S. M. Sun, and B.-Y. Zhang, "A nonhomogeneous boundary-value problem for the Korteweg–de Vries equation posed on a finite domain", *Comm. Partial Differential Equations* **28**:7-8 (2003), 1391–1436. MR Zbl

[Bona et al. 2009] J. L. Bona, S. M. Sun, and B.-Y. Zhang, "A non-homogeneous boundary-value problem for the Korteweg–de Vries equation posed on a finite domain, II", *J. Differential Equations* **247**:9 (2009), 2558–2596. MR Zbl

[Boussinesq 1877] J. Boussinesq, *Essai sur la théorie des eaux courantes*, Mémoires présentés par divers savants à l'Acad. des Sci. Inst. Nat. France **23**, 1877. JFM

[Brockett 1983] R. W. Brockett, "Asymptotic stability and feedback stabilization", pp. 181–191 in *Differential geometric control theory* (Houghton, MI, 1982), edited by R. W. Brockett and R. S. Millman, Progr. Math. **27**, Birkhäuser, Boston, 1983. MR Zbl

[Capistrano-Filho et al. 2015] R. A. Capistrano-Filho, A. F. Pazoto, and L. Rosier, "Internal controllability of the Korteweg–de Vries equation on a bounded domain", *ESAIM Control Optim. Calc. Var.* **21**:4 (2015), 1076–1107. MR Zbl

[Cerpa 2007] E. Cerpa, "Exact controllability of a nonlinear Korteweg–de Vries equation on a critical spatial domain", *SIAM J. Control Optim.* **46**:3 (2007), 877–899. MR Zbl

[Cerpa 2014] E. Cerpa, "Control of a Korteweg–de Vries equation: a tutorial", *Math. Control Relat. Fields* **4**:1 (2014), 45–99. MR Zbl

[Cerpa and Coron 2013] E. Cerpa and J.-M. Coron, "Rapid stabilization for a Korteweg–de Vries equation from the left Dirichlet boundary condition", *IEEE Trans. Automat. Control* **58**:7 (2013), 1688–1695. MR

[Cerpa and Crépeau 2009a] E. Cerpa and E. Crépeau, "Boundary controllability for the nonlinear Korteweg–de Vries equation on any critical domain", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **26**:2 (2009), 457–475. MR Zbl

[Cerpa and Crépeau 2009b] E. Cerpa and E. Crépeau, "Rapid exponential stabilization for a linear Korteweg–de Vries equation", *Discrete Contin. Dyn. Syst. Ser. B* **11**:3 (2009), 655–668. MR Zbl

[Chowdhury and Ervedoza 2017] S. Chowdhury and S. Ervedoza, "Open loop stabilization of incompressible Navier–Stokes equations in a 2d channel with a normal control using power series expansion", preprint, 2017.

[Chu et al. 2015] J. Chu, J.-M. Coron, and P. Shang, "Asymptotic stability of a nonlinear Korteweg–de Vries equation with critical lengths", *J. Differential Equations* **259**:8 (2015), 4045–4085. MR Zbl

[Constantin and Saut 1988] P. Constantin and J.-C. Saut, "Local smoothing properties of dispersive equations", *J. Amer. Math. Soc.* **1**:2 (1988), 413–439. MR Zbl

[Coron 1990] J.-M. Coron, "A necessary condition for feedback stabilization", *Systems Control Lett.* **14**:3 (1990), 227–232. MR Zbl

[Coron 1995] J.-M. Coron, "On the stabilization in finite time of locally controllable systems by means of continuous time-varying feedback law", *SIAM J. Control Optim.* **33**:3 (1995), 804–833. MR Zbl

[Coron 2007] J.-M. Coron, *Control and nonlinearity*, Mathematical Surveys and Monographs **136**, American Mathematical Society, Providence, RI, 2007. MR Zbl

[Coron and Crépeau 2004] J.-M. Coron and E. Crépeau, "Exact boundary controllability of a nonlinear KdV equation with critical lengths", *J. Eur. Math. Soc.* **6**:3 (2004), 367–398. MR Zbl

[Coron and Lü 2014] J.-M. Coron and Q. Lü, "Local rapid stabilization for a Korteweg–de Vries equation with a Neumann boundary control on the right", *J. Math. Pures Appl.* (9) **102**:6 (2014), 1080–1120. MR Zbl

[Coron and Lü 2015] J.-M. Coron and Q. Lü, "Fredholm transform and local rapid stabilization for a Kuramoto–Sivashinsky equation", *J. Differential Equations* **259**:8 (2015), 3683–3729. MR Zbl

[Coron and Rivas 2016] J.-M. Coron and I. Rivas, "Quadratic approximation and time-varying feedback laws", preprint, 2016, available at https://hal.archives-ouvertes.fr/hal-01402747/.

[Coron and Rosier 1994] J.-M. Coron and L. Rosier, "A relation between continuous time-varying and discontinuous feedback stabilization", *J. Math. Systems Estim. Control* **4**:1 (1994), 67–84. MR Zbl

[Coron et al. 2016] J.-M. Coron, L. Gagnon, and M. Morancey, "Rapid stabilization of a linearized bilinear 1-D Schrödinger equation", preprint, 2016, available at https://hal.archives-ouvertes.fr/hal-01408179.

[Craig et al. 1992] W. Craig, T. Kappeler, and W. Strauss, "Gain of regularity for equations of KdV type", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **9**:2 (1992), 147–186. MR Zbl

[Doronin and Natali 2014] G. G. Doronin and F. M. Natali, "An example of non-decreasing solution for the KdV equation posed on a bounded interval", *C. R. Math. Acad. Sci. Paris* **352**:5 (2014), 421–424. MR Zbl

[Gagnon 2016] L. Gagnon, "Lagrangian controllability of the 1-dimensional Korteweg–de Vries equation", *SIAM J. Control Optim.* **54**:6 (2016), 3152–3173. MR Zbl

[Gardner et al. 1967] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, "Method for solving the Korteweg–de Vries equation", *Phys. Rev. Lett.* **19** (1967), 1095–1097. Zbl

[Glass and Guerrero 2010] O. Glass and S. Guerrero, "Controllability of the Korteweg–de Vries equation from the right Dirichlet boundary condition", *Systems Control Lett.* **59**:7 (2010), 390–395. MR Zbl

[Goubet and Shen 2007] O. Goubet and J. Shen, "On the dual Petrov–Galerkin formulation of the KdV equation on a finite interval", *Adv. Differential Equations* **12**:2 (2007), 221–239. MR Zbl

[Jia and Zhang 2012] C. Jia and B.-Y. Zhang, "Boundary stabilization of the Korteweg–de Vries equation and the Korteweg–de Vries–Burgers equation", *Acta Appl. Math.* **118** (2012), 25–47. MR Zbl

[Komornik 1997] V. Komornik, "Rapid boundary stabilization of linear distributed systems", *SIAM J. Control Optim.* **35**:5 (1997), 1591–1613. MR Zbl

[Korteweg and de Vries 1895] D. J. Korteweg and G. de Vries, "On the change of form of long waves advancing in a rectangular canal, and on a new type of long stationary waves", *Philos. Mag.* (5) **39**:240 (1895), 422–443. MR Zbl

[Krstic and Smyshlyaev 2008] M. Krstic and A. Smyshlyaev, *Boundary control of PDEs*, Advances in Design and Control **16**, Society for Industrial and Applied Mathematics, Philadelphia, 2008. MR Zbl

[Laurent et al. 2010] C. Laurent, L. Rosier, and B.-Y. Zhang, "Control and stabilization of the Korteweg–de Vries equation on a periodic domain", *Comm. Partial Differential Equations* **35**:4 (2010), 707–744. MR Zbl

[Massarolo et al. 2007] C. P. Massarolo, G. P. Menzala, and A. F. Pazoto, "On the uniform decay for the Korteweg–de Vries equation with weak damping", *Math. Methods Appl. Sci.* **30**:12 (2007), 1419–1435. MR Zbl

[Morancey 2014] M. Morancey, "Simultaneous local exact controllability of 1D bilinear Schrödinger equations", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **31**:3 (2014), 501–529. MR Zbl

[Murray 1978] A. C. Murray, "Solutions of the Korteweg–de Vries equation from irregular data", *Duke Math. J.* **45**:1 (1978), 149–181. MR Zbl

[Pazoto 2005] A. F. Pazoto, "Unique continuation and decay for the Korteweg–de Vries equation with localized damping", *ESAIM Control Optim. Calc. Var.* **11**:3 (2005), 473–486. MR Zbl

[Perla Menzala et al. 2002] G. Perla Menzala, C. F. Vasconcellos, and E. Zuazua, "Stabilization of the Korteweg–de Vries equation with localized damping", *Quart. Appl. Math.* **60**:1 (2002), 111–129. MR Zbl

[Rivas et al. 2011] I. Rivas, M. Usman, and B.-Y. Zhang, "Global well-posedness and asymptotic behavior of a class of initial-boundary-value problem of the Korteweg–de Vries equation on a finite domain", *Math. Control Relat. Fields* **1**:1 (2011), 61–81. MR Zbl

[Rosier 1997] L. Rosier, "Exact boundary controllability for the Korteweg–de Vries equation on a bounded domain", *ESAIM Control Optim. Calc. Var.* **2** (1997), 33–55. MR Zbl

[Rosier 2004] L. Rosier, "Control of the surface of a fluid by a wavemaker", *ESAIM Control Optim. Calc. Var.* **10**:3 (2004), 346–380. MR Zbl

[Rosier and Zhang 2006] L. Rosier and B.-Y. Zhang, "Global stabilization of the generalized Korteweg–de Vries equation posed on a finite domain", *SIAM J. Control Optim.* **45**:3 (2006), 927–956. MR Zbl

[Rosier and Zhang 2009] L. Rosier and B.-Y. Zhang, "Control and stabilization of the Korteweg–de Vries equation: recent progresses", *J. Syst. Sci. Complex.* **22**:4 (2009), 647–682. MR Zbl

[Russell and Zhang 1995] D. L. Russell and B. Y. Zhang, "Smoothing and decay properties of solutions of the Korteweg–de Vries equation on a periodic domain with point dissipation", *J. Math. Anal. Appl.* **190**:2 (1995), 449–488. MR Zbl

[Russell and Zhang 1996] D. L. Russell and B. Y. Zhang, "Exact controllability and stabilizability of the Korteweg–de Vries equation", *Trans. Amer. Math. Soc.* **348**:9 (1996), 3643–3672. MR Zbl

[Slemrod 1974] M. Slemrod, "A note on complete controllability and stabilizability for linear control systems in Hilbert space", *SIAM J. Control* **12** (1974), 500–508. MR Zbl

[Tang et al. 2016] S. Tang, J. Chu, P. Shang, and J.-M. Coron, "Asymptotic stability of a Korteweg–de Vries equation with a two-dimensional center manifold", *Adv. Nonlinear Anal.* (online publication October 2016).

[Temam 1969] R. Temam, "Sur un problème non linéaire", *J. Math. Pures Appl.* (9) **48** (1969), 159–172. MR Zbl

[Tucsnak and Weiss 2009] M. Tucsnak and G. Weiss, *Observation and control for operator semigroups*, Birkhäuser, Basel, 2009. MR Zbl

[Urquiza 2005] J. M. Urquiza, "Rapid exponential feedback stabilization with unbounded control operators", *SIAM J. Control Optim.* **43**:6 (2005), 2233–2244. MR Zbl

[Whitham 1974] G. B. Whitham, *Linear and nonlinear waves*, Wiley, New York, 1974. MR Zbl

[Zhang 1999] B.-Y. Zhang, "Exact boundary controllability of the Korteweg–de Vries equation", *SIAM J. Control Optim.* **37**:2 (1999), 543–565. MR Zbl

Received 2 May 2016. Revised 29 Sep 2016. Accepted 7 Mar 2017.

JEAN-MICHEL CORON: coron@ann.jussieu.fr
Université Pierre et Marie Curie - Paris 6, UMR 7598 Laboratoire Jacques-Louis Lions, 75005 Paris, France
and
ETH Zürich, Institute for Theoretical Studies, 8092 Zürich, Switzerland

IVONNE RIVAS: ivonne.rivas@correounivalle.edu.co
Universidad del Valle, Departamento de Matemáticas, Cali, AA 25360, Colombia

SHENGQUAN XIANG: shengquan.xiang@ens.fr
Université Pierre et Marie Curie - Paris 6, UMR 7598 Laboratoire Jacques-Louis Lions, 75005 Paris, France
and
ETH Zürich, Institute for Theoretical Studies and Forschungsinstitut für Mathematik, 8092 Zürich, Switzerland

ANALYSIS AND PDE
Vol. 10, No. 5, 2017
dx.doi.org/10.2140/apde.2017.10.1123

ON THE GROWTH OF SOBOLEV NORMS FOR NLS ON 2- AND 3-DIMENSIONAL MANIFOLDS

FABRICE PLANCHON, NIKOLAY TZVETKOV AND NICOLA VISCIGLIA

Using suitable modified energies, we study higher-order Sobolev norms' growth in time for the nonlinear Schrödinger equation (NLS) on a generic 2- or 3-dimensional compact manifold. In two dimensions, we extend earlier results that dealt only with cubic nonlinearities, and get polynomial-in-time bounds for any higher-order nonlinearities. In three dimensions, we prove that solutions to the cubic NLS grow at most exponentially, while for the subcubic NLS we get polynomial bounds on the growth of the H^2 norm.

1. Introduction

We are interested in long-time qualitative properties of solutions to the family of nonlinear Schrödinger equations

$$\begin{cases} i\,\partial_t u + \Delta_g u = |u|^{p-1}u, & (t,x) \in \mathbb{R} \times M^d, \\ u(0,x) = \varphi \in H^m(M^d), \end{cases} \tag{1}$$

where Δ_g is the Laplace–Beltrami operator associated with a d-dimensional compact Riemannian manifold (M^d, g) and $H^m(M^d)$, the standard Sobolev space associated to Δ_g, where $m \in \mathbb{N}$ with $m \geq 2$. More specifically we are interested in the analysis of the possible growth of higher-order Sobolev norms for large times, namely the behavior of the quantity $\|u(t,x)\|_{H^m(M^d)}$ for $m \geq 2$ and $t \gg 1$.

This issue of growth of higher-order Sobolev norms has garnered a lot of attention in recent years, mainly because of its connection with the so-called *weak wave turbulence*, e.g., a cascade of energy from low to high frequencies. In fact two main issues have been extensively studied in the literature: the first one concerns a priori bounds on how fast higher-order Sobolev norms can grow along the flow associated with Hamiltonian PDEs (see [Bourgain 1993; 1996; 1999a; 1999b; Colliander et al. 2012; Delort 2014; Sohinger 2011a; 2011b; 2012; Staffilani 1997; Thirouin 2017; Zhong 2008]); the second one concerns the existence of global solutions whose higher-order Sobolev norms are unbounded (see [Colliander et al. 2010; Gérard and Grellier 2016; 2015; Guardia 2014; Guardia et al. 2016; Guardia and Kaloshin 2015; Hani 2014; Hani et al. 2015; Haus and Procesi 2015; Xu 2015]).

Here, we aim at dealing with the first problem, namely to provide a priori bounds on the growth of higher-order Sobolev norms, or equivalently to understand how fast the dynamical system under consideration can move energy from the low frequencies to the high frequencies.

Planchon was partially supported by ANR grant GEODISP, ERC grant SCAPDE and ERC grant BLOWDISOL, Tzvetkov was partially supported by the ERC grant DISPEQ, and Visciglia was supported by the grant PRA 2016 Problemi di Evoluzione: Studio Qualitativo e Comportamento Asintotico.
MSC2010: 35Q55.
Keywords: growth of Sobolev norms, NLS on compact manifolds.

First of all we point out that solutions to (1) enjoy so-called mass and energy conservation laws:

$$\int_{M^d} |u(t,x)|^2 \,\mathrm{dvol}_g = \int_{M^d} |\varphi(x)|^2 \,\mathrm{dvol}_g,$$
$$\int_{M^d} \Big(|\nabla_g u(t,x)|_g^2 + \frac{1}{p+1} |u(t,x)|^{p+1} \Big) \,\mathrm{dvol}_g = \int_{M^d} \Big(|\nabla_g \varphi(x)|_g^2 + \frac{1}{p+1} |\varphi(x)|^{p+1} \Big) \,\mathrm{dvol}_g,$$

where ∇_g and $|\cdot|_g$ are respectively the gradient and the norm associated with the metric g, and $|\cdot|$ denotes the modulus of any complex number. These conservation laws immediately imply that

$$\sup_{\mathbb{R}} \|u(t,x)\|_{H^1(M^d)} < \infty, \tag{2}$$

and therefore the growth in time of H^m norms is only of interest for $m \geq 2$.

In the sequel, with notation as above, we shall be interested in the following cases:

- $(d, p) = (2, 2n+1)$ with $n \in \mathbb{N}$, $n \geq 1$ (2-dimensional manifold and odd integer nonlinearity),
- $(d, p) = (3, 3)$ (3-dimensional manifold and cubic nonlinearity),
- $(d, p) = (3, p)$ with $2 < p < 3$ (3-dimensional manifold and subcubic nonlinearity).

In those settings, existence of local solutions follows by classical arguments, provided one assumes the initial datum to be H^2. On the other hand, following [Burq et al. 2004], one can establish local (and hence global) Cauchy theory in H^1 for generic nonlinear potentials in the 2-dimensional case, as well as local (and global) Cauchy theory in $H^{1+\epsilon}$ for the cubic and subcubic NLS in the 3-dimensional case (see [Burq et al. 2003; 2004]). From now on and for the sake of simplicity, we shall assume existence and uniqueness of a global solution, and focus on estimating the growth of higher-order Sobolev norms. However, we point out that our argument not only provides polynomial bounds of such growth, but also yields an alternative proof of global existence in three dimensions.

We will use as a basic tool (in fact, as a black box) available Strichartz estimates on manifolds (see [Burq et al. 2004; Staffilani and Tataru 2002]) together with the introduction of suitable *modified energies*, which is the main new ingredient in this context. For this reason we will not discuss further the issue of global existence, which is indeed guaranteed by aforementioned previous results.

We first start with the 2-dimensional case. It is worth mentioning that, to the authors' knowledge, no results were available in the literature about growth of higher-order Sobolev norms for NLS with higher than cubic nonlinearities, although one may reasonably believe that this problem could be addressed, at least in two dimensions, by adapting the strategy pioneered by Bourgain (see for instance [Zhong 2008]). Nevertheless as a warm up we show how this problem can be handled by a completely different strategy, based on the introduction of suitable *modified energies*: its benefit relies on a clear decoupling between higher-order energy estimates relying on clever integration by parts and the (deep) input provided by dispersive estimates of Strichartz type. Moreover by using modified energies, one can deal as well with generic nonlinear potential $V(|u|^2)$ rather than $|u|^{p-1}$, where V may not necessarily be a pure power (see also Remark 1.7 below).

We emphasize that modified energies have proved useful in different contexts (see, for instance, [Chiron and Rousset 2009; Hunter et al. 2015; Koch and Tataru 2016; Kwon 2008; Ozawa and Visciglia 2016;

Raphaël and Szeftel 2009; Tsutsumi 1989]), but the present work seems to provide the first example where they are combined with dispersive bounds in order to get results on the growth of higher-order Sobolev norms.

We underline that our argument, being essentially based on integration by parts, relies on the time derivative of suitable higher-order energies $\mathcal{E}_m$, whose leading term is essentially the norm $\|u(t,x)\|^2_{H^m}$. In fact, for $m = 2k$ an even integer, one should think of $\|\partial_t^k u(t,x)\|^2_{L^2}$ as a good prototype of modified energy, up to lower-order terms. In other words, one should think of replacing Δ_g by ∂_t rather than the other way around when using the equation satisfied by u.

A direct consequence of this privileged use of ∂_t is that in our approach the geometry of the manifold is not directly involved in the computation, and integration by parts in the space variables, when required, is performed thanks to the following elementary identity, available on any generic manifold:

$$\Delta_g(fh) = h\Delta_g f + 2(\nabla_g f, \nabla_g h)_g + f\Delta_g h.$$

We also underline that the aforementioned energy $\mathcal{E}_m$ is not preserved along the flow; however, by computing its time derivative along solutions, we may estimate the resulting space-time integral taking advantage of dispersive bounds, namely Strichartz estimates with loss, which are available on a generic manifold (or better ones when available).

In order to state our result in two dimensions, we recall Strichartz estimates with loss:

$$\|e^{it\Delta_g}\varphi\|_{L^4((0,1)\times M^2)} \lesssim \|\varphi\|_{H^{s_0}(M^2)}. \tag{3}$$

It is well known that estimate (3) holds on $\mathbb{T}^2$ for any $s_0 > 0$ (see [Bourgain 1993; 1999a]) and on the sphere $\mathbb{S}^2$ for any $s_0 > \frac{1}{8}$ (see [Burq et al. 2004]). We can now state our first result, where we assume (3) to be satisfied for some s_0 in the range $\left[0, \frac{1}{4}\right]$. We recall that the existence of such an s_0 is guaranteed on every compact manifold M^2 by [Burq et al. 2004].

Theorem 1.1. *For every $\epsilon > 0$, $m \in \mathbb{N}$ with $m \geq 2$ and for every solution $u(t,x) \in \mathcal{C}_t(H^m(M^2))$ to* (1)*, where $d = 2$ and $p = 2n+1$ for $n \geq 1$, we get*

$$\sup_{(0,T)} \|u(t,x)\|_{H^m(M^2)} \leq C(\max\{1,T\})^{\frac{m-1}{1-2s_0}+\epsilon}, \tag{4}$$

where $C = C(\epsilon, m, \|\varphi\|_{H^m}) > 0$ and $s_0 \in \left[0, \frac{1}{4}\right]$ is given in (3)*.*

Notice that bounds from Theorem 1.1 also apply to solutions of NLS on $\mathbb{T}$. In fact the dynamics of NLS on $\mathbb{T}$ is a subset of the dynamics on $\mathbb{T}^2$, and this framework is covered by Theorem 1.1, where we can choose $s_0 = 0$. In particular, Theorem 1.1 recovers results from [Colliander et al. 2012] for solutions to NLS on $\mathbb{T}$ with $p > 5$. Notice that paper obtains a better $\mathbb{T}^{\frac{m-1}{2}+\epsilon}$ growth for $p = 5$ by implementing a normal-form method. We will address this better growth for all $p > 5$ with a suitable modification of our argument in a later work.

Remark 1.2. We underline that the main point in order to establish Theorem 1.1 is the following bound: for all $\tau \in (0,1)$, $\epsilon > 0$,

$$\|u(\tau)\|^2_{H^m(M^2)} - \|u(0)\|^2_{H^m(M^2)} \lesssim \sqrt{\tau}\,\|u\|^{\frac{2m-3+2s_0}{m-1}+\epsilon}_{L^\infty((0,\tau);H^m(M^2))} + \|u\|^{\frac{2m-4}{m-1}+\epsilon}_{L^\infty((0,\tau);H^m(M^2))}. \tag{5}$$

Once this bound is established, a classical argument (which in turn requires the local well-posedness of the Cauchy problem in the energy space H^1) leads to the polynomial growth. More specifically notice that the exponent $(m-1)/(1-2s_0)+\epsilon$ (which appears in the right-hand side of (4)) can be computed as the quantity $\frac{1}{2\gamma}$, where $2-2\gamma = (2m-3+2s_0)/(m-1)+\epsilon$ is the power of the first term in the right-hand side of (5). Next we choose $\tau = \tau(\|u(0)\|_{H^1})$ to be the time of existence provided by the H^1 local Cauchy theory. Then (5) gives

$$\|u(t+\tau)\|^2_{H^m(M^2)} \le \|u(t)\|^2_{H^m(M^2)} + C\big(\|u\|^{2-2\gamma}_{L^\infty((t,t+\tau);H^m(M^2))} + 1\big).$$

As a byproduct of the local existence theory in H^1, and conservation of the energy, we get

$$\|u(t+\tau)\|^2_{H^m(M^2)} \le \|u(t)\|^2_{H^m(M^2)} + C\big(\|u(t)\|^{2-2\gamma}_{H^m(M^2)} + 1\big).$$

Therefore the sequence $\alpha_n = 1 + \|u(n\tau)\|^2_{H^m(M^2)}$ satisfies $\alpha_{n+1} \le \alpha_n + C\alpha_n^{1-\gamma}$, which in turn implies $\alpha_n \lesssim n^{1/\gamma}$, leading to (4) by induction on n.

Next we present our result on the growth of higher-order Sobolev norms for the cubic NLS on a generic 3-dimensional compact manifold M^3. We recall that, following [Burq et al. 2004], the Cauchy problem is globally well-posed for every initial data $\varphi \in H^{1+\epsilon_0}(M^3)$, and that, following the crucial use of logarithmic Sobolev type inequalities, one can get the following double exponential bound,

$$\sup_{(0,T)} \|u(t,x)\|_{H^m(M^3)} \le C\exp(\exp(CT)).$$

Our main contribution is an improvement on the bound above; indeed, we will replace the double exponential with a single one. It should be emphasized that, in the 3-dimensional case, it is at best unclear to us how Bourgain's original argument and derivatives thereof could be used in order to get Theorem 1.3. More specifically, in three dimensions our use of modified energies appears to be a key tool in order to eliminate one of the two exponentials.

Theorem 1.3. *For every $m \in \mathbb{N}$ with $m \ge 2$ and for every solution $u(t,x) \in \mathcal{C}_t(H^m(M^3))$ to* (1), *where $(d,p) = (3,3)$, we have*

$$\sup_{(0,T)} \|u(t,x)\|_{H^m(M^3)} \le C\exp(CT),$$

where $C = C(m, \|\varphi\|_{H^m}) > 0$.

Remark 1.4. The proof of Theorem 1.3 follows by a straightforward iteration once the following bound is established: for all $\tau \in (0,1)$,

$$\|u(\tau)\|^2_{H^m(M^3)} - \|u(0)\|^2_{H^m(M^3)} \lesssim \tau\|u\|^2_{L^\infty((0,\tau);H^m(M^3))} + \|u\|^\gamma_{L^\infty((0,\tau);H^m(M^3))}, \tag{6}$$

where $\gamma \in (0,2)$ is a suitable number and the implicit constant depends only on the energy of u. Indeed, using (6) for τ small enough and the fact that $\gamma < 2$, we get the bound

$$\|u\|^2_{L^\infty((0,\tau);H^m(M^3))} \le 2\|u(0)\|^2_{H^m(M^3)} + C.$$

Therefore the sequence $\alpha_n = \|u(n\tau)\|^2_{H^m(M^3)}$ satisfies $\alpha_{n+1} \le 2\alpha_n + C$, which implies the claimed exponential bound.

Remark 1.5. Notice that in Theorems 1.1 and 1.3 we provide bounds on the growth of the H^m Sobolev norm for initial data of regularity H^m, for a given, odd or even, integer m. We point out that most of the paper will be devoted to the case of even integers. In the last section we sketch how to adapt the argument to odd integers. Of course, if we assume the initial datum φ to be H^{m+1}, then the growth of H^m, with m odd, can be obtained by interpolation between the growth of the two norms H^{m+1} and H^{m-1}, with $m-1$ and $m+1$ even. Hence if the initial datum is smooth enough, it is not necessary to deal separately with the case m odd. However the situation is more delicate since we assume only the regularity H^m (with m odd) on the initial datum.

Finally, we end our presentation with a result dealing with NLS on a 3-dimensional compact manifold M^3 with subcubic nonlinearity, establishing polynomial growth for the H^2 Sobolev norm. It makes no sense to consider higher-order Sobolev norms, given that the nonlinearity is not smooth enough to guarantee that regularity H^m, with $m > 2$, is preserved along the evolution.

Nevertheless we emphasize that the next result appears to be the first one available in the literature about polynomial growth of any Sobolev norms above the energy, on a generic 3-dimensional compact manifold.

Theorem 1.6. *For every solution $u(t,x) \in \mathcal{C}_t(H^2(M^3))$ to (1) with $d = 3$ and $p \in (2,3)$ we have*

$$\sup_{(0,T)} \|u(t,x)\|_{H^2(M^3)} \leq C(\max\{1,T\})^{\frac{4}{3-p}},$$

where $C = C(\|\varphi\|_{H^2}) > 0$.

Remark 1.7. The proof of Theorem 1.6 follows once the following local bound is established: for all $\tau \in (0,1)$,

$$\|u(\tau)\|^2_{H^2(M^3)} - \|u(0)\|^2_{H^2(M^3)} \lesssim \tau \|u\|^{\frac{p+5}{4}}_{L^\infty((0,\tau);H^2(M^3))} + \|u\|^{\gamma}_{L^\infty((0,\tau);H^2(M^3))} \tag{7}$$

for some $\gamma \in \left(0, \frac{p+5}{4}\right)$. In order to conclude the polynomial growth from (7), we can combine Remarks 1.2 and 1.4. In fact arguing as in Remark 1.4 we get

$$\|u\|^2_{L^\infty((0,\tau);H^m(M^3))} \leq 2\|u(0)\|^2_{H^m(M^3)} + C.$$

Once this bound is established, the polynomial growth follows by using (7) and arguing exactly as in Remark 1.2.

Remark 1.8. Following our approach to proving (7), there is no need to restrict oneself to pure power nonlinearities. In particular, polynomial growth for solutions to NLS on generic 3-dimensional compact manifolds could be established for general higher-order Sobolev norms (namely H^m with $m \geq 2$), provided the subcubic nonlinearity is suitably regularized in order to guarantee that H^m regularity is preserved along the flow. Nevertheless, for the sake of simplicity we elected not to deal with the full generality in this work.

2. Linear Strichartz estimates

Strichartz estimates on M^2. In the sequel we shall make use without any further comment of the following Strichartz estimate, which was already recalled in the Introduction:

$$\|e^{it\Delta_g}\varphi\|_{L^4((0,1)\times M^2)} \lesssim \|\varphi\|_{H^{s_0}(M^2)}. \tag{8}$$

By using Duhamel formula we also have at our disposal an inhomogeneous estimate that we state as an independent proposition.

Proposition 2.1. *Let $v(t,x)$ be solution to*

$$\begin{cases} i\,\partial_t v + \Delta_g v = F, \quad (t,x) \in \mathbb{R}\times M^2, \\ u(0,x) = \varphi \in H^{s_0}(M^2). \end{cases}$$

Then we have, for $T \in (0,1)$,

$$\|v\|_{L^4((0,T)\times M^2)} \lesssim \|\varphi\|_{H^{s_0}(M^2)} + T\|F\|_{L^\infty((0,T);H^{s_0}(M^2))}. \tag{9}$$

Strichartz estimates on M^3. In the proofs of Theorems 1.3 and 1.6 we shall make use of the following suitable version of the endpoint Strichartz estimate:

Proposition 2.2. *Let $v(t,x)$ be solution to*

$$i\,\partial_t v + \Delta_g v = F, \quad (t,x) \in \mathbb{R}\times M^3.$$

Then we have, for $\tau \in (0,1)$,

$$\|v\|_{L^2((0,\tau);L^6(M^3))} \lesssim_\epsilon \|v\|_{L^\infty((0,\tau);H^\epsilon(M^3))} + \|v\|_{L^2((0,\tau);H^{1/2}(M^3))} + \|F\|_{L^2((0,\tau);L^{6/5}(M^3))}. \tag{10}$$

Notice that the above estimate may look somewhat unusual compared with the classical version of Strichartz estimates, where on the right-hand side one expects a norm involving the initial datum $v(0,x)$ and another norm involving the forcing term $F(t,x)$.

Nevertheless we underline that in the case $F = 0$, the estimate above reduces to the usual Strichartz estimate with loss of half of a derivative (see [Burq et al. 2004; Staffilani and Tataru 2002]). On the other hand, the main point of (10) is that no derivative losses occur on the forcing term $F(t,x)$ when this term is not identically zero, and the loss of derivative indeed occurs only for the solution $v(t,x)$ on the right-hand side. Estimates in this spirit are also of crucial importance in the low regularity well-posedness theory for quasilinear dispersive PDEs (see, e.g., [Koch and Tzvetkov 2003]). We emphasize that the estimate (10) comes from the following spectrally localized version (see [Burq et al. 2004; Staffilani and Tataru 2002] and for more details Proposition 5.4 in [Bouclet and Tzvetkov 2007]):

$$\begin{aligned}\|\pi_N v\|_{L^2((0,1);L^6(M^3))} &\\ \lesssim \|\pi_N v\|_{L^\infty((0,1);L^2(M^3))} &+ \|\pi_N v\|_{L^2((0,1);H^{1/2}(M^3))} + \|\pi_N F\|_{L^2((0,1);L^{6/5}(M^3))},\end{aligned}$$

where (π_N) is the usual Littlewood–Paley spectral projector and N ranges over dyadic numbers. In fact by taking squares and summing over N we get (10), provided that we make use of the bound

$$\sum_N \|\pi_N v\|^2_{L^\infty((0,1);L^2(M^3))} \le \sum_N \frac{1}{N^\epsilon}\|v\|^2_{L^\infty((0,1);H^\epsilon(M^3))}$$

together with the equivalence of the L^r norm of v with the L^r $(1 < r < \infty)$ norm of its squared function $\left(\sum_N |\pi_N v|^2\right)^{\frac{1}{2}}$.

3. Modified energies associated with even Sobolev norms

Modified energies. In this subsection we consider the general Cauchy problem

$$\begin{cases} i\,\partial_t u + \Delta_g u = |u|^{p-1}u, \quad (t,x) \in \mathbb{R}\times M^d, \\ u(0,x) = \varphi \in H^{2k}(M^d), \end{cases} \tag{11}$$

where (M^d, g) is a compact d-dimensional Riemannian manifold.

In the sequel we shall extensively make use of the following bound without further notice:

$$\|u\|_{L^\infty(\mathbb{R};H^1(M^d))} \lesssim_{p,\|\varphi\|_{H^1}} 1. \tag{12}$$

For every solution $u(t,x)$ to the Cauchy problem (11) we introduce the following energy, to be used in connection with growth of the Sobolev norm H^{2k}:

$$\mathcal{E}_{2k}(u) = \|\partial_t^k u\|^2_{L^2(M^d)} - \frac{p-1}{4}\int_{M^d} \big|\partial_t^{k-1}\nabla_g(|u|^2)\big|_g^2\,|u|^{p-3}\,\mathrm{dvol}_g - \int_{M^d}\big|\partial_t^{k-1}(|u|^{p-1}u)\big|^2\,\mathrm{dvol}_g.$$

We have the following key identity.

Proposition 3.1. *Let $u(t,x)$ be a solution to* (11)*, where $p = 2n+1 \geq 3$, with initial data $\varphi \in H^{2k}(M^d)$. Then we have*

$$\begin{aligned}
\frac{d}{dt}\mathcal{E}_{2k}(u(t,x)) = &-\frac{p-1}{4}\int_{M^d}\big|\partial_t^{k-1}\nabla_g(|u|^2)\big|_g^2\,\partial_t(|u|^{p-3})\,\mathrm{dvol}_g + 2\int_{M^d}\partial_t^k(|u|^{p-1})\,\partial_t^{k-1}(|\nabla_g u|_g^2)\,\mathrm{dvol}_g \\
&+\sum_{j=0}^{k-1} c_j\int_{M^d}\partial_t^j\nabla_g(|u|^2)\,\partial_t^{k-1}\nabla_g(|u|^2)_g\,\partial_t^{k-j}(|u|^{p-3})\,\mathrm{dvol}_g \\
&+\mathrm{Re}\sum_{j=0}^{k-1} c_j\int_{M^d}\partial_t^j(|u|^{p-1})\,\partial_t^{k-j}u\,\partial_t^{k-1}(|u|^{p-1}\bar u)\,\mathrm{dvol}_g \\
&+\mathrm{Re}\sum_{j=0}^{k-2} c_j\int_{M^d}\partial_t^k(|u|^{p-1})\,\partial_t^j(\Delta_g\bar u)\,\partial_t^{k-1-j}u\,\mathrm{dvol}_g \\
&+\mathrm{Im}\sum_{j=1}^{k-1} c_j\int_{M^d}\partial_t^j(|u|^{p-1})\,\partial_t^{k-j}u\,\partial_t^k\bar u\,\mathrm{dvol}_g,
\end{aligned} \tag{13}$$

where c_j denote explicit constants that may change from line to line.

Proof. We start with the following computation:

$$\begin{aligned}
\frac{d}{dt}\|\partial_t^k u\|^2_{L^2(M^d)} &= 2\,\mathrm{Re}(\partial_t^{k+1}u, \partial_t^k u) = 2\,\mathrm{Re}\big(\partial_t^k(-\Delta_g u + |u|^{p-1}u), i\,\partial_t^k u\big) \\
&= 2\,\mathrm{Im}\int_{M^d}(\partial_t^k\nabla_g u, \partial_t^k\nabla_g u)_g\,\mathrm{dvol}_g + 2\,\mathrm{Re}\big(\partial_t^k(|u|^{p-1}u), i\,\partial_t^k u\big),
\end{aligned}$$

where (f,g) denotes the usual $L^2(M^d)$ scalar product $\int_{M^d} f\cdot\bar g\,\mathrm{dvol}_g$. Since the first term on the right-hand side vanishes identically we get

$$\begin{aligned}\frac{d}{dt}\|\partial_t^k u\|^2_{L^2(M^d)}&=2\,\mathrm{Re}\big(\partial_t^k(|u|^{p-1}u),i\,\partial_t^k u\big)\\&=2\,\mathrm{Re}\big(\partial_t^k(|u|^{p-1})u,i\,\partial_t^k u\big)+2\,\mathrm{Re}\big(|u|^{p-1}\,\partial_t^k u,i\,\partial_t^k u\big)+\mathrm{Re}\sum_{j=1}^{k-1}c_j\big(\partial_t^j(|u|^{p-1})\partial_t^{k-j}u,i\,\partial_t^k u\big),\end{aligned}$$

where c_j are suitable integers. Notice that the second term on the right-hand side vanishes identically and if we substitute for the equation again then we get

$$\begin{aligned}&\frac{d}{dt}\|\partial_t^k u\|^2_{L^2(M^d)}\\&\quad=2\,\mathrm{Re}\big(\partial_t^k(|u|^{p-1})u,-\Delta_g(\partial_t^{k-1}u)\big)+2\,\mathrm{Re}\big(\partial_t^k(|u|^{p-1})u,\partial_t^{k-1}(|u|^{p-1}u)\big)\\&\qquad+\mathrm{Re}\sum_{j=1}^{k-1}c_j\big(\partial_t^j|u|^{p-1}\,\partial_t^{k-j}u,i\,\partial_t^k u\big)\\&\quad=2\,\mathrm{Re}\big(\partial_t^k(|u|^{p-1})u,-\Delta_g(\partial_t^{k-1}u)\big)+2\,\mathrm{Re}\big(\partial_t^k(|u|^{p-1}u),\partial_t^{k-1}(|u|^{p-1}u)\big)\\&\qquad+\mathrm{Re}\sum_{j=0}^{k-1}c_j\big(\partial_t^j(|u|^{p-1})\,\partial_t^{k-j}u,\partial_t^{k-1}(|u|^{p-1}u)\big)+\mathrm{Re}\sum_{j=1}^{k-1}c_j\big(\partial_t^j|u|^{p-1}\,\partial_t^{k-j}u,i\,\partial_t^k u\big)\\&\quad=2\,\mathrm{Re}\big(\partial_t^k(|u|^{p-1})u,-\Delta_g(\partial_t^{k-1}u)\big)+\int_{M^d}\partial_t|\partial_t^{k-1}(|u|^{p-1}u)|^2\,\mathrm{dvol}_g\\&\qquad+\mathrm{Re}\sum_{j=0}^{k-1}c_j\big(\partial_t^j(|u|^{p-1})\,\partial_t^{k-j}u,\partial_t^{k-1}(|u|^{p-1}u)\big)+\mathrm{Re}\sum_{j=1}^{k-1}c_j\big(\partial_t^j(|u|^{p-1})\,\partial_t^{k-j}u,i\,\partial_t^k u\big).\end{aligned}\tag{14}$$

Next we focus on the first term on the right-hand side

$$2\,\mathrm{Re}\big(\partial_t^k(|u|^{p-1})u,-\Delta_g(\partial_t^{k-1}u)\big)=\int_{M^d}\partial_t^k(|u|^{p-1})\big(-\bar u\,\partial_t^{k-1}(\Delta_g u)-u\,\partial_t^{k-1}(\Delta_g\bar u)\big)\,\mathrm{dvol}_g$$

and we notice

$$-\bar u\Delta_g(\partial_t^{k-1}u)-u\Delta_g(\partial_t^{k-1}\bar u)=\partial_t^{k-1}(-\bar u\Delta_g u-u\Delta_g\bar u)+\mathrm{Re}\sum_{j=0}^{k-2}c_j\,\partial_t^j(\Delta_g u)\,\partial_t^{k-1-j}\bar u.$$

Moreover we have the identity

$$\Delta_g(|u|^2)=u\Delta_g\bar u+\bar u\Delta_g u+2|\nabla_g u|_g^2.$$

Hence,

$$\begin{aligned}&2\,\mathrm{Re}\big(\partial_t^k(|u|^{p-1})u,-\Delta_g\,\partial_t^{k-1}u\big)\\&\quad=-\int_{M^d}\partial_t^k(|u|^{p-1})\,\partial_t^{k-1}\Delta_g(|u|^2)\,\mathrm{dvol}_g+2\int_{M^d}\partial_t^k(|u|^{p-1})\,\partial_t^{k-1}(|\nabla_g u|_g^2)\,\mathrm{dvol}_g\\&\qquad+\mathrm{Re}\sum_{j=0}^{k-2}c_j\int_{M^d}\partial_t^k(|u|^{p-1})\,\partial_t^j(\Delta_g u)\,\partial_t^{k-1-j}\bar u\,\mathrm{dvol}_g\end{aligned}$$

$$= \int_{M^d} \big(\partial_t^k \nabla_g(|u|^{p-1}), \partial_t^{k-1}\nabla_g(|u|^2)\big)_g \,\mathrm{dvol}_g + 2\int_{M^d} \partial_t^k(|u|^{p-1})\,\partial_t^{k-1}(|\nabla_g u|_g^2)\,\mathrm{dvol}_g + \mathrm{Re}\sum_{j=0}^{k-2} c_j \int_{M^d} \partial_t^k(|u|^{p-1})\,\partial_t^j(\Delta_g u)\,\partial_t^{k-1-j}\bar{u}\,\mathrm{dvol}_g,$$

and by elementary computations we get

$$\cdots = \frac{p-1}{2}\int_{M^d} \big(\partial_t^k(\nabla_g(|u|^2)|u|^{p-3}), \partial_t^{k-1}\nabla_g(|u|^2)\big)_g \,\mathrm{dvol}_g + 2\int_{M^d} \partial_t^k(|u|^{p-1})\,\partial_t^{k-1}(|\nabla_g u|_g^2)\,\mathrm{dvol}_g + \mathrm{Re}\sum_{j=0}^{k-2} c_j \int_{M^d} \partial_t^k(|u|^{p-1})\,\partial_t^j(\Delta_g u)\,\partial_t^{k-1-j}\bar{u}\,\mathrm{dvol}_g.$$

Using the Leibniz rule to develop ∂_t^k we get

$$\begin{aligned}
\cdots &= \frac{p-1}{2}\int_{M^d} (\partial_t^k \nabla_g(|u|^2)|u|^{p-3}, \partial_t^{k-1}\nabla_g(|u|^2))_g \,\mathrm{dvol}_g \\
&\quad + \sum_{j=0}^{k-1} c_j \int_{M^d} (\partial_t^j \nabla_g(|u|^2), \partial_t^{k-1}\nabla_g(|u|^2))_g\,\partial_t^{k-j}(|u|^{p-3})\,\mathrm{dvol}_g \\
&\quad + 2\int_{M^d} \partial_t^k(|u|^{p-1})\,\partial_t^{k-1}(|\nabla_g u|_g^2)\,\mathrm{dvol}_g + \mathrm{Re}\sum_{j=0}^{k-2} c_j \int_{M^d} \partial_t^k(|u|^{p-1})\,\partial_t^j(\Delta_g u)\,\partial_t^{k-1-j}\bar{u}\,\mathrm{dvol}_g \\
&= \frac{p-1}{4}\int_{M^d} \partial_t |\partial_t^{k-1}\nabla_g(|u|^2)|_g^2 |u|^{p-3}\,\mathrm{dvol}_g \\
&\quad + \sum_{j=0}^{k-1} c_j \int_{M^d} (\partial_t^j \nabla_g(|u|^2), \partial_t^{k-1}\nabla_g(|u|^2))_g\,\partial_t^{k-j}(|u|^{p-3})\,\mathrm{dvol}_g \\
&\quad + 2\int_{M^d} \partial_t^k(|u|^{p-1})\,\partial_t^{k-1}(|\nabla_g u|_g^2)\,\mathrm{dvol}_g + \mathrm{Re}\sum_{j=0}^{k-2} c_j \int_{M^d} \partial_t^k(|u|^{p-1})\,\partial_t^j(\Delta_g u)\,\partial_t^{k-1-j}\bar{u}\,\mathrm{dvol}_g,
\end{aligned}$$

and we conclude by combining this identity with (14). □

Remark 3.2. In the specific case of the cubic NLS (i.e., (11) with $p = 3$) we have some simplifications; more precisely we get

$$\mathcal{E}_{2k}(u) = \|\partial_t^k u\|_{L^2(M^d)}^2 - \frac{1}{2}\int_{M^d} \big|\partial_t^{k-1}\nabla_g(|u|^2)\big|_g^2 \,\mathrm{dvol}_g - \int_{M^d} \big|\partial_t^{k-1}(|u|^2 u)\big|^2 \,\mathrm{dvol}_g$$

and also

$$\begin{aligned}
&\frac{d}{dt}\mathcal{E}_{2k}(u(t,x)) \\
&= 2\int_{M^d} \partial_t^k(|u|^2)\,\partial_t^{k-1}(|\nabla_g u|_g^2)\,\mathrm{dvol}_g + \mathrm{Re}\sum_{j=0}^{k-2} c_j \int_{M^d} \partial_t^k(|u|^2)\,\partial_t^j(\Delta_g u)\,\partial_t^{k-1-j}\bar{u}\,\mathrm{dvol}_g \\
&\quad + \mathrm{Re}\sum_{j=0}^{k-1} c_j \int_{M^d} \partial_t^j(|u|^2)\,\partial_t^{k-j} u\,\partial_t^{k-1}(|u|^2\bar{u})\,\mathrm{dvol}_g + \mathrm{Im}\sum_{j=1}^{k-1} c_j \int_{M^d} \partial_t^j(|u|^2)\,\partial_t^{k-j} u\,\partial_t^k \bar{u}\,\mathrm{dvol}_g. \quad (15)
\end{aligned}$$

The norms $\|\partial_t^k u\|_{L^2}$ and $\|u\|_{H^{2k}}$ are comparable. The aim of this subsection is indeed to prove that the leading term in our modified energy $\mathcal{E}_{2k}(u)$ is equivalent to the Sobolev norm $\|u\|_{H^{2k}}$, provided that $u(t,x)$ is a solution to (11) with $d=2$ and $p\geq 3$ or $d=3$ and $p=3$.

Proposition 3.3. *Let $u(t,x)$ be solution to* (11)*, where either $d=2$ and $p\geq 3$ is an integer, or $d=3$ and $p=3$. Then for every $k,s\in\mathbb{N}$ we have*

$$\|\partial_t^k u - i^k \Delta_g^k u\|_{H^s(M^d)} \lesssim_{\|\varphi\|_{H^1}} \|u\|_{H^{s+2k-1}(M^d)}. \tag{16}$$

Proof. We shall use the following identity (satisfied by every solution to (11) in any dimension d):

$$\partial_t^h u = i^h \Delta_g^h u + \sum_{j=0}^{h-1} c_j\, \partial_t^j \Delta_g^{h-j-1}(u|u|^{p-1}), \tag{17}$$

where $c_j\in\mathbb{C}$ are suitable coefficients. The elementary proof follows by induction on h and by using the equation solved by $u(t,x)$.

First case: $d=2$, $p\geq 3$. We argue by induction on k, and hence we shall prove $k\Rightarrow k+1$. By (17) we aim at proving

$$\|\partial_t^j(u|u|^{p-1})\|_{H^{2k-2j+s}(M^2)} \lesssim \|u\|_{H^{s+2k+1}(M^2)}, \quad j=0,\dots,k, \tag{18}$$

by assuming the property (16) is true for k. By expanding the time and space derivatives on the left-hand side above, we deduce (18) by the chain of inequalities

$$\begin{aligned}\prod_{\substack{j_1+\cdots+j_p=j\\ s_1+\cdots+s_p=2k-2j+s}} \|\partial_t^{j_l} u\|_{W^{s_l,2p}(M^2)} &\lesssim \prod_{\substack{j_1+\cdots+j_p=j\\ s_1+\cdots+s_p=2k-2j+s}} \|\partial_t^{j_l} u\|_{H^{s_l+1}(M^2)}\\ &\lesssim \prod_{\substack{j_1+\cdots+j_p=j\\ s_1+\cdots+s_p=2k-2j+s}} \|u\|_{H^{2j_l+s_l+1}(M^2)},\end{aligned}$$

where we used the Sobolev embedding $H^1(M^2)\subset L^{2p}(M^2)$ and we have used the induction hypothesis at the last step. We can continue the estimate by a trivial interpolation argument as follows:

$$\cdots \lesssim \Big(\prod_{l=1,\dots,p} \|u\|^{\theta_l}_{H^{s+2k+1}(M^2)}\, \|u\|^{(1-\theta_l)}_{H^1(M^2)}\Big),$$

where

$$\theta_l(s+2k+1)+(1-\theta_l) = 2j_l + s_l + 1.$$

We conclude using (12), since $\sum_{l=1}^p \theta_l = 1$ for $j=0,\dots,k$.

Second case: $d=3$, $p=3$. Arguing as above, and by assuming the result true for k, we are reduced to proving

$$\|\partial_t^j(u|u|^2)\|_{H^{2k-2j+s}(M^3)} \lesssim \|u\|_{H^{s+2k+1}(M^3)}, \quad j=0,\dots,k. \tag{19}$$

Expanding again the time and space derivatives on the left-hand side, we are reduced to the estimate

$$\begin{aligned}\|\partial_t^{j_1}u\|_{W^{k_1,6}(M^3)}\times\|\partial_t^{j_2}u\|_{W^{k_2,6}(M^3)}\times\|\partial_t^{j_3}u\|_{W^{k_3,6}(M^3)}&\\ \lesssim\|\partial_t^{j_1}u\|_{H^{k_1+1}(M^3)}\times\|\partial_t^{j_2}u\|_{H^{k_2+1}(M^3)}\times\|\partial_t^{j_3}u\|_{H^{k_3+1}(M^3)}&\\ \lesssim\|u\|_{H^{2j_1+k_1+1}(M^3)}\|u\|_{H^{2j_2+k_2+1}(M^3)}\|u\|_{H^{2j_3+k_3+1}(M^3)},&\end{aligned}$$

where

$$j_1+j_2+j_3=j,\quad k_1+k_2+k_3=2k-2j+s.$$

Notice that we have used the Sobolev embedding $H^1(M^3)\subset L^6(M^3)$ and the induction hypothesis at the last step. By interpolation we have

$$\|u\|_{H^{2j_l+k_l+1}(M^3)}\lesssim\|u\|^{\theta_l}_{H^{s+2k+1}(M^3)}\|u\|^{1-\theta_l}_{H^1(M^3)},\quad l=1,2,3,$$

where

$$\theta_l(s+2k+1)+(1-\theta_l)=2j_l+k_l+1,$$

and we conclude as above since $\sum_{l=1}^3\theta_l=1$ for $j=0,\dots,k$. □

Strichartz estimates for nonlinear solutions. In this subsection we get a priori bounds for the Strichartz norms of solutions to (11) in dimension $d=2$, with a general nonlinearity, and in dimension $d=3$, with cubic nonlinearity. In the sequel we denote by $L^p_\tau X$ the space $L^p((0,\tau);X)$, where X is a Banach space and $p\in[1,\infty]$.

Proposition 3.4. *We have the following estimate for every solution $u(t,x)$ to (11) for $d=2$ and $p=2n+1\geq 3$ is an integer: for any $\epsilon>0$ and $\tau\in(0,1)$,*

$$\|\partial_t^j u\|_{L^4_\tau W^{s,4}(M^2)}\lesssim_{\epsilon,\|\varphi\|_{H^1}}\|u\|^{1-s_0}_{L^\infty_\tau H^{2j+s}(M^2)}\|u\|^{s_0}_{L^\infty_\tau H^{2j+s+1}(M^2)}\|u\|^{\epsilon}_{L^\infty_\tau H^{2j+2}(M^2)}.\tag{20}$$

Proof. We use (9), together with the equation solved by $\partial_t^j u$, and we get

$$\begin{aligned}\|\partial_t^j u\|_{L^4_\tau W^{s,4}(M^2)}&\\ \lesssim\|\partial_t^j u(0)\|_{H^{s+s_0}(M^2)}+\tau\|\partial_t^j(u|u|^{p-1})\|_{L^\infty_\tau H^{s+s_0}(M^2)}&\\ \lesssim\|\partial_t^j u(0)\|^{1-s_0}_{H^s}\|\partial_t^j u(0)\|^{s_0}_{H^{s+1}(M^2)}+T\|\partial_t^j(u|u|^{p-1})\|^{1-s_0}_{L^\infty_\tau H^s(M^2)}\|\partial_t^j(u|u|^{p-1})\|^{s_0}_{L^\infty_\tau H^{s+1}(M^2)}.&\end{aligned}$$

Notice that the first term on the right-hand side can be estimated by Proposition 3.3. Hence we shall complete the proof provided that for every $\epsilon>0$,

$$\|\partial_t^j(u|u|^{p-1})\|_{H^s(M^2)}\lesssim_{\epsilon,\|\varphi\|_{H^1}}\|u\|_{H^{2j+s}(M^2)}\|u\|^{\epsilon}_{H^{2j+2}(M^2)}\quad\text{for all } j,s=1,2,\dots.$$

Expanding the time derivative ∂_t^j and using

$$\|fg\|_{H^r(M^2)}\lesssim\|f\|_{H^r(M^2)}\|g\|_{L^\infty(M^2)}+\|g\|_{H^r(M^2)}\|f\|_{L^\infty(M^2)},$$

we are reduced to estimating

$$\|\partial_t^{j_1}u\|_{H^s(M^2)}\times\|\partial_t^{j_2}u\|_{L^\infty(M^2)}\times\cdots\times\|\partial_t^{j_p}u\|_{L^\infty(M^2)},$$

where $j_1 + \cdots + j_p = j$. Notice that from

$$\|v\|_{L^\infty(M^2)} \lesssim_\epsilon \|v\|^{1-\epsilon}_{H^1(M^2)} \|v\|^{\epsilon}_{H^2(M^2)} \tag{21}$$

we get

$$\begin{aligned}
&\|\partial_t^{j_1} u\|_{H^s(M^2)} \times \|\partial_t^{j_2} u\|_{L^\infty(M^2)} \times \cdots \times \|\partial_t^{j_p} u\|_{L^\infty(M^2)} \\
&\quad \lesssim_\epsilon \|\partial_t^{j_1} u\|_{H^s(M^2)} \times \|\partial_t^{j_2} u\|^{1-\epsilon}_{H^1(M^2)} \times \|\partial_t^{j_2} u\|^{\epsilon}_{H^2} \times \cdots \times \|\partial_t^{j_p} u\|^{1-\epsilon}_{H^1(M^2)} \times \|\partial_t^{j_p} u\|^{\epsilon}_{H^2(M^2)},
\end{aligned}$$

and hence by (16)

$$\begin{aligned}
\cdots &\lesssim \|u\|_{H^{2j_1+s}(M^2)} \times \|u\|^{1-\epsilon}_{H^{2j_2+1}(M^2)} \times \|u\|^{\epsilon}_{H^{2j_2+2}(M^2)} \times \cdots \times \|u\|^{1-\epsilon}_{H^{2j_p+1}(M^2)} \times \|u\|^{\epsilon}_{H^{2j_p+2}(M^2)} \\
&\lesssim \|u\|^{\theta_1}_{H^{2j+s}(M^2)} \|u\|^{1-\theta_1}_{H^1(M^2)} \times \|u\|^{\theta_2(1-\epsilon)}_{H^{2j+s}} \|u\|^{(1-\theta_2)(1-\epsilon)}_{H^1(M^2)} \\
&\qquad \times \cdots \times \|u\|^{\theta_p(1-\epsilon)}_{H^{2j+s}(M^2)} \|u\|^{(1-\theta_p)(1-\epsilon)}_{H^1(M^2)} \times \|u\|^{\epsilon(p-1)}_{H^{2j+2}(M^2)},
\end{aligned}$$

where at the last step we have used an interpolation argument with

$$\theta_1(2j+s) + (1-\theta_1) = 2j_1 + s, \qquad \theta_l(2j+s) + (1-\theta_l) = 2j_l + 1, \quad l = 2, \ldots, p.$$

Notice that we get $\sum_{l=1}^p \theta_l = 1$ and we conclude by (12). □

Proposition 3.5. *We have the following estimate for every solution $u(t,x)$ to (11) for $(p,l) = (3,3)$ and for every $\epsilon > 0$, $\tau \in (0,1)$:*

$$\begin{aligned}
\|\partial_t^j u\|_{L^2_\tau L^6(M^3)} &\lesssim_{\epsilon, \|\varphi\|_{H^1}} \|\partial_t^j u\|^{1-\epsilon}_{L^\infty_\tau L^2(M^3)} \|\partial_t^j u\|^{\epsilon}_{L^\infty_\tau H^1(M^3)} + \sqrt{\tau} \|u\|^{1/2}_{L^\infty_\tau H^{2j}(M^3)} \|u\|^{1/2}_{L^\infty_\tau H^{2j+1}(M^3)} \\
&\quad + \sqrt{\tau} \sum_{\substack{j_1+j_2+j_3=j \\ j_1 = \max\{j_1,j_2,j_3\}}} \|u\|_{L^\infty_\tau H^{2j_1}(M^3)} \|u\|_{L^\infty_\tau H^{2j_2+1}(M^3)} \|u\|_{L^\infty_\tau H^{2j_3+1}(M^3)},
\end{aligned} \tag{22}$$

and

$$\begin{aligned}
\|\partial_t^j u\|_{L^2_\tau W^{1,6}(M^3)} &\lesssim_{\epsilon, \|\varphi\|_{H^1}} \|\partial_t^j u\|^{1-\epsilon}_{L^\infty_\tau H^1(M^3)} \|\partial_t^j u\|^{\epsilon}_{L^\infty_\tau H^2(M^3)} + \sqrt{\tau} \|u\|^{1/2}_{L^\infty_\tau H^{2j+1}(M^3)} \|u\|^{1/2}_{L^\infty_\tau H^{2j+2}(M^3)} \\
&\quad + \sqrt{\tau} \sum_{j_1+j_2+j_3=j} \|u\|_{L^\infty_\tau H^{2j_1+1}(M^3)} \|u\|_{L^\infty_\tau H^{2j_2+1}(M^3)} \|u\|_{L^\infty_\tau H^{2j_3+1}(M^3)}.
\end{aligned} \tag{23}$$

Proof. We prove (23), the proof of (22) being similar. By using Strichartz estimates and the equation solved by $\partial_t^j u$ we get

$$\begin{aligned}
\|\partial_t^j u\|_{L^2_\tau W^{1,6}(M^3)} &\lesssim \|\partial_t^j u\|_{L^\infty_\tau H^{1+\epsilon}(M^3)} + \sqrt{T} \|\partial_t^j u\|_{L^\infty_\tau H^{3/2}(M^3)} + \|\partial_t^j (u|u|^2)\|_{L^2_\tau W^{1,6/5}(M^3)} \\
&\lesssim \|\partial_t^j u\|^{1-\epsilon}_{L^\infty_\tau H^1(M^3)} \|\partial_t^j u\|^{\epsilon}_{L^\infty_\tau H^2(M^3)} \\
&\qquad + \|\partial_t^j u\|^{1/2}_{L^\infty_\tau H^1(M^3)} \|\partial_t^j u\|^{1/2}_{L^\infty_\tau H^2(M^3)} + \|\partial_t^j (u|u|^2)\|_{L^2_\tau W^{1,6/5}(M^3)}.
\end{aligned}$$

Notice that by expanding the time derivative, and by using Hölder we get

$$\begin{aligned}\|\partial_t^j(u|u|^2)\|_{W^{1,6/5}(M^3)} &\lesssim \sum_{\substack{j_1+j_2+j_3=j\\ j_1=\max\{j_1,j_2,j_3\}}} \|\partial_t^{j_1}u\|_{H^1(M^3)}\,\|\partial_t^{j_2}u\|_{L^6(M^3)}\,\|\partial_t^{j_3}u\|_{L^6(M^3)}\\ &\lesssim \sum_{\substack{j_1+j_2+j_3=j\\ j_1=\max\{j_1,j_2,j_3\}}} \|\partial_t^{j_1}u\|_{H^1(M^3)}\,\|\partial_t^{j_2}u\|_{H^1(M^3)}\,\|\partial_t^{j_3}u\|_{H^1(M^3)}.\end{aligned}$$

We then conclude by using Proposition 3.3 in the special case of the cubic NLS on M^3. □

4. Polynomial growth of H^{2k} for pure power NLS on M^2

This section is devoted to the proof of Theorem 1.1 in the case $m = 2k$. We shall need the following estimate.

Proposition 4.1. *Let us assume that $u(t,x)$ solves* (11) *with $d = 2$ and $p = 2n+1 \geq 3$. Then we have the following bound for every $\tau \in (0,1)$:*

$$\int_0^\tau |\textit{right-hand side of } (13)|\,ds \lesssim \sqrt{\tau}\|u\|^{\frac{4k-3+2s_0}{2k-1}+\epsilon}_{L^\infty_\tau H^{2k}(M^2)} + \|u\|^{\frac{4k-4}{2k-1}+\epsilon}_{L^\infty_\tau H^{2k}(M^2)}.$$

Proof. Since we work on a 2-dimensional compact manifold we simplify notation as follows: L^q, $W^{s,q}$, H^s denote the spaces $L^q(M^2)$, $W^{s,q}(M^2)$, $H^s(M^2)$. Moreover in the sequel we shall denote by $\epsilon > 0$ any arbitrary small constant whose value can change from line to line. We shall also make use of the inequality

$$\|u\|_{L^\infty_T H^s} \lesssim_{\|\varphi\|_{H^1}} \|u\|^{\frac{s-1}{2k-1}}_{L^\infty_\tau H^{2k}}, \quad s \in [1,2k], \tag{24}$$

which in turn follows by combining an elementary interpolation inequality with (12).

Let I, II, III, IV, V, VI be the successive terms on each line of the right-hand side in (13). Estimating I can be reduced to controlling the terms

$$\int_0^T \|\partial_t^{k_1}u\|^2_{W^{1,4}}\,\|\partial_t^{k_2}u\|^2_{L^\infty}\,\|\partial_t u\|_{L^2}\,\|u\|^{p-4}_{L^\infty}\,ds, \quad k_1+k_2 = k-1, \tag{25}$$

and we have, by combining (21), Proposition 3.3, Proposition 3.4 and the Hölder inequality,

$$\begin{aligned}(25) &\lesssim \sqrt{T}\|u\|^{2(1-\epsilon)}_{L^\infty_\tau H^{2k_2+1}}\,\|u\|^{2\epsilon}_{L^\infty H^{2k_2+2}}\,\|u\|_{L^\infty_\tau H^2}\,\|\partial_t^{k_1}u\|^2_{L^4_T W^{1,4}}\\ &\lesssim \sqrt{T}\|u\|^{2(1-\epsilon)}_{L^\infty_\tau H^{2k_2+1}}\,\|u\|_{L^\infty H^2}\,\|u\|^{2(1-s_0)}_{L^\infty_\tau H^{2k_1+1}}\,\|u\|^{2s_0}_{L^\infty_\tau H^{2k_1+2}}\,\|u\|^{\epsilon}_{L^\infty_\tau H^{2k}} \lesssim \sqrt{\tau}\|u\|^{\frac{4k-3+2s_0}{2k-1}+\epsilon}_{L^\infty_\tau H^{2k}},\end{aligned}$$

where at the last step we have used (24). Notice that the value of $\epsilon > 0$ changes at each line, but can be chosen arbitrarily small. Concerning II, we are reduced to controlling

$$\int_0^T \|\partial_t^{j_1}u\|_{L^2}\Big(\prod_{h=2,\dots,p-1}\|\partial_t^{j_h}u\|_{L^\infty}\Big)\|\partial_t^{k_1}u\|_{W^{1,4}}\,\|\partial_t^{k_2}u\|_{W^{1,4}}, \tag{26}$$

where we assume $j_1 = \max\{j_1, j_2, \dots, j_{p-1}\}$ and

$$j_1 + \cdots + j_{p-1} = k, \quad k_1 + k_2 = k - 1.$$

By using the interpolation estimate (21) together with Proposition 3.3, Proposition 3.4 and the Hölder inequality, we get

$$\begin{aligned}(26) &\lesssim \sqrt{T}\,\|u\|^{\epsilon}_{L^\infty_T H^{2k}}\,\|u\|_{L^\infty_T H^{2j_1}}\Big(\prod_{h=2,\dots,p-1}\|u\|^{1-\epsilon}_{L^\infty_T H^{2j_h+1}}\Big)\\ &\qquad\times \|u\|^{1-s_0}_{L^\infty_T H^{2k_1+1}}\,\|u\|^{s_0}_{L^\infty_T H^{2k_1+2}}\,\|u\|^{1-s_0}_{L^\infty_T H^{2k_2+1}}\,\|u\|^{s_0}_{L^\infty_T H^{2k_2+2}}\\ &\lesssim \sqrt{\tau}\,\|u\|^{\frac{4k-3+2s_0}{2k-1}+\epsilon}_{L^\infty_\tau H^{2k}},\end{aligned}$$

where we used (24) at the last step. Next we deal with III, and it is sufficient to control

$$\int_0^\tau \|\partial_t^{h_1}u\|_{L^\infty}\,\|\partial_t^{h_2}u\|_{W^{1,4}}\,\|\partial_t^{m_1}u\|_{L^2}\Big(\prod_{i=2,\dots,p-3}\|\partial_t^{m_i}u\|_{L^\infty}\Big)\|\partial_t^{l_1}u\|_{L^\infty}\,\|\partial_t^{l_2}u\|_{W^{1,4}}, \tag{27}$$

where we assume $m_1 = \max\{m_1, m_2, \dots, m_{p-3}\}$ and

$$h_1 + h_2 = j \in [0, k-1], \quad m_1 + \cdots + m_{p-3} = k - j, \quad l_1 + l_2 = k - 1.$$

Arguing as above, it can be estimated by

$$\begin{aligned}(27) &\lesssim \sqrt{\tau}\|u\|^{\epsilon}_{L^\infty_\tau H^{2k}}\,\|u\|_{L^\infty_\tau H^{2h_1+1}}\,\|u\|^{1-s_0}_{L^\infty_\tau H^{2h_2+1}}\,\|u\|^{s_0}_{L^\infty_\tau H^{2h_2+2}}\,\|u\|_{L^\infty_\tau H^{2m_1}}\\ &\qquad\times\Big(\prod_{i=2,\dots,p-3}\|u\|^{1-\epsilon}_{L^\infty H^{2m_i+1}}\Big)\|u\|^{1-\epsilon}_{L^\infty_\tau H^{2l_1+1}}\,\|u\|^{1-s_0}_{L^\infty_\tau H^{2l_2+1}}\,\|u\|^{s_0}_{L^\infty_\tau H^{2l_2+2}}\\ &\lesssim \sqrt{\tau}\|u\|^{\frac{4k-3+2s_0}{2k-1}+\epsilon}_{L^\infty_\tau H^{2k}}.\end{aligned}$$

In order to treat IV we are reduced to controlling

$$\int_0^T \|\partial_t^{j_1}u\|_{L^2}\Big(\prod_{h=2,\dots,p-1}\|\partial_t^{j_h}u\|_{L^\infty}\Big)\|\partial_t^{j}(\Delta_g u)\|_{L^4}\,\|\partial_t^{k-1-j}\bar{u}\|_{L^4}, \tag{28}$$

where we assume $j_1 = \max\{j_1, j_2, \dots, j_{p-1}\}$ and

$$j_1 + \cdots + j_{p-1} = k,$$

and by an argument similar to those above we have

$$\begin{aligned}(28) &\lesssim \sqrt{\tau}\|u\|^{\epsilon}_{L^\infty_\tau H^{2k}}\,\|u\|_{L^\infty_T H^{2j_1}}\Big(\prod_{h=2,\dots,p-1}\|u\|^{1-\epsilon}_{L^\infty_\tau H^{2j_h+1}}\Big)\\ &\qquad\times \|u\|^{1-s_0}_{L^\infty_\tau H^{2j+2}}\,\|u\|^{s_0}_{L^\infty_\tau H^{2j+3}}\,\|u\|^{1-s_0}_{L^\infty_\tau H^{2k-2-2j}}\,\|u\|^{s_0}_{L^\infty_\tau H^{2k-2j-1}}\\ &\lesssim \sqrt{\tau}\|u\|^{\frac{4k-3+2s_0}{2k-1}+\epsilon}_{L^\infty_\tau H^{2k}}.\end{aligned}$$

In order to estimate V it is sufficient to control the terms

$$\int_0^T \|\partial_t^{m_1} u\|_{L^4} \Big(\prod_{i=2,\dots,p-1} \|\partial_t^{m_i} u\|_{L^\infty} \Big) \|\partial_t^{k-j} u\|_{L^4} \, \|\partial_t^k u\|_{L^2}, \tag{29}$$

where we assume $m_1 = \max\{m_1, m_2, \dots, m_{p-1}\}$ and

$$m_1 + \cdots + m_{p-1} = j,$$

and as usual we get

$$\begin{aligned}(29) &\lesssim \sqrt{\tau} \|u\|^{1+\epsilon}_{L^\infty_\tau H^{2k}} \, \|u\|^{1-s_0}_{L^\infty_\tau H^{2m_1}} \, \|u\|^{s_0}_{L^\infty_\tau H^{2m_1+1}} \Big(\prod_{i=2,\dots,p-1} \|u\|^{1-\epsilon}_{L^\infty_\tau H^{2m_i+1}} \Big) \\ &\qquad \times \|u\|^{1-s_0}_{L^\infty_\tau H^{2k-2j}} \, \|u\|^{s_0}_{L^\infty_\tau H^{2k-2j+1}} \\ &\lesssim \sqrt{\tau} \|u\|^{\frac{4k-3+2s_0}{2k-1}+\epsilon}_{L^\infty_\tau H^{2k}}.\end{aligned}$$

We conclude with the estimate of VI, which in turn can be reduced to controlling

$$\int_0^T \|\partial_t^{m_1} u\|_{L^2} \Big(\prod_{i=2,\dots,p-1} \|\partial_t^{m_i} u\|_{L^\infty} \Big) \|\partial_t^{k-j} u\|_{L^\infty} \, \|\partial_t^{l_1} u\|_{L^2} \Big(\prod_{i=2,\dots,p} \|\partial_t^{l_i} u\|_{L^\infty} \Big), \tag{30}$$

where we assume $m_1 = \max\{m_1, m_2, \dots, m_{p-1}\}$ and

$$m_1 + \cdots + m_{p-1} = j, \quad l_1 + \cdots + l_p = k-1,$$

and we get

$$\begin{aligned}(30) &\lesssim \tau \|u\|^{\epsilon}_{L^\infty_\tau H^{2k}} \, \|u\|_{L^\infty_\tau H^{2m_1}} \Big(\prod_{i=2,\dots,p-1} \|u\|^{1-\epsilon}_{L^\infty_\tau H^{2m_i+1}} \Big) \\ &\qquad \times \|u\|^{1-\epsilon}_{L^\infty_\tau H^{2k-2j+1}} \, \|u\|_{L^\infty_\tau H^{2l_1}} \Big(\prod_{i=2,\dots,p-1} \|u\|^{1-\epsilon}_{L^\infty_\tau H^{2l_i+1}} \Big) \\ &\lesssim \tau \|u\|^{\frac{4k-4}{2k-1}+\epsilon}_{L^\infty_\tau H^{2k}}.\end{aligned}$$

□

The key estimate to deduce Theorem 1.1 is the following one (see Remark 1.2).

Proposition 4.2. *Let us assume that $u(t,x)$ solves (11) with $d = 2$ and $p \geq 3$. Then we have the following bound for every $\tau \in (0,1)$ and for every $\epsilon > 0$:*

$$\|u(\tau)\|^2_{H^{2k}} - \|u(0)\|^2_{H^{2k}} \lesssim \sqrt{\tau} \|u\|^{\frac{4k-3+2s_0}{2k-1}+\epsilon}_{L^\infty_\tau H^{2k}} + \|u\|^{\frac{4k-4}{2k-1}+\epsilon}_{L^\infty_\tau H^{2k}}.$$

Proof. We write $\mathcal{E}_{2k}(u) = \|\partial_t^k u\|^2_{L^2} + \mathcal{R}_{2k}(u)$, where

$$\mathcal{R}_{2k}(u) = -\frac{p-1}{4} \int \big|\partial_t^{k-1} \nabla_g (|u|^2)\big|^2_g \, |u|^{p-3} \, \mathrm{dvol}_g - \int \big|\partial_t^{k-1} (|u|^{p-1} u)\big|^2 \, \mathrm{dvol}_g.$$

We claim that

$$|\mathcal{R}_{2k}(u)| \lesssim_\epsilon \|u\|^{\frac{4k-4}{2k-1}+\epsilon}_{H^{2k}} + \|u\|^{\frac{4k-6}{2k-1}+\epsilon}_{H^{2k}}. \tag{31}$$

In fact notice that arguing as in the proof of Proposition 4.1 we get

$$\int \left|\partial_t^{k-1}\nabla_g(|u|^2)\right|_g^2 |u|^{p-3}\,\mathrm{dvol}_g \lesssim \sum_{k_1+k_2=k-1} \|\partial_t^{k_1}u\|_{W^{1,2}}^2 \|\partial_t^{k_2}u\|_{L^\infty}^2 \|u\|_{L^\infty}^{p-3}$$
$$\lesssim \sum_{k_1+k_2=k-1} \|u\|_{H^{2k_1+1}}^2 \|u\|_{H^{2k_2+1}}^2 \|u\|_{H^{2k}}^{\epsilon} \lesssim \|u\|_{H^{2k}}^{\frac{4k-4}{2k-1}+\epsilon}$$

and also

$$\int \left|\partial_t^{k-1}(|u|^{p-1}u)\right|^2 \mathrm{dvol}_g \lesssim \sum_{j_1+\cdots+j_p=k-1} \|\partial_t^{j_1}u\|_{L^2}^2 \Big(\prod_{h=1,\dots,p} \|\partial_t^{j_h}u\|_{L^\infty}^2\Big)$$
$$\lesssim \sum_{j_1+\cdots+j_p=k-1} \|u\|_{H^{2j_1}}^2 \Big(\prod_{h=1,\dots,p} \|u\|_{H^{2j_h+1}(M^2)}^2\Big) \|u\|_{L^\infty H^{2k}}^{\epsilon} \lesssim \|u\|_{L^\infty H^{2k}}^{\frac{4k-6}{2k-1}+\epsilon}.$$

Next notice that if we integrate the identity (13) and we use Proposition 4.1 then

$$\|\partial_t^k u(\tau)\|_{L^2}^2 - \|\partial_t^k u(0)\|_{L^2}^2 \lesssim \sup_{(0,\tau)} |\mathcal{R}_{2k}(u)| + \sqrt{\tau}\|u\|_{L_\tau^\infty H^{2k}}^{\frac{4k-3+2s_0}{2k-1}+\epsilon} + \|u\|_{L_\tau^\infty H^{2k}}^{\frac{4k-4}{2k-1}+\epsilon}.$$

We conclude by using (31) and Proposition 3.3. □

5. Exponential growth for H^{2k} norms of solutions to the cubic NLS on M^3

The aim of this section is the proof of Theorem 1.3 in the case $m = 2k$.

The following is the 3-dimensional version of Proposition 4.1 for the cubic NLS.

Proposition 5.1. *Let us assume that $u(t,x)$ solves* (11) *with $d = 3$ and $p = 3$. Then we have the following bound for every $\tau \in (0,1)$*

$$\int_0^\tau |\textit{right-hand side of } (13)|\, ds \lesssim \tau \|u\|_{L_\tau^\infty H^{2k}}^2 + \|u\|_{L_\tau^\infty H^{2k}}^\gamma$$

for some $\gamma \in (0,2)$.

Proof. Since we work on a 3-dimensional compact manifold we simplify the notation as follows: L^q, $W^{s,q}$, H^s denote the spaces $L^q(M^3)$, $W^{s,q}(M^3)$, $H^s(M^3)$. In the sequel we shall also make use of the following inequalities, which in turn follow by combining an elementary interpolation inequality with (12). We also notice that by combining Proposition 3.3 and Proposition 3.5 with (24) we get

$$\|\partial_t^j u\|_{L_\tau^2 L^6} \lesssim_\epsilon \|u\|_{L_\tau^\infty H^{2j}}^{1-\epsilon} \|u\|_{L_\tau^\infty H^{2j+1}}^{\epsilon} + \sqrt{\tau}\|u\|_{L_\tau^\infty H^{2j}}^{1/2} \|u\|_{L_\tau^\infty H^{2j+1}}^{1/2}$$
$$+ \sqrt{\tau} \sum_{\substack{j_1+j_2+j_3=j \\ j_1=\max\{j_1,j_2,j_3\}}} \|u\|_{L_\tau^\infty H^{2j_1}} \|u\|_{L_\tau^\infty H^{2j_2+1}} \|u\|_{L_\tau^\infty H^{2j_3+1}}$$
$$\lesssim \|u\|_{L_\tau^\infty H^{2k}}^{\frac{2j-1+\epsilon}{2k-1}} + \sqrt{\tau}\|u\|_{L_\tau^\infty H^{2k}}^{\frac{4j-1}{4k-2}} + \sqrt{\tau}\|u\|_{L_\tau^\infty H^{2k}}^{\frac{2j-1}{2k-1}}, \tag{32}$$

provided that $j \geq 1$, and

$$\begin{aligned}\|\partial_t^j u\|_{L^2_\tau W^{1,6}} \lesssim_\epsilon \|u\|^{1-\epsilon}_{L^\infty_\tau H^{2j+1}} \|u\|^{\epsilon}_{L^\infty_\tau H^{2j+2}} + \sqrt{\tau}\,\|u\|^{1/2}_{L^\infty_\tau H^{2j+1}} \|u\|^{1/2}_{L^\infty_\tau H^{2j+2}} \\ + \sqrt{\tau} \sum_{j_1+j_2+j_3=j} \|u\|_{L^\infty_\tau H^{2j_1+1}} \|u\|_{L^\infty_\tau H^{2j_2+1}} \|u\|_{L^\infty_\tau H^{2j_3+1}} \\ \lesssim \|u\|^{\frac{2j+\epsilon}{2k-1}}_{L^\infty_\tau H^{2k}} + \sqrt{\tau}\,\|u\|^{\frac{4j+1}{4k-2}}_{L^\infty_\tau H^{2k}} + \sqrt{\tau}\,\|u\|^{\frac{2j}{2k-1}}_{L^\infty_\tau H^{2k}}. \end{aligned} \tag{33}$$

We denote by I, II, III, IV the four terms on each line of the right-hand side in (15). We first estimate the term I. By developing the time derivatives ∂_t^k and ∂_t^{k-1}, and by using the Hölder inequality, we are reduced to estimating

$$\int_0^T \|\partial_t^{k_1} u\|_{L^2} \|\partial_t^{k_2} u\|_{L^6} \|\partial_t^{j_1} u\|_{W^{1,6}} \|\partial_t^{j_2} u\|_{W^{1,6}}\, ds, \tag{34}$$

where we can assume $k_1 \geq k_2$ and

$$j_1 + j_2 = k-1, \quad k_1 + k_2 = k.$$

Notice that by combining the Sobolev embedding $H^1(M^3) \subset L^6(M^3)$ with Proposition 3.3 for $d = 3$ and $p = 3$, and (24) we have

$$\begin{aligned}(34) &\lesssim \|u\|_{L^\infty_\tau H^{2k_1}} \|u\|_{L^\infty_\tau H^{2k_2+1}} \|\partial_t^{j_1} u\|_{L^2_\tau W^{1,6}} \|\partial_t^{j_2} u\|_{L^2_\tau W^{1,6}} \\ &\lesssim \|u\|_{L^\infty_\tau H^{2k}} \|\partial_t^{j_1} u\|_{L^2_\tau W^{1,6}} \|\partial_t^{j_2} u\|_{L^2_\tau W^{1,6}}, \end{aligned}$$

and we can continue the estimate by using (33). Indeed we should estimate $\|\partial_t^j u\|_{L^2_\tau W^{1,6}}$ by three terms on the right-hand side in (33). However, we can consider only the term that gives the worst growth with respect to the power of $\|u\|_{L^\infty_\tau H^{2k}}$ (i.e., only the second term on the right-hand side of (33), as all the other terms give a smaller power of $\|u\|_{L^\infty_\tau H^{2k}}$). Summarizing we get

$$(34) \lesssim \tau \|u\|^2_{L^\infty_\tau H^{2k}} + \|u\|^{\gamma}_{L^\infty_\tau H^{2k}}$$

for a suitable $\gamma \in (0, 2)$. Next we estimate the term II, which can be reduced to estimating the terms

$$\int_0^T \|\partial_t^{k_1} u\|_{L^2} \|\partial_t^{k_2} u\|_{L^6} \|\partial_t^{j} \Delta_g u\|_{L^6} \|\partial_t^{k-1-j} u\|_{L^6}, \tag{35}$$

where we can assume $k_1 \leq k_2$ and

$$j = 0, \ldots, k-2, \quad k_1 + k_2 = k.$$

By using the Sobolev embedding $H^1(M^3) \subset L^6(M^3)$ in conjunction with Proposition 3.3 we get

$$(35) \lesssim \|u\|_{L^\infty_\tau H^{2k_1}} \|\partial_t^{k_2} u\|_{L^2_\tau L^6} \|u\|_{L^\infty_\tau H^{2j+3}} \|\partial_t^{k-1-j} u\|_{L^2_\tau L^6}.$$

By using (32) and (24) we get

$$(35) \lesssim \|u\|_{L^\infty_\tau H^{2k_1}} \|u\|^{\frac{4k_2-1}{4k-2}}_{L^\infty_\tau H^{2k}} \|u\|_{L^\infty_\tau H^{2j+3}} \|u\|^{\frac{4(k-1-j)-1}{4k-2}}_{L^\infty_\tau H^{2k}} \lesssim \tau \|u\|^2_{L^\infty_\tau H^{2k}} + \|u\|^{\gamma}_{L^\infty_\tau H^{2k}},$$

where $\gamma \in (0,2)$. Concerning the term III we are reduced to

$$\int_0^T \|\partial_t^{j_1} u\|_{L^\infty} \|\partial_t^{j_2} u\|_{L^\infty} \|\partial_t^{k-j} u\|_{L^2} \|\partial_t^{k_1} u\|_{L^6} \|\partial_t^{k_2} u\|_{L^6} \|\partial_t^{k_3} u\|_{L^6},$$
$$j_1 + j_2 = j, \quad 0 \le j \le k-1, \quad k_1 + k_2 + k_3 = k-1. \tag{36}$$

By the Sobolev embeddings $H^1(M^3) \subset L^6(M^3)$ and $H^2(M^3) \subset L^\infty(M^3)$ and Proposition 3.3 we get

$$(36) \lesssim \|u\|_{L^\infty_\tau H^{2j_1+2}} \|u\|_{L^\infty_\tau H^{2j_2+2}} \|u\|_{L^\infty_\tau H^{2k-2j}} \|\partial_t^{k_1} u\|_{L^2_\tau L^6} \|\partial_t^{k_2} u\|_{L^2_\tau L^6} \|u\|_{L^\infty_\tau H^{2k_3+1}}.$$

By combining (32) with (24) we get

$$(36) \lesssim \tau \|u\|^2_{L^\infty_\tau H^{2k}} + \|u\|^\gamma_{L^\infty_\tau H^{2k}}$$

for $\gamma \in (0,2)$. Concerning IV, it is sufficient to estimate

$$\int_0^T \|\partial_t^k u\|_{L^2} \|\partial_t^{k-j} u\|_{L^6} \|\partial_t^{j_1} u\|_{L^6} \|\partial_t^{j_2} u\|_{L^6},$$
$$j_1 + j_2 = j, \quad 1 \le j \le k-1. \tag{37}$$

We can control it by using $H^1(M^3) \subset L^6(M^3)$ and Proposition 3.3:

$$(37) \lesssim \|u\|_{L^\infty_\tau H^{2k}} \|u\|_{L^\infty_\tau H^{2k-2j+1}} \|\partial_t^{j_1} u\|_{L^2_\tau L^6} \|\partial_t^{j_1} u\|_{L^2_\tau L^6}.$$

Again by (32) and (24) we get

$$(37) \lesssim \tau \|u\|^2_{L^\infty_\tau H^{2k}} + \|u\|^\gamma_{L^\infty_\tau H^{2k}}$$

for some $\gamma \in (0,2)$. □

In order to conclude the proof of Theorem 1.3, following the same argument as in the proof of Theorem 1.1, we have to split $\mathcal{E}_{2k}(u)$ as $\mathcal{E}_{2k}(u) = \|\partial_t^k u\|^2_{L^2} + \mathcal{R}_{2k}(u)$, where

$$\mathcal{R}_{2k}(u) = -\frac{1}{2} \int \left|\partial_t^{k-1} \nabla_g (|u|^2)\right|^2_g \operatorname{dvol}_g - \int \left|\partial_t^{k-1} (|u|^2 u)\right|^2 \operatorname{dvol}_g,$$

and we need to estimate the term $\mathcal{R}_{2k}(u)$, namely

$$|\mathcal{R}_{2k}(u)| \lesssim \|u\|^{\frac{4k-3}{2k-1}+\epsilon}_{H^{2k}} + \|u\|^{\frac{4k-5}{2k-1}+\epsilon}_{H^{2k}},$$

which is a version of (31) in three dimensions. Once we prove this estimate, the conclusion is similar to Theorem 1.1. Notice that

$$\int \left|\partial_t^{k-1} \nabla_g (|u|^2)\right|^2_g \operatorname{dvol}_g \lesssim \sum_{k_1+k_2=k-1} \|\partial_t^{k_1} u\|^2_{W^{1,2}} \|\partial_t^{k_2} u\|^2_{L^\infty}$$
$$\lesssim \sum_{k_1+k_2=k-1} \|u\|^2_{H^{2k_1+1}} \|u\|^{1-\epsilon}_{H^{2k_2+1}} \|u\|^{1+\epsilon}_{H^{2k_2+2}} \lesssim \|u\|^{\frac{4k-3}{2k-1}+\epsilon}_{H^{2k}},$$

where we used the estimate

$$\|v\|_{L^\infty} \lesssim \|v\|_{H^1}^{\frac{1-\epsilon}{2}} \|v\|_{H^2}^{\frac{1+\epsilon}{2}}, \tag{38}$$

which in turn follows by combining interpolation with Sobolev embedding, Proposition 3.3 and (24). Moreover we have

$$\begin{aligned}\int \left|\partial_t^{k-1}(|u|^2u)\right|^2 \operatorname{dvol}_g &\\ &\lesssim \sum_{j_1+j_2+j_3=k-1} \|\partial_t^{j_1}u\|_{L^2}^2 \|\partial_t^{j_2}u\|_{L^\infty}^2 \|\partial_t^{j_3}u\|_{L^\infty}^2\\ &\lesssim \sum_{j_1+j_2+j_3=k-1} \|u\|_{H^{2j_1}}^2 \|u\|_{H^{2j_2+1}}^{1-\epsilon} \|u\|_{H^{2j_3+1}}^{1-\epsilon} \|u\|_{H^{2j_2+2}}^{1+\epsilon} \|u\|_{H^{2j_3+2}}^{1+\epsilon} \lesssim \|u\|_{H^{2k}}^{\frac{4k-5}{2k-1}+\epsilon},\end{aligned}$$

where we used (38), Proposition 3.3 and (24).

6. Polynomial growth of H^2 for the subcubic NLS on M^3

Next we prove Theorem 1.6. We introduce the energy

$$\mathcal{F}_2(v(t,x)) = \int_{M^3} |\partial_t v|^2 \operatorname{dvol}_g - (p-1)\int_{M^3} |v|^{p-1}\left|\nabla_g|v|\right|^2 \operatorname{dvol}_g - \frac{p-1}{p}\int_{M^3} |v|^{2p} \operatorname{dvol}_g.$$

Proposition 6.1. *Let $u(t,x)$ be solution to* (11) *for $d=3$ and $2<p<3$. Then we have*

$$\begin{aligned}&\frac{d}{dt}\mathcal{F}_2u(t,x)\\ &\quad= (p-1)(p-3)\int_{M^3} |u|^{p-2}\,\partial_t|u|\left|\nabla_g|u|\right|^2 \operatorname{dvol}_g + 2(p-1)\int_{M^3} |u|^{p-2}\,\partial_t|u|\,|\nabla_g u|_g^2 \operatorname{dvol}_g.\end{aligned} \tag{39}$$

Proof. We start with the following computation:

$$\begin{aligned}\frac{d}{dt}\|\partial_t u\|_{L^2}^2 &= 2\operatorname{Re}(\partial_t^2 u, \partial_t u)\\ &= 2\operatorname{Re}\big(\partial_t(-\Delta_g u + |u|^{p-1}u), i\,\partial_t u\big)\\ &= 2\operatorname{Im}\int_{M^3} (\partial_t\nabla_g u, \partial_t\nabla_g u)_g \operatorname{dvol}_g + 2\operatorname{Re}\big(\partial_t(|u|^{p-1}u), i\,\partial_t u\big),\end{aligned}$$

where $(f,g) = \int_{M^3} f\bar{g} \operatorname{dvol}_g$. Since the first term vanishes, we get

$$\begin{aligned}\frac{d}{dt}\|\partial_t u\|_{L^2}^2 &= 2\operatorname{Re}\big(\partial_t(|u|^{p-1})u, i\,\partial_t u\big) + 2\operatorname{Re}\big(|u|^{p-1}\,\partial_t u, i\,\partial_t u\big)\\ &= 2\operatorname{Re}\big(\partial_t(|u|^{p-1})u, -\Delta_g u\big) + 2\operatorname{Re}\big(\partial_t(|u|^{p-1})u, |u|^{p-1}u\big)\\ &= 2\operatorname{Re}\big(\partial_t(|u|^{p-1})u, -\Delta_g u\big) + \frac{p-1}{p}\frac{d}{dt}\int_{M^3} |u|^{2p} \operatorname{dvol}_g.\end{aligned}$$

By using the identity

$$\Delta_g(|u|^2) = u\Delta_g\bar{u} + \bar{u}\Delta_g u + 2|\nabla_g u|_g^2,$$

we get

$$
\begin{aligned}
&2\,\mathrm{Re}\big(\partial_t(|u|^{p-1})u, -\Delta_g u\big)\\
&\quad= -(\partial_t|u|^{p-1}, \Delta_g|u|^2) + 2(\partial_t|u|^{p-1}, |\nabla_g u|_g^2)\\
&\quad= (\partial_t\nabla_g|u|^{p-1}, \nabla_g|u|^2) + 2(\partial_t|u|^{p-1}, |\nabla_g u|_g^2)\\
&\quad= 2(p-1)\big(\partial_t(|u|^{p-2}\nabla_g|u|), |u|\nabla_g|u|\big) + 2(\partial_t|u|^{p-1}, |\nabla_g u|_g^2)\\
&\quad= 2(p-1)\frac{d}{dt}(|u|^{p-2}\nabla_g|u|, |u|\nabla_g|u|) - 2(p-1)(|u|^{p-2}\nabla_g|u|, \partial_t|u|\nabla_g|u|)\\
&\qquad\qquad - 2(p-1)(|u|^{p-2}\nabla_g|u|, |u|\nabla_g\,\partial_t|u|) + 2(\partial_t|u|^{p-1}, |\nabla_g u|_g^2)\\
&\quad= 2(p-1)\frac{d}{dt}(|u|^{p-2}\nabla_g|u|, |u|\nabla_g|u|) - 2(p-1)(|u|^{p-2}\nabla_g|u|, \partial_t|u|\nabla_g|u|)\\
&\qquad - (p-1)\frac{d}{dt}\big(|u|^{p-1}, \big|\nabla_g|u|\big|^2\big) + (p-1)\big(\partial_t|u|^{p-1}, \big|\nabla_g|u|\big|^2\big) + 2(\partial_t|u|^{p-1}, |\nabla_g u|_g^2). \quad \square
\end{aligned}
$$

The following proposition is a substitute for Proposition 5.1 in the subcubic case.

Proposition 6.2. *We have for every* $\tau \in (0,1)$

$$
\int_0^\tau |\textit{right-hand side of } (39)|\, ds \lesssim \tau \|u\|^{\frac{p+5}{4}}_{L^\infty_\tau H^2} + \|u\|^{\gamma}_{L^\infty_\tau H^2}
$$

for some $\gamma \in \big(0, \frac{p+5}{4}\big)$.

Proof. We can write the terms on the right-hand side of (39) as I and II. We estimate I and the estimate of II is similar. We estimate I as follows (we shall use the diamagnetic inequality in order to remove $|\cdot|$ inside the derivatives ∇_g and ∂_t) by the Hölder inequality:

$$
|I| \lesssim \|\partial_t u\|_{L^\infty_\tau L^2}\, \|u\|^2_{L^2_\tau W^{1,\frac{12}{5-p}}}\, \|u\|^{p-2}_{L^6} \lesssim \tau^{\frac{6-2p}{8}} \|\partial_t\, u\|_{L^\infty_\tau L^2}\, \|u\|^2_{L^{\frac{8}{p+1}}_\tau W^{1,\frac{12}{5-p}}},
$$

where the pair $\big(\frac{8}{p+1}, \frac{12}{5-p}\big)$ is Strichartz admissible. Notice that by using the equation solved by $u(t,x)$, we are allowed to replace $\|\partial_t u\|_{L^\infty_\tau L^2}$ with $\|u\|_{L^\infty_\tau H^2}$ and hence

$$
|I| \lesssim \tau^{\frac{6-2p}{8}} \|u\|_{L^\infty_\tau H^2}\, \|u\|^2_{L^{\frac{8}{p+1}}_\tau W^{1,\frac{12}{5-p}}}.
$$

Next notice that we have the bound

$$
\|u\|_{L^{\frac{8}{p+1}}_\tau W^{1,\frac{12}{5-p}}} \lesssim \|u\|^{\frac{3-p}{4}}_{L^\infty_\tau H^1}\, \|u\|^{\frac{p+1}{4}}_{L^2_\tau W^{1,6}},
$$

and hence due to the conservation of the energy, we can continue the estimate above as

$$
|I| \lesssim \tau^{\frac{6-2p}{8}} \|u\|_{L^\infty_\tau H^2}\, \|u\|^{\frac{p+1}{2}}_{L^2_\tau W^{1,6}}.
$$

We can continue the estimate by using the Strichartz estimates (33) for $j = 0$ (which are still available for solutions to the subcubic NLS):

$$
|I| \lesssim \tau \|u\|_{L^\infty_\tau H^2}\, \|u\|^{\frac{p+1}{4}}_{L^\infty_\tau H^2} + \|u\|^{\gamma}_{L^\infty_\tau H^2}
$$

for some $\gamma \in \left(0, \frac{p+5}{4}\right)$ (indeed we have estimated the term $\|u\|_{L^2_\tau W^{1,6}}$ with the middle term on the right-hand side in (33) since it is the one that involves the larger power of $\|u\|_{L^\infty_\tau H^2}$, and the lower powers are absorbed in the term $\|u\|^\gamma_{L^\infty_\tau H^2}$). □

Integrating (39) on $[0, \tau]$ and arguing exactly as in the proofs of Theorems 1.1 and 1.3, we get the bound

$$\|u(\tau)\|^2_{H^2(M^3)} - \|u(0)\|^2_{H^2(M^3)} \lesssim \tau \|u\|^{\frac{p+5}{4}}_{L^\infty((0,\tau);H^2(M^3))} + \|u\|^\gamma_{L^\infty((0,\tau);H^2(M^3))}$$

for some $\gamma \in \left(0, \frac{p+5}{4}\right)$. This is sufficient to conclude Remark 1.7.

7. Growth of odd Sobolev norms H^{2k+1}

We point out that if we assume the initial datum φ to be H^{2k+2}, then the estimate

$$\sup_{(0,T)} \|u(t,x)\|_{H^{2k+1}(M^2)} \le C(\max\{1,T\})^{\frac{2k}{1-2s_0}+\epsilon},$$

stated in Theorem 1.1, follows by interpolation between the following bounds, which have been already proved by looking at growth of even Sobolev norms:

$$\sup_{(0,T)} \|u(t,x)\|_{H^{2k+2}(M^2)} \le C(\max\{1,T\})^{\frac{2k+1}{1-2s_0}+\epsilon},$$
$$\sup_{(0,T)} \|u(t,x)\|_{H^{2k}(M^2)} \le C(\max\{1,T\})^{\frac{2k-1}{1-2s_0}+\epsilon}.$$

A similar argument follows in order to prove Theorem 1.3 for $m = 2k+1$.

However, the main point in this section is that we assume the initial datum φ to be only in H^{2k+1}, and hence the argument above cannot be applied.

The proofs of Theorems 1.1 and 1.3 (which have been proved in the case $m = 2k$) can be adapted to the case $m = 2k+1$ by using the modified energies

$$\begin{aligned}\mathcal{E}_{2k+1}(u) = {} & \frac{1}{2}\|\partial_t^k \nabla_g u\|^2_{L^2} + \frac{1}{2}\int |u|^{p-1} |\partial_t^k u|^2 \,\mathrm{dvol}_g + \frac{p-1}{8}\int |u|^{p-3} \left|\partial_t^k (|u|^2)\right|^2 \mathrm{dvol}_g \\ & - \mathrm{Re}\sum_{j=1}^{k-1} c_j \int \partial_t^j u \, \partial_t^{k-j}(|u|^{p-1}) \, \partial_t^k \bar{u} \,\mathrm{dvol}_g \\ & - \sum_{j=1}^{k-1} c_j \int \partial_t^{k-j}(|u|^{p-3}) \, \partial_t^j (|u|^2) \, \partial_t^k (|u|^2) \,\mathrm{dvol}_g. \end{aligned} \tag{40}$$

Indeed we have the following proposition, from which one may conclude the proof of Theorems 1.1 and 1.3 in the case $m = 2k+1$, exactly as we did in the case $m = 2k$. We leave details to the reader.

Proposition 7.1. *Let $u(t,x)$ be a solution to* (1) *with initial datum φ in H^{2k+1}. Then we have the identity*

$$\begin{aligned}\frac{d}{dt}\mathcal{E}_{2k+1}(u(t,x)) &= \frac{1}{2}\int \partial_t(|u|^{p-1})|\partial_t^k u|^2\,\mathrm{dvol}_g - \mathrm{Re}\sum_{j=1}^{k-1} c_j\int \partial_t^{j+1}u\,\partial_t^{k-j}(|u|^{p-1})\,\partial_t^k\bar{u}\,\mathrm{dvol}_g\\ &\quad -\mathrm{Re}\sum_{j=1}^{k-1} c_j\int \partial_t^{j}u\,\partial_t^{k-j+1}(|u|^{p-1})\,\partial_t^k\bar{u}\,\mathrm{dvol}_g + \frac{p-1}{8}\int_{M^2}\partial_t(|u|^{p-3})\big|\partial_t^k(|u|^2)\big|^2\,\mathrm{dvol}_g\\ &\quad +\sum_{j=1}^{k-1} c_j\int \partial_t^{k-j+1}(|u|^{p-3})\,\partial_t^j(|u|^2)\,\partial_t^k(|u|^2)\,\mathrm{dvol}_g\\ &\quad +\sum_{j=1}^{k-1} c_j\int \partial_t^{k-j}(|u|^{p-3})\,\partial_t^{j+1}(|u|^2)\,\partial_t^k(|u|^2)\,\mathrm{dvol}_g\\ &\quad +\sum_{j=1}^{k} c_j\int \partial_t^{k}(|u|^{p-1})\,\partial_t^{j}u\,\partial_t^{k+1-j}\bar{u}\,\mathrm{dvol}_g,\end{aligned}$$

where $c_j \in \mathbb{R}$ are explicit real numbers that can change in different lines.

Proof. First of all notice that we have

$$\begin{aligned}\mathrm{Re}(i\,\partial_t^{k+1}u, \partial_t^k u) &= \mathrm{Re}(\partial_t^k(-\Delta_g u), \partial_t^k u) + \mathrm{Re}(\partial_t^k(u|u|^{p-1}), \partial_t^k u)\\ &= \|\partial_t^k\nabla_g u\|_{L^2}^2 + \mathrm{Re}(\partial_t^k(u|u|^{p-1}), \partial_t^k u).\end{aligned}$$

Due to the identity above and by taking the time derivative, we get

$$\begin{aligned}\frac{d}{dt}\big(\|\partial_t^k\nabla_g u\|_{L^2}^2 + \mathrm{Re}(\partial_t^k(u|u|^{p-1}), \partial_t^k u)\big) &= \frac{d}{dt}\,\mathrm{Re}(i\,\partial_t^{k+1}u, \partial_t^k u) = \mathrm{Re}(i\,\partial_t^{k+2}u, \partial_t^k u)\\ &= \mathrm{Re}(\partial_t^{k+1}(-\Delta_g u), \partial_t^k u) + \mathrm{Re}(\partial_t^{k+1}(|u|^{p-1}u), \partial_t^k u)\\ &= \frac{1}{2}\frac{d}{dt}\|\partial_t^k\nabla_g u\|_{L^2}^2 + \mathrm{Re}(\partial_t^{k+1}(|u|^{p-1}u), \partial_t^k u).\end{aligned}$$

Next we focus on the second term on the right-hand side:

$$\begin{aligned}&\mathrm{Re}(\partial_t^{k+1}(|u|^{p-1}u), \partial_t^k u)\\ &\quad= \frac{d}{dt}\,\mathrm{Re}(\partial_t^k(|u|^{p-1}u), \partial_t^k u) - \mathrm{Re}(\partial_t^k(|u|^{p-1}u), \partial_t^{k+1}u)\\ &\quad= \frac{d}{dt}\,\mathrm{Re}(\partial_t^k(|u|^{p-1}u), \partial_t^k u) - \mathrm{Re}(\partial_t^k(|u|^{p-1})u, \partial_t^{k+1}u) - \mathrm{Re}(|u|^{p-1}\partial_t^k u, \partial_t^{k+1}u)\\ &\qquad+ \mathrm{Re}\sum_{j=1}^{k-1} c_j(\partial_t^j u\,\partial_t^{k-j}(|u|^{p-1}), \partial_t^{k+1}u)\\ &\quad= \frac{d}{dt}\,\mathrm{Re}(\partial_t^k(|u|^{p-1}u), \partial_t^k u) - \mathrm{Re}(\partial_t^k(|u|^{p-1})u, \partial_t^{k+1}u) - \frac{1}{2}\frac{d}{dt}\int |u|^{p-1}|\partial_t^k u|^2\,\mathrm{dvol}_g\\ &\qquad+\frac{1}{2}\int \partial_t(|u|^{p-1})|\partial_t^k u|^2\,\mathrm{dvol}_g + \mathrm{Re}\sum_{j=1}^{k-1} c_j(\partial_t^j u\,\partial_t^{k-j}(|u|^{p-1}), \partial_t^{k+1}u)\end{aligned}$$

$$= \frac{d}{dt}\operatorname{Re}(\partial_t^k(|u|^{p-1}u), \partial_t^k u) - \operatorname{Re}(\partial_t^k(|u|^{p-1})u, \partial_t^{k+1}u) - \frac{1}{2}\frac{d}{dt}\int |u|^{p-1}|\partial_t^k u|^2 \,\mathrm{dvol}_g$$
$$+ \frac{1}{2}\int \partial_t(|u|^{p-1})|\partial_t^k u|^2 \,\mathrm{dvol}_g + \frac{d}{dt}\operatorname{Re}\sum_{j=1}^{k-1} c_j(\partial_t^j u\, \partial_t^{k-j}(|u|^{p-1}), \partial_t^k u)$$
$$- \operatorname{Re}\sum_{j=1}^{k-1} c_j(\partial_t^{j+1} u\, \partial_t^{k-j}(|u|^{p-1}), \partial_t^k u) - \operatorname{Re}\sum_{j=1}^{k-1} c_j(\partial_t^j u\, \partial_t^{k-j+1}(|u|^{p-1}), \partial_t^k u).$$

Next we deal with the third term on the right-hand side:

$$-\operatorname{Re}(\partial_t^k(|u|^{p-1})u, \partial_t^{k+1}u)$$
$$= -\frac{1}{2}\int \partial_t^k(|u|^{p-1})\,\partial_t^{k+1}(|u|^2)\,\mathrm{dvol}_g + \sum_{j=1}^{k} c_j \int \partial_t^k(|u|^{p-1})\,\partial_t^j u\,\partial_t^{k+1-j}\bar{u}\,\mathrm{dvol}_g,$$

and we notice that $\partial_t^k(|u|^{p-1}) = \frac{1}{2}(p-1)\,\partial_t^{k-1}(\partial_t(|u|^2)|u|^{p-3})$. Hence we can continue the identity above as follows:

$$\cdots = -\frac{p-1}{4}\int |u|^{p-3}\,\partial_t^k(|u|^2)\,\partial_t^{k+1}(|u|^2)\,\mathrm{dvol}_g + \sum_{j=1}^{k-1} c_j \int \partial_t^{k-j}(|u|^{p-3})\,\partial_t^j(|u|^2)\,\partial_t^{k+1}(|u|^2)\,\mathrm{dvol}_g$$
$$+ \sum_{j=1}^{k} c_j \int \partial_t^k(|u|^{p-1})\,\partial_t^j u\,\partial_t^{k+1-j}\bar{u}\,\mathrm{dvol}_g$$
$$= -\frac{p-1}{8}\frac{d}{dt}\int |u|^{p-3}\big|\partial_t^k(|u|^2)\big|^2\,\mathrm{dvol}_g + \frac{p-1}{8}\int \partial_t(|u|^{p-3})\big|\partial_t^k(|u|^2)\big|^2\,\mathrm{dvol}_g$$
$$+ \sum_{j=1}^{k-1} c_j \int \partial_t^{k-j}(|u|^{p-3})\,\partial_t^j(|u|^2)\,\partial_t^{k+1}(|u|^2)\,\mathrm{dvol}_g + \sum_{j=1}^{k} c_j \int \partial_t^k(|u|^{p-1})\,\partial_t^j u\,\partial_t^{k+1-j}\bar{u}\,\mathrm{dvol}_g.$$

Then by elementary considerations

$$\cdots = -\frac{p-1}{8}\frac{d}{dt}\int |u|^{p-3}\big|\partial_t^k(|u|^2)\big|^2\,\mathrm{dvol}_g + \frac{p-1}{8}\int \partial_t(|u|^{p-3})\big|\partial_t^k(|u|^2)\big|^2\,\mathrm{dvol}_g$$
$$+ \frac{d}{dt}\sum_{j=1}^{k-1} c_j \int \partial_t^{k-j}(|u|^{p-3})\,\partial_t^j(|u|^2)\,\partial_t^k(|u|^2)\,\mathrm{dvol}_g + \sum_{j=1}^{k-1} c_j \int \partial_t^{k-j+1}(|u|^{p-3})\,\partial_t^j(|u|^2)\,\partial_t^k(|u|^2)\,\mathrm{dvol}_g$$
$$+ \sum_{j=1}^{k-1} c_j \int \partial_t^{k-j}(|u|^{p-3})\,\partial_t^{j+1}(|u|^2)\,\partial_t^k(|u|^2)\,\mathrm{dvol}_g + \sum_{j=1}^{k} c_j \int \partial_t^k(|u|^{p-1})\,\partial_t^j u\,\partial_t^{k+1-j}\bar{u}\,\mathrm{dvol}_g. \qquad \square$$

Acknowledgement

The authors are grateful to the referee for interesting remarks and suggestions to improve this paper.

References

[Bouclet and Tzvetkov 2007] J.-M. Bouclet and N. Tzvetkov, "Strichartz estimates for long range perturbations", *Amer. J. Math.* **129**:6 (2007), 1565–1609. MR Zbl

[Bourgain 1993] J. Bourgain, "Fourier transform restriction phenomena for certain lattice subsets and applications to nonlinear evolution equations, II: The KdV-equation", *Geom. Funct. Anal.* **3**:3 (1993), 209–262. MR Zbl

[Bourgain 1996] J. Bourgain, "On the growth in time of higher Sobolev norms of smooth solutions of Hamiltonian PDE", *Internat. Math. Res. Notices* 6 (1996), 277–304. MR Zbl

[Bourgain 1999a] J. Bourgain, *Global solutions of nonlinear Schrödinger equations*, American Mathematical Society Colloquium Publications **46**, Amer. Math. Soc., Providence, RI, 1999. MR Zbl

[Bourgain 1999b] J. Bourgain, "On growth of Sobolev norms in linear Schrödinger equations with smooth time dependent potential", *J. Anal. Math.* **77** (1999), 315–348. MR Zbl

[Burq et al. 2003] N. Burq, P. Gérard, and N. Tzvetkov, "The Cauchy problem for the nonlinear Schrödinger equation on a compact manifold", *J. Nonlinear Math. Phys.* **10**:suppl. 1 (2003), 12–27. MR

[Burq et al. 2004] N. Burq, P. Gérard, and N. Tzvetkov, "Strichartz inequalities and the nonlinear Schrödinger equation on compact manifolds", *Amer. J. Math.* **126**:3 (2004), 569–605. MR Zbl

[Chiron and Rousset 2009] D. Chiron and F. Rousset, "Geometric optics and boundary layers for nonlinear-Schrödinger equations", *Comm. Math. Phys.* **288**:2 (2009), 503–546. MR Zbl

[Colliander et al. 2010] J. Colliander, M. Keel, G. Staffilani, H. Takaoka, and T. Tao, "Transfer of energy to high frequencies in the cubic defocusing nonlinear Schrödinger equation", *Invent. Math.* **181**:1 (2010), 39–113. MR Zbl

[Colliander et al. 2012] J. Colliander, S. Kwon, and T. Oh, "A remark on normal forms and the 'upside-down' I-method for periodic NLS: growth of higher Sobolev norms", *J. Anal. Math.* **118**:1 (2012), 55–82. MR Zbl

[Delort 2014] J.-M. Delort, "Growth of Sobolev norms for solutions of time dependent Schrödinger operators with harmonic oscillator potential", *Comm. Partial Differential Equations* **39**:1 (2014), 1–33. MR Zbl

[Gérard and Grellier 2016] P. Gérard and S. Grellier, "On the growth of Sobolev norms for the cubic Szegő equation", exposé 11, 20 pp. in *Séminaire Laurent Schwartz: équations aux dérivées partielles et applications,* 2014–2015, Ed. Éc. Polytech., Palaiseau, 2016. MR

[Grellier and Gerard 2015] S. Grellier and P. Gerard, "The cubic Szegő equation and Hankel operators", preprint, 2015. arXiv

[Guardia 2014] M. Guardia, "Growth of Sobolev norms in the cubic nonlinear Schrödinger equation with a convolution potential", *Comm. Math. Phys.* **329**:1 (2014), 405–434. MR Zbl

[Guardia and Kaloshin 2015] M. Guardia and V. Kaloshin, "Growth of Sobolev norms in the cubic defocusing nonlinear Schrödinger equation", *J. Eur. Math. Soc.* (*JEMS*) **17**:1 (2015), 71–149. MR Zbl

[Guardia et al. 2016] M. Guardia, E. Haus, and M. Procesi, "Growth of Sobolev norms for the analytic NLS on $\mathbb{T}^2$", *Adv. Math.* **301** (2016), 615–692. MR Zbl

[Hani 2014] Z. Hani, "Long-time instability and unbounded Sobolev orbits for some periodic nonlinear Schrödinger equations", *Arch. Ration. Mech. Anal.* **211**:3 (2014), 929–964. MR Zbl

[Hani et al. 2015] Z. Hani, B. Pausader, N. Tzvetkov, and N. Visciglia, "Modified scattering for the cubic Schrödinger equation on product spaces and applications", *Forum Math. Pi* **3** (2015), art. id. e4. MR Zbl

[Haus and Procesi 2015] E. Haus and M. Procesi, "Growth of Sobolev norms for the quintic NLS on T^2", *Anal. PDE* **8**:4 (2015), 883–922. MR Zbl

[Hunter et al. 2015] J. K. Hunter, M. Ifrim, D. Tataru, and T. K. Wong, "Long time solutions for a Burgers–Hilbert equation via a modified energy method", *Proc. Amer. Math. Soc.* **143**:8 (2015), 3407–3412. MR Zbl

[Koch and Tataru 2016] H. Koch and D. Tataru, "Conserved energies for the cubic NLS in 1-d", preprint, 2016. arXiv

[Koch and Tzvetkov 2003] H. Koch and N. Tzvetkov, "On the local well-posedness of the Benjamin–Ono equation in $H^s(\mathbb{R})$", *Int. Math. Res. Not.* **2003**:26 (2003), 1449–1464. MR Zbl

[Kwon 2008] S. Kwon, "On the fifth-order KdV equation: local well-posedness and lack of uniform continuity of the solution map", *J. Differential Equations* **245**:9 (2008), 2627–2659. MR Zbl

[Ozawa and Visciglia 2016] T. Ozawa and N. Visciglia, "An improvement on the Brézis–Gallouët technique for 2D NLS and 1D half-wave equation", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **33**:4 (2016), 1069–1079. MR Zbl

[Raphaël and Szeftel 2009] P. Raphaël and J. Szeftel, "Standing ring blow up solutions to the N-dimensional quintic nonlinear Schrödinger equation", *Comm. Math. Phys.* **290**:3 (2009), 973–996. MR Zbl

[Sohinger 2011a] V. Sohinger, "Bounds on the growth of high Sobolev norms of solutions to nonlinear Schrödinger equations on $\mathbb{R}$", *Indiana Univ. Math. J.* **60**:5 (2011), 1487–1516. MR Zbl

[Sohinger 2011b] V. Sohinger, "Bounds on the growth of high Sobolev norms of solutions to nonlinear Schrödinger equations on S^1", *Differential Integral Equations* **24**:7-8 (2011), 653–718. MR Zbl

[Sohinger 2012] V. Sohinger, "Bounds on the growth of high Sobolev norms of solutions to 2D Hartree equations", *Discrete Contin. Dyn. Syst.* **32**:10 (2012), 3733–3771. MR Zbl

[Staffilani 1997] G. Staffilani, "On the growth of high Sobolev norms of solutions for KdV and Schrödinger equations", *Duke Math. J.* **86**:1 (1997), 109–142. MR Zbl

[Staffilani and Tataru 2002] G. Staffilani and D. Tataru, "Strichartz estimates for a Schrödinger operator with nonsmooth coefficients", *Comm. Partial Differential Equations* **27**:7-8 (2002), 1337–1372. MR Zbl

[Thirouin 2017] J. Thirouin, "On the growth of Sobolev norms of solutions of the fractional defocusing NLS equation on the circle", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **34**:2 (2017), 509–531. MR Zbl

[Tsutsumi 1989] M. Tsutsumi, "On smooth solutions to the initial-boundary value problem for the nonlinear Schrödinger equation in two space dimensions", *Nonlinear Anal.* **13**:9 (1989), 1051–1056. MR Zbl

[Xu 2015] H. Xu, "Unbounded Sobolev trajectories and modified scattering theory for a wave guide nonlinear Schrödinger equation", preprint, 2015. arXiv

[Zhong 2008] S. Zhong, "The growth in time of higher Sobolev norms of solutions to Schrödinger equations on compact Riemannian manifolds", *J. Differential Equations* **245**:2 (2008), 359–376. MR Zbl

Received 29 Jul 2016. Revised 5 Mar 2017. Accepted 24 Apr 2017.

FABRICE PLANCHON: fabrice.planchon@unice.fr
Université Côte d'Azur, CNRS, LJAD, Parc Valrose, 06108 Nice, France

NIKOLAY TZVETKOV: nikolay.tzvetkov@u-cergy.fr
Department of Mathematics, Université de Cergy-Pontoise, 2, Avenue A. Chauvin, 95302 Cergy-Pontoise Cedex, France

NICOLA VISCIGLIA: viscigli@dm.unipi.it
Dipartimento di Matematica, Università Degli Studi di Pisa, Largo Bruno Pontecorvo 5, I-56127 Pisa, Italy

ANALYSIS AND PDE
Vol. 10, No. 5, 2017
dx.doi.org/10.2140/apde.2017.10.1149

A SUFFICIENT CONDITION FOR GLOBAL EXISTENCE OF SOLUTIONS TO A GENERALIZED DERIVATIVE NONLINEAR SCHRÖDINGER EQUATION

NORIYOSHI FUKAYA, MASAYUKI HAYASHI AND TAKAHISA INUI

We give a sufficient condition for global existence of the solutions to a generalized derivative nonlinear Schrödinger equation (gDNLS) by a variational argument. The variational argument is applicable to a cubic derivative nonlinear Schrödinger equation (DNLS). For (DNLS), Wu (2015) proved that the solution with the initial data u_0 is global if $\|u_0\|_{L^2}^2 < 4\pi$ by the sharp Gagliardo–Nirenberg inequality. The variational argument gives us another proof of the global existence for (DNLS). Moreover, by the variational argument, we can show that the solution to (DNLS) is global if the initial data u_0 satisfies $\|u_0\|_{L^2}^2 = 4\pi$ and the momentum $P(u_0)$ is negative.

1. Introduction

1A. ***Background.*** The following equation is known as a derivative nonlinear Schrödinger equation:

$$i\partial_t v + \partial_x^2 v + i\partial_x(|v|^2 v) = 0, \quad (t,x) \in \mathbb{R}\times\mathbb{R}. \tag{1-1}$$

This equation appears in plasma physics [Mio et al. 1976; Mjølhus 1976] and as a model for ultrashort optical pulses [Moses et al. 2007]. Using the gauge transformation

$$u(t,x) = v(t,x)\exp\left(\frac{i}{2}\int_{-\infty}^{x}|v(t,x)|^2\,dx\right),$$

we get a Hamiltonian form of (1-1):

$$i\partial_t u + \partial_x^2 u + i|u|^2\partial_x u = 0, \quad (t,x) \in \mathbb{R}\times\mathbb{R}. \tag{DNLS}$$

Namely, this equation can be written as $i\partial_t u = E'(u)$ (see below for the definition of the Hamiltonian E). The Cauchy problem for (DNLS) (or equivalently (1-1)) has been studied by many researchers. It is known that (DNLS) is locally well-posed in the energy space $H^1(\mathbb{R})$. See [Tsutsumi and Fukuda 1980; Hayashi

MSC2010: 35Q55.
Keywords: variational structure, generalized derivative nonlinear Schrödinger equation, global existence.

and Ozawa 1992; Hayashi 1993; Hayashi and Ozawa 1994a; 1994b]. Hayashi and Ozawa [1994a] proved that the solution is global if $\|u_0\|_{L^2}^2 < 2\pi$. See also [Ozawa 1996]. Wu [2013; 2015] proved that it holds if $\|u_0\|_{L^2}^2 < 4\pi$. Recently, Miao, Tang, and Xu obtained the global well-posedness by a variational argument (see the remark on page 1156). For the initial data with low regularity, there are also many references. Takaoka [1999] proved that (DNLS) is locally well-posed in $H^s(\mathbb{R})$ when $s \geq \frac{1}{2}$ by the Fourier restricted method. Biagioni and Linares [2001] proved that the solution map from $H^s(\mathbb{R})$ to $C([-T, T] : H^s(\mathbb{R}))$, where $T > 0$, for (DNLS) is not locally uniformly continuous when $s < \frac{1}{2}$. Colliander, Keel, Staffilani, Takaoka, and Tao [Colliander et al. 2002] proved that the H^s-solution is global if $\|u_0\|_{L^2}^2 < 2\pi$ when $s > \frac{1}{2}$ by the I-method (see also [Colliander et al. 2001; Takaoka 2001]). Recently, Miao, Wu, and Xu [Miao et al. 2011] showed that $H^{1/2}$-solution is global if $\|u_0\|_{L^2}^2 < 2\pi$. Guo and Wu [2017] improved their result; that is, they proved that $H^{1/2}$-solution is global if $\|u_0\|_{L^2}^2 < 4\pi$. The orbital stability of solitary waves has been also studied. It is known that (DNLS) has a two-parameter family of the solitary waves $u_{\omega,c}(t, x) = e^{i\omega t}\phi_{\omega,c}(x - ct)$, where (ω, c) satisfies $\omega > c^2/4$, or $\omega = c^2/4$ and $c > 0$ (see below for the explicit formula of $\phi_{\omega,c}$). Guo and Wu [1995] proved that the solitary waves $u_{\omega,c}$ are orbitally stable when $\omega > c^2/4$ and $c < 0$ by the abstract theory of Grillakis, Shatah, and Strauss [Grillakis et al. 1987; 1990] and the spectral analysis of the linearized operators. Colin and Ohta [2006] proved that the solitary waves $u_{\omega,c}$ are orbitally stable when $\omega > c^2/4$ by characterizing the solitary waves from the viewpoint of a variational structure. The case of $\omega = c^2/4$ and $c > 0$ was treated by Kwon and Wu [2016]. Recently, the stability of the multisolitons was studied by Miao, Tang, and Xu [Miao et al. 2017b] and Le Coz and Wu [2016].

To understand the structural properties of (DNLS), Liu, Simpson, and Sulem [Liu et al. 2013] introduced an extension of (DNLS) with general power nonlinearity. The generalized derivative nonlinear Schrödinger equation is

$$\begin{cases} i\partial_t u + \partial_x^2 u + i|u|^{2\sigma}\partial_x u = 0, & (t, x) \in \mathbb{R} \times \mathbb{R}, \\ u(0, x) = u_0(x), & x \in \mathbb{R}, \end{cases} \tag{gDNLS}$$

where $\sigma > 0$. Equation (gDNLS) is invariant under the scaling transformation

$$u_\gamma(t, x) := \gamma^{1/(2\sigma)} u(\gamma^2 t, \gamma x), \quad \gamma > 0.$$

This implies that its critical Sobolev exponent is $s_c = \frac{1}{2} - 1/(2\sigma)$. In particular, (DNLS) is L^2-critical. Liu et al. [2013] investigated the orbital stability of a two-parameter family of solitary waves

$$u_{\omega,c}(t, x) = e^{i\omega t}\phi_{\omega,c}(x - ct),$$

where (ω, c) satisfies $\omega > c^2/4$, or $\omega = c^2/4$ and $c > 0$, and

$$\phi_{\omega,c}(x) = \Phi_{\omega,c}(x) \exp\left(i\frac{c}{2}x - \frac{i}{2\sigma + 2}\int_0^x \Phi_{\omega,c}(y)^{2\sigma}\, dy \right), \tag{1-2}$$

$$\Phi_{\omega,c}(x) = \begin{cases} \left\{ \dfrac{(\sigma + 1)(4\omega - c^2)}{2\sqrt{\omega}\cosh(\sigma\sqrt{4\omega - c^2}x) - c} \right\}^{1/(2\sigma)} & \text{if } \omega > c^2/4, \\ \left\{ \dfrac{2(\sigma + 1)c}{\sigma^2(cx)^2 + 1} \right\}^{1/(2\sigma)} & \text{if } \omega = c^2/4 \text{ and } c > 0. \end{cases} \tag{1-3}$$

We note that $\Phi_{\omega,c}$ is the positive even solution of

$$-\Phi'' + (\omega - \tfrac{1}{4}c^2)\Phi + \tfrac{1}{2}c|\Phi|^{2\sigma}\Phi - \frac{2\sigma+1}{(2\sigma+2)^2}|\Phi|^{4\sigma}\Phi = 0, \quad x \in \mathbb{R}, \tag{1-4}$$

and then the complex-valued function $\phi_{\omega,c}$ satisfies

$$-\phi'' + \omega\phi + ic\phi' - i|\phi|^{2\sigma}\phi' = 0, \quad x \in \mathbb{R}.$$

Liu et al. [2013] proved that the solitary waves are orbitally stable if $-2\sqrt{\omega} < c < 2z_0\sqrt{\omega}$, and orbitally unstable if $2z_0\sqrt{\omega} < c < 2\sqrt{\omega}$ when $1 < \sigma < 2$, where the constant $z_0 = z_0(\sigma) \in (-1, 1)$ is the solution of

$$F_\sigma(z) := (\sigma-1)^2\left\{\int_0^\infty (\cosh y - z)^{-1/\sigma}\, dy\right\}^2 - \left\{\int_0^\infty (\cosh y - z)^{-1/\sigma-1}(z\cosh y - 1)\, dy\right\}^2 = 0.$$

Moreover, they also proved that the solitary waves for all $\omega > c^2/4$ are orbitally unstable when $\sigma \geq 2$ and orbitally stable when $0 < \sigma < 1$. Recently, Fukaya [2016] proved that the solitary waves are orbitally unstable if $c = 2z_0\sqrt{\omega}$ when $\frac{7}{6} < \sigma < 2$. More recently, Tang and Xu investigated stability of the sum of two solitary waves for (gDNLS) (see [Tang and Xu 2017] for more details). Before Liu et al. [2013], Hao [2007] considered (gDNLS) and proved the local well-posedness in $H^{1/2}(\mathbb{R})$ when $\sigma \geq \frac{5}{2}$. Santos [2015] proved the existence and uniqueness of a solution $u \in C([0, T]; H^{1/2}(\mathbb{R}))$ for sufficiently small initial data when $\sigma > 1$. Recently, Hayashi and Ozawa [2016] proved local well-posedness in $H^1(\mathbb{R})$ when $\sigma \geq 1$ and that the following quantities are conserved:

$$E(u) := \tfrac{1}{2}\|\partial_x u\|_{L^2}^2 - \frac{1}{2\sigma+2}\operatorname{Re}\int_{\mathbb{R}} i|u|^{2\sigma}\bar{u}\partial_x u\, dx, \tag{Energy}$$

$$M(u) := \|u\|_{L^2}^2, \tag{Mass}$$

$$P(u) := \operatorname{Re}\int_{\mathbb{R}} i\partial_x u\bar{u}\, dx. \tag{Momentum}$$

Moreover, they proved global well-posedness for small initial data. They also constructed global solutions for any initial data in $H^1(\mathbb{R})$ in the case $0 < \sigma < 1$ (L^2-subcritical case). However, in the case $\sigma \geq 1$ (L^2-critical or supercritical case), there has been no global existence result for large data. In the present paper, we investigate global well-posedness for (gDNLS) in the case $\sigma \geq 1$ by a variational argument. More precisely, we give a variational characterization of solitary waves and a sufficient condition for global existence of solutions to (gDNLS) by using the characterization. Such an argument was done for nonlinear hyperbolic partial differential equations by Sattinger [1968] (see also [Tsutsumi 1972; Payne and Sattinger 1975]). Our argument is also applicable to (DNLS). Indeed, the variational argument gives another proof of the result by Wu [2015]. Moreover, we prove that the solution of (DNLS) is global if the initial data u_0 satisfies $\|u_0\|_{L^2}^2 = 4\pi$ and $P(u_0) < 0$.

1B. ***Main results.*** To state our main results, we introduce some notations. Let (ω, c) satisfy

$$\omega > c^2/4 \quad \text{or} \quad \omega = c^2/4 \text{ and } c > 0. \tag{1-5}$$

For (ω, c) satisfying (1-5), we define

$$S_{\omega,c}(\varphi) := E(\varphi) + \tfrac{1}{2}\omega M(\varphi) + \tfrac{1}{2}cP(\varphi).$$

We denote the nonlinear term by

$$N(\varphi) := \mathrm{Re}\int_{\mathbb{R}} i|\varphi|^{2\sigma}\bar{\varphi}\partial_x\varphi\,dx.$$

We define

$$\widetilde{S}_{\omega,c}(\psi) := \tfrac{1}{2}\|\partial_x\psi\|_{L^2}^2 + \tfrac{1}{2}(\omega - \tfrac{1}{4}c^2)\|\psi\|_{L^2}^2 + \frac{c}{2(2\sigma+2)}\|\psi\|_{L^{2\sigma+2}}^{2\sigma+2} - \frac{1}{2\sigma+2}N(\psi).$$

Then, we have $S_{\omega,c}(\varphi) = \widetilde{S}_{\omega,c}(e^{-(c/2)ix}\varphi)$ by using the identities

$$cP(\varphi) = -\|\partial_x\varphi\|_{L^2}^2 - \tfrac{1}{4}c^2\|\varphi\|_{L^2}^2 + \|\partial_x(e^{-(c/2)ix}\varphi)\|_{L^2}^2, \tag{1-6}$$

$$N(\varphi) = -\tfrac{1}{2}c\|\varphi\|_{L^{2\sigma+2}}^{2\sigma+2} + N(e^{-(c/2)ix}\varphi). \tag{1-7}$$

We denote the scaling transformation by $f_\lambda^{\alpha,\beta}(x) := e^{\alpha\lambda}f(e^{-\beta\lambda}x)$ for $(\alpha,\beta)\in\mathbb{R}^2$ and any function f. For $(\alpha,\beta)\in\mathbb{R}^2$, we define

$$\widetilde{K}_{\omega,c}^{\alpha,\beta}(\psi) := \partial_\lambda\widetilde{S}_{\omega,c}(\psi_\lambda^{\alpha,\beta})|_{\lambda=0},$$
$$K_{\omega,c}^{\alpha,\beta}(\varphi) := \widetilde{K}_{\omega,c}^{\alpha,\beta}(e^{-(c/2)ix}\varphi).$$

By a direct calculation, we have the explicit formulae

$$\begin{aligned}
\widetilde{K}_{\omega,c}^{\alpha,\beta}(\psi) &= \langle \widetilde{S}'_{\omega,c}(\psi), \alpha\psi - \beta x\partial_x\psi\rangle \\
&= \frac{2\alpha-\beta}{2}\|\partial_x\psi\|_{L^2}^2 + \frac{2\alpha+\beta}{2}\left(\omega - \frac{c^2}{4}\right)\|\psi\|_{L^2}^2 + \frac{\{(2\sigma+2)\alpha+\beta\}c}{2(2\sigma+2)}\|\psi\|_{L^{2\sigma+2}}^{2\sigma+2} - \alpha N(\psi), \\
K_{\omega,c}^{\alpha,\beta}(\varphi) &= \langle \widetilde{S}'_{\omega,c}(e^{-(c/2)ix}\varphi), \alpha e^{-(c/2)ix}\varphi - \beta x\partial_x(e^{-(c/2)ix}\varphi)\rangle \\
&= \langle S'_{\omega,c}(\varphi), \alpha\varphi + \tfrac{1}{2}ci\beta x\varphi - \beta x\partial_x\varphi\rangle \\
&= \frac{2\alpha-\beta}{2}\|\partial_x\varphi\|_{L^2}^2 + \left(\frac{2\alpha+\beta}{2}\omega - \frac{c^2}{4}\beta\right)\|\varphi\|_{L^2}^2 + \frac{2\alpha-\beta}{2}cP(\varphi) + \frac{\beta c}{2(2\sigma+2)}\|\varphi\|_{L^{2\sigma+2}}^{2\sigma+2} - \alpha N(\varphi),
\end{aligned}$$

where we have used (1-6) and (1-7).

Remark. (1) If $\beta\neq 0$, then $K_{\omega,c}^{\alpha,\beta}$ is different from $I_{\omega,c}^{\alpha,\beta}(\varphi) := \partial_\lambda S_{\omega,c}(\varphi_\lambda^{\alpha,\beta})|_{\lambda=0}$. Indeed, the explicit formula of $I_{\omega,c}^{\alpha,\beta}$ is

$$I_{\omega,c}^{\alpha,\beta}(\varphi) = \frac{2\alpha-\beta}{2}\|\partial_x\varphi\|_{L^2}^2 + \frac{2\alpha+\beta}{2}\omega\|\varphi\|_{L^2}^2 + c\alpha P(\varphi) - \alpha N(\varphi).$$

We note that $K_{\omega,c}^{\alpha,0}$ coincides with $I_{\omega,c}^{\alpha,0}$, and especially $K_{\omega,c}^{1,0} = I_{\omega,c}^{1,0}$ is nothing but the Nehari functional.

(2) It is not clear whether the momentum P is positive or not. That is why we introduce $\widetilde{S}_{\omega,c}$ by using (1-6). Such an argument can be seen in [Bellazzini et al. 2014b] (see (14) therein for the details).

(3) The functional $K_{\omega,c}^{\alpha,\beta}$ is more useful to obtain the characterization of the solitary waves when $\omega = c^2/4$ and $c>0$ than $I_{\omega,c}^{\alpha,\beta}$ since $K_{\omega,c}^{\alpha,\beta}$ contains the $L^{2\sigma+2}$-norm (see the proof in Section 2B).

(4) $\widetilde{S}_{\omega,c}$ and $\widetilde{K}^{\alpha,\beta}_{\omega,c}$ are relevant to the elliptic equation

$$-\psi'' + (\omega - \tfrac{1}{4}c^2)\psi + \tfrac{1}{2}c|\psi|^{2\sigma}\psi - i|\psi|^{2\sigma}\psi' = 0, \quad x \in \mathbb{R}.$$

We define the following function space for (ω, c) satisfying (1-5):

$$X_{\omega,c} := \begin{cases} H^1(\mathbb{R}) & \text{if } \omega > c^2/4, \\ \dot{H}^1(\mathbb{R}) \cap L^{2\sigma+2}(\mathbb{R}) & \text{if } \omega = c^2/4 \text{ and } c > 0. \end{cases}$$

We consider the following minimization problem:

$$\begin{aligned} \mu^{\alpha,\beta}_{\omega,c} &:= \inf\{S_{\omega,c}(\varphi) : e^{-(c/2)ix}\varphi \in X_{\omega,c} \setminus \{0\},\ K^{\alpha,\beta}_{\omega,c}(\varphi) = 0\} \\ &= \inf\{\widetilde{S}_{\omega,c}(\psi) : \psi \in X_{\omega,c} \setminus \{0\},\ \widetilde{K}^{\alpha,\beta}_{\omega,c}(\psi) = 0\}. \end{aligned}$$

Remark. (1) We note that the solitary waves $\phi_{c^2/4,c}$ do not belong to $L^2(\mathbb{R})$ when $\sigma \geq 2$. Therefore, we define $X_{c^2/4,c} := \dot{H}^1(\mathbb{R}) \cap L^{2\sigma+2}(\mathbb{R})$ to characterize the solitary waves $\phi_{c^2/4,c}$ (cf. [Kwon and Wu 2016]).

(2) $S_{c^2/4,c}$ seems meaningless on the function space $\{\varphi : e^{-(c/2)ix}\varphi \in X_{c^2/4,c}\}$ since $S_{c^2/4,c}$ contains L^2-norm. However, in fact, $S_{c^2/4,c}$ is well-defined on the function space since $\widetilde{S}_{c^2/4,c}$ is defined on $\dot{H}^1(\mathbb{R}) \cap L^{2\sigma+2}(\mathbb{R})$ and the equality $S_{c^2/4,c}(\varphi) = \widetilde{S}_{c^2/4,c}(e^{-(c/2)ix}\varphi)$ holds. Similarly, $K^{\alpha,\beta}_{c^2/4,c}$ is well-defined on this function space.

(3) Since $\varphi \in H^1(\mathbb{R})$ if and only if $e^{-(c/2)ix}\varphi \in H^1(\mathbb{R})$, when $\omega > c^2/4$, we have

$$\mu^{\alpha,\beta}_{\omega,c} = \inf\{S_{\omega,c}(\varphi) : \varphi \in H^1(\mathbb{R}) \setminus \{0\},\ K^{\alpha,\beta}_{\omega,c}(\varphi) = 0\}.$$

However, when $\omega = c^2/4$ and $c > 0$, the above equality does not hold.

We assume that $(\alpha, \beta) \in \mathbb{R}^2$ satisfies

$$\begin{cases} 2\alpha - \beta > 0,\ 2\alpha + \beta > 0, \text{ and } \beta c \leq 0 & \text{when } \omega > c^2/4, \\ 2\alpha - \beta > 0,\ 2\alpha + \beta > 0, \text{ and } \beta < 0 & \text{when } \omega = c^2/4 \text{ and } c > 0. \end{cases} \tag{1-8}$$

We define some function spaces:

$$\begin{aligned} \mathscr{M}^{\alpha,\beta}_{\omega,c} &:= \{\varphi : e^{-(c/2)ix}\varphi \in X_{\omega,c} \setminus \{0\},\ S_{\omega,c}(\varphi) = \mu^{\alpha,\beta}_{\omega,c},\ K^{\alpha,\beta}_{\omega,c}(\varphi) = 0\}, \\ \mathscr{G}_{\omega,c} &:= \{\varphi : e^{-(c/2)ix}\varphi \in X_{\omega,c} \setminus \{0\},\ S'_{\omega,c}(\varphi) = 0\}. \end{aligned}$$

We give the following characterization of the solitary waves.

Theorem 1.1. *Let $\sigma \geq 1$, (ω, c) satisfy* (1-5), *and (α, β) satisfy* (1-8). *Then*,

$$\mathscr{M}^{\alpha,\beta}_{\omega,c} = \mathscr{G}_{\omega,c} = \{e^{i\theta_0}\phi_{\omega,c}(\cdot - x_0) : \theta_0 \in [0, 2\pi),\ x_0 \in \mathbb{R}\}.$$

Theorem 1.1 also means that $\mu^{\alpha,\beta}_{\omega,c}$ and $\mathscr{M}^{\alpha,\beta}_{\omega,c}$ are independent of (α, β) and $\mathscr{M}^{\alpha,\beta}_{\omega,c}$ is not empty. Thus, we denote $\mu^{\alpha,\beta}_{\omega,c}$ by $\mu_{\omega,c}$.

We define

$$\begin{aligned} \mathscr{K}^{\alpha,\beta,+}_{\omega,c} &:= \{\varphi \in H^1(\mathbb{R}) : S_{\omega,c}(\varphi) \leq \mu_{\omega,c},\ K^{\alpha,\beta}_{\omega,c}(\varphi) \geq 0\}, \\ \mathscr{K}^{\alpha,\beta,-}_{\omega,c} &:= \{\varphi \in H^1(\mathbb{R}) : S_{\omega,c}(\varphi) \leq \mu_{\omega,c},\ K^{\alpha,\beta}_{\omega,c}(\varphi) < 0\}. \end{aligned}$$

The characterization by Theorem 1.1 gives us the following sufficient condition for global existence.

Theorem 1.2. *Let $\sigma \geq 1$, (ω, c) satisfy* (1-5), *and (α, β) satisfy* (1-8). *Then, $\mathcal{K}_{\omega,c}^{\alpha,\beta,\pm}$ are invariant under the flow of* (gDNLS). *Namely, if the initial data u_0 belongs to $\mathcal{K}_{\omega,c}^{\alpha,\beta,\pm}$, then the solution $u(t)$ of* (gDNLS) *also belongs to $\mathcal{K}_{\omega,c}^{\alpha,\beta,\pm}$ for all $t \in I_{\max}$, where $I_{\max}$ denotes the maximal existence time.*

Moreover, if the initial data u_0 belongs to $\mathcal{K}_{\omega,c}^{\alpha,\beta,+}$ for some (ω, c) satisfying (1-5) *and (α, β) satisfying* (1-8), *then the corresponding solution u of* (gDNLS) *exists globally in time and*

$$\|u\|_{L^\infty(\mathbb{R}:H^1(\mathbb{R}))} \leq C(\|u_0\|_{H^1}),$$

where $C : [0, \infty) \to \mathbb{R}$ is continuous.

Recently, Miao et al. [2017a] independently obtained the results similar to Theorems 1.1 and 1.2 when $\sigma = 1$. We will compare their method with our argument in the remark on page 1156.

We show that Theorem 1.2 gives us some interesting corollaries for (DNLS).

Corollary 1.3. *Let $\sigma = 1$. If the initial data $u_0 \in H^1(\mathbb{R})$ satisfies $\|u_0\|_{L^2}^2 < 4\pi$, then the solution of* (DNLS) *is global.*

Two proofs have been known for Corollary 1.3. One was obtained by Wu [2015] and another one by Guo and Wu [2017]. We give another proof by Theorem 1.2. We compare the methods of [Wu 2015; Guo and Wu 2017], which depend on the sharp Gagliardo–Nirenberg-type inequality, with our variational argument. Using the gauge transformation to the solution of (DNLS)

$$u(t, x) = w(t, x) \exp\left(-\frac{i}{4}\int_{-\infty}^{x} |w(t, x)|^2\, dx\right), \tag{1-9}$$

then w satisfies the equation

$$\begin{cases} i\partial_t w + \partial_x^2 w + \frac{1}{2}i|w|^2\partial_x w - \frac{1}{2}iw^2\partial_x \overline{w} + \frac{3}{16}|w|^4 w = 0, & (t, x) \in \mathbb{R} \times \mathbb{R}, \\ w(0, x) = w_0(x), & x \in \mathbb{R}. \end{cases} \tag{1-10}$$

The energy and the momentum are transformed as

$$\mathcal{E}(w) = \tfrac{1}{2}\|\partial_x w\|_{L^2}^2 - \tfrac{1}{32}\|w\|_{L^6}^6,$$
$$\mathcal{P}(w) = \mathrm{Re}\int_{\mathbb{R}} i\partial_x w \overline{w}\, dx + \tfrac{1}{4}\|w\|_{L^4}^4.$$

Hayashi and Ozawa [1992] used the sharp Gagliardo–Nirenberg inequality

$$\|f\|_{L^6}^6 \leq \frac{4}{\pi^2}\|f\|_{L^2}^4 \|\partial_x f\|_{L^2}^2 \tag{1-11}$$

in order to obtain an a priori estimate in $\dot{H}^1(\mathbb{R})$. We note that the optimizer for the inequality (1-11) is given by $Q := \Phi_{1,0}$ and Q satisfies the elliptic equation

$$-Q'' + Q - \tfrac{3}{16}Q^5 = 0. \tag{1-12}$$

Hayashi and Ozawa [1992] proved the H^1-solution of (DNLS) is global if the initial data u_0 satisfies $\|u_0\|_{L^2}^2 = \|w_0\|_{L^2}^2 < \|Q\|_{L^2}^2 = 2\pi$ (see also [Weinstein 1982]). Wu [2015] used not only the energy but

also the momentum and the sharp Gagliardo–Nirenberg inequality

$$\|f\|_{L^6}^6 \le 3(2\pi)^{-2/3}\|f\|_{L^4}^{16/3}\|\partial_x f\|_{L^2}^{2/3}. \tag{1-13}$$

We note that the optimizer for the inequality (1-13) is given by $W := \Phi_{1/4,1}$ and W satisfies the elliptic equation

$$-W'' + \tfrac{1}{2}W^3 - \tfrac{3}{16}W^5 = 0. \tag{1-14}$$

Wu [2015] proved that the H^1-solution of (DNLS) is global if the initial data u_0 satisfies $\|u_0\|_{L^2}^2 = \|w_0\|_{L^2}^2 < \|W\|_{L^2}^2 = 4\pi$. His proof depends on a contradiction argument. Supposing that there exists a time sequence $\{t_n\}_{n\in\mathbb{N}}$ with $t_n \to T_{\max}$ or $-T_{\min}$ such that $\|\partial_x w(t_n)\|_{L^2} \to \infty$ as $n \to \infty$, where $(-T_{\min}, T_{\max})$ is the maximal time interval, he mainly proved that $X = \|w(t_n)\|_{L^4}^8/\|w(t_n)\|_{L^6}^6$ satisfies $X^3 - \|w\|_{L^2}^2X^2 + 16\{3(2\pi)^{-2/3}\}^{-3}\|w\|_{L^2}^2 < 0$, but this does not hold when $\|w\|_{L^2}^2 < 4\pi$. On the other hand, Guo and Wu [2017] gave an a priori estimate directly for (1-10) by the sharp Gagliardo–Nirenberg inequality (1-13). More precisely, they showed in [Guo and Wu 2017, Lemma 2.1] the inequality

$$\mathcal{P}(w) \ge \tfrac{1}{4}\|w\|_{L^4}^4\left(1 - \frac{\|w\|_{L^2}}{2\sqrt{\pi}}\right) - \frac{8\sqrt{\pi}\mathcal{E}(w)\|w\|_{L^2}}{\|w\|_{L^4}^4}, \tag{1-15}$$

and thus, $\|\partial_x w\|_{L^2}^2$ is bounded by $\mathcal{P}$ and $\mathcal{E}$ if $\|w\|_{L^2}^2 < 4\pi$ [Guo and Wu 2017, Lemma 2.2]. In our variational argument, we do not use a contradiction argument, the gauge transformation like (1-9), or any sharp Gagliardo–Nirenberg inequality.

We give the global existence result in the threshold case by Theorem 1.2.

Corollary 1.4. *Let $\sigma = 1$. We assume that the initial data $u_0 \in H^1(\mathbb{R})$ satisfies $\|u_0\|_{L^2}^2 = 4\pi$. If $P(u_0) < 0$, then the solution of* (DNLS) *is global.*

After submitting the present paper, Guo pointed out that Corollary 1.4 can be obtained by (1-15). We also give the proof by (1-15) for the reader's convenience.

The following corollary means that there exist global solutions with any large mass.

Corollary 1.5. *Let $\sigma \ge 1$. Given $\psi \in H^1(\mathbb{R})$, set the initial data as $u_{0,c} = e^{(c/2)ix}\psi$. Then there exists $c_0 > 0$ such that, if $c \ge c_0$, then the corresponding solution u_c of* (gDNLS) *is global.*

Remark. The existence of blow-up solutions in finite time is still an open problem. It might be a very interesting problem whether finite-time blow-up occurs when the initial data u_0 satisfies $\|u_0\|_{L^2}^2 = 4\pi$ and $P(u_0) > 0$.

1C. ***Compare DNLS with mass-critical NLS.*** Equation (DNLS) is L^2-critical in the sense that the equation and L^2-norm are invariant under the scaling transformation

$$u_\gamma(t, x) := \gamma^{1/2}u(\gamma^2 t, \gamma x), \quad \gamma > 0.$$

The same invariance holds for the quintic nonlinear Schrödinger equation in one-dimensional space:

$$i\partial_t u + \partial_x^2 u + \tfrac{3}{16}|u|^4 u = 0, \quad (t, x) \in \mathbb{R} \times \mathbb{R}. \tag{1-16}$$

This equation has the same energy as (1-10). It is known that (1-16) is locally well-posed in the energy space $H^1(\mathbb{R})$ and the solution is global if the initial data u_0 satisfies $\|u_0\|_{L^2}^2 < \|Q\|_{L^2}^2$, where Q is the ground state of the same elliptic equation (1-12). The condition $\|u_0\|_{L^2}^2 < \|Q\|_{L^2}^2$ is equivalent to the condition obtained by the variational argument. In this argument, the momentum is not essential since (1-16) is invariant under the Galilean transformation, and thus, we may assume that the momentum is zero. On the other hand, (DNLS) is not invariant under the Galilean transformation. Therefore, the condition by the variational argument is better than the assumption $\|u_0\|_{L^2}^2 < \|W\|_{L^2}^2 = 4\pi$. Indeed, the momentum and the parameter c play important roles in Corollaries 1.4 and 1.5.

1D. ***Idea of proofs.*** The proof of Theorem 1.1 is based on the method of Colin and Ohta [2006] (concentration compactness method). They characterized the solitary waves for $\omega > c^2/4$ when $\sigma = 1$ by the Nehari functional $I_{\omega,c}^{1,0}$. However, in the case $\omega = c^2/4$ and $c > 0$, we cannot apply their argument directly since the L^2-norm in $I_{\omega,c}^{1,0}$ disappears by (1-6). Therefore, we introduce the new functional $K_{\omega,c}^{\alpha,\beta}$ for (α, β) satisfying (1-8). We can use the $L^{2\sigma+2}$-norm instead of the L^2-norm by using $K_{\omega,c}^{\alpha,\beta}$. That is why we introduce the function space $X_{\omega,c}$ as $\dot{H}^1 \cap L^{2\sigma+2}$ in the massless case (i.e., $\omega = c^2/4$ and $c > 0$). Noting that the solitary waves $\phi_{c^2/4,c}$ do not belong to $L^2(\mathbb{R})$ when $\sigma \geq 2$, the function space $X_{\omega,c}$ is essential to obtain the characterization of the solitary waves $\phi_{c^2/4,c}$. Based on the argument of Colin and Ohta [2006], we characterize the solitary waves $\phi_{c^2/4,c}$ by $K_{\omega,c}^{\alpha,\beta}$. By the conservation laws and the characterization of the solitary waves, we get an a priori estimate and thus obtain Theorem 1.2. The corollaries follow from Theorem 1.2. In their proofs, the parameter c plays an important role. More precisely, taking $c > 0$ large, we get the corollaries. At last, we emphasize that we do not use the sharp Gagliardo–Nirenberg inequality and we do not apply the gauge transformation to (gDNLS) since the equation after applying the transformation is complicated unlike (DNLS).

Remark. Miao et al. [2017a] treated the case of $\sigma = 1$. They considered (1-10) by using the gauge transformation and defined the action by $\mathcal{S}_{\omega,c} := \mathcal{E} + \omega M/2 + c\mathcal{P}/2$. They applied a concentration compactness argument to give the variational characterization of the solitary waves. Then, they use the Nehari functional $\mathcal{K}_{\omega,c}$ derived from the action $\mathcal{S}_{\omega,c}$. The explicit formula of $\mathcal{K}_{\omega,c}$ is

$$\mathcal{K}_{\omega,c}(w) := \|\partial_x w\|_{L^2}^2 - \tfrac{3}{16}\|w\|_{L^6}^6 + \omega\|w\|_{L^2}^2 + c\,\mathrm{Re}\int_{\mathbb{R}} i\,\partial_x w\,\overline{w}\,dx + \tfrac{1}{2}c\|w\|_{L^4}^4.$$

They defined

$$\mathscr{A}_{\omega,c}^{\pm} := \{\varphi \in H^1(\mathbb{R}) : \mathcal{S}_{\omega,c}(\varphi) \leq \mathcal{S}_{\omega,c}(\phi_{\omega,c}),\ \mathcal{K}_{\omega,c}(\varphi) \gtrless 0\},$$

and they also showed that $\mathscr{A}_{\omega,c}^{\pm}$ are invariant under the flow of (1-10) and the solution to (1-10) is global if $w_0 \in \mathscr{A}_{\omega,c}^{+}$ for some (ω, c). The functional $\mathcal{K}_{\omega,c}$ is useful to characterize the solitary waves $\phi_{c^2/4,c}$ since it contains L^4-norm. Namely, one can use the Nehari functional by the gauge transformation. On the other hand, we cannot use the Nehari functional when we do not apply the gauge transformation, and thus, we introduce the new functionals $K_{\omega,c}^{\alpha,\beta}$.

The rest of the present paper is as follows. In Section 2A, we prepare some lemmas to obtain the characterization of the solitary waves and prove the a priori estimate (see (2-2)). In Section 2B, we give

the characterization of the solitary waves $\phi_{c^2/4,c}$. We remark that the characterization of the solitary waves $\phi_{\omega,c}$ for $\omega > c^2/4$ can be obtained in the same manner as in [Colin and Ohta 2006], and then we omit the proof. Section 3 is devoted to the proof of Theorem 1.2 and the corollaries. In the Appendix, we show that there is no nontrivial solution of the nonlinear elliptic equation (1-4) if $\omega < c^2/4$, or $\omega = c^2/4$ and $c \le 0$.

2. Variational characterization of the solitary waves

2A. ***Preliminaries.*** We define function spaces

$$\tilde{\mathscr{M}}^{\alpha,\beta}_{\omega,c} := \{\psi \in X_{\omega,c} \setminus \{0\} : \widetilde{S}_{\omega,c}(\psi) = \mu^{\alpha,\beta}_{\omega,c},\ \widetilde{K}^{\alpha,\beta}_{\omega,c}(\psi) = 0\},$$
$$\widetilde{\mathscr{G}}_{\omega,c} := \{\psi \in X_{\omega,c} \setminus \{0\} : \widetilde{S}'_{\omega,c}(\psi) = 0\}.$$

In this section, we prove the following proposition, which gives Theorem 1.1.

Proposition 2.1. *Let* (ω, c) *satisfy* (1-5) *and* (α, β) *satisfy* (1-8)*. Then*

$$\tilde{\mathscr{M}}^{\alpha,\beta}_{\omega,c} = \widetilde{\mathscr{G}}_{\omega,c} = \{e^{i\theta} e^{-(c/2)ix} \phi_{\omega,c}(\cdot - y) : \theta \in [0, 2\pi),\ y \in \mathbb{R}\}.$$

Indeed, Theorem 1.1 follows from Proposition 2.1 and the following properties:

$$\varphi \in \mathscr{M}^{\alpha,\beta}_{\omega,c} \iff e^{-(c/2)ix}\varphi \in \tilde{\mathscr{M}}^{\alpha,\beta}_{\omega,c},$$
$$\varphi \in \mathscr{G}_{\omega,c} \iff e^{-(c/2)ix}\varphi \in \widetilde{\mathscr{G}}_{\omega,c},$$

where we note that $\widetilde{S}'_{\omega,c}(e^{-(c/2)ix}\varphi) = e^{-(c/2)ix} S'_{\omega,c}(\varphi)$ holds.

To prove Proposition 2.1, we prepare some basic lemmas. We have the Gagliardo–Nirenberg-type inequality.

Lemma 2.2. *Let* $p \ge 1$*. We have the estimate*

$$\|f\|^{2p}_{L^\infty} \le 2p \|f\|^{2p-1}_{L^{4p-2}} \|\partial_x f\|_{L^2}. \tag{2-1}$$

Proof. By the Hölder inequality,

$$\begin{aligned} |f(x)|^{2p} &= \int_{-\infty}^{x} \frac{d}{dx}(|f(y)|^{2p})\, dy \\ &= \int_{-\infty}^{x} 2p|f(y)|^{2p-2} \operatorname{Re}(\overline{f(y)}(\partial_x f)(y))\, dy \\ &\le 2p \||f|^{2p-1}\|_{L^2} \|\partial_x f\|_{L^2} \\ &= 2p \|f\|^{2p-1}_{L^{4p-2}} \|\partial_x f\|_{L^2}. \end{aligned}$$

Taking the supremum, we obtain (2-1). □

We have the Lieb compactness lemma. See [Lieb 1983] for $p = 2$ and [Bellazzini et al. 2014a, Lemma 2.1] for more general setting.

Lemma 2.3. *Let* $p \ge 2$ *and* $d \in \mathbb{N}$*. Let* $\{f_n\}$ *be a bounded sequence in* $\dot{H}^1(\mathbb{R}^d) \cap L^p(\mathbb{R}^d)$*. Assume that there exists* $q \in (p, 2^*)$ *such that* $\limsup_{n\to\infty} \|f_n\|_{L^q} > 0$*. Then there exist* $\{y_n\}$ *and* $f \in \dot{H}^1(\mathbb{R}^d) \cap L^p(\mathbb{R}^d) \setminus \{0\}$ *such that* $\{f_n(\cdot - y_n)\}$ *has a subsequence that converges to* f *weakly in* $\dot{H}^1(\mathbb{R}^d) \cap L^p(\mathbb{R}^d)$*.*

We have the Brézis–Lieb lemma [1983].

Lemma 2.4. *Let $d \in \mathbb{N}$ and $1 < p < \infty$. Let $\{f_n\}$ be a bounded sequence in $L^p(\mathbb{R}^d)$ and $f_n \to f$ a.e. in $\mathbb{R}^d$. Then*

$$\|f_n\|_{L^p}^p - \|f_n - f\|_{L^p}^p - \|f\|_{L^p}^p \to 0.$$

If $\{f_n\}$ is a bounded sequence in $L^2(\mathbb{R}^d)$ and f_n converges to f weakly in $L^2(\mathbb{R}^d)$, then the statement with $p = 2$ holds.

A direct calculation gives us the following relation.

Lemma 2.5. *We have*

$$\alpha(2\sigma+2)\widetilde{S}_{\omega,c}(\psi) = \widetilde{K}_{\omega,c}^{\alpha,\beta}(\psi) + \frac{2\sigma\alpha+\beta}{2}\|\partial_x\psi\|_{L^2}^2 + (\omega - \tfrac{1}{4}c^2)\frac{2\sigma\alpha-\beta}{2}\|\psi\|_{L^2}^2 - \frac{\beta c}{2(2\sigma+2)}\|\psi\|_{L^{2\sigma+2}}^{2\sigma+2}. \tag{2-2}$$

We denote the difference $\alpha(2\sigma+2)\widetilde{S}_{\omega,c}(\psi) - \widetilde{K}_{\omega,c}^{\alpha,\beta}(\psi)$ by

$$\tilde{J}_{\omega,c}^{\alpha,\beta}(\psi) := \frac{2\sigma\alpha+\beta}{2}\|\partial_x\psi\|_{L^2}^2 + (\omega - \tfrac{1}{4}c^2)\frac{2\sigma\alpha-\beta}{2}\|\psi\|_{L^2}^2 - \frac{\beta c}{2(2\sigma+2)}\|\psi\|_{L^{2\sigma+2}}^{2\sigma+2}.$$

2B. ***Variational characterization.*** First we consider the case of $\omega = c^2/4$ and $c > 0$. Then (α, β) satisfies

$$2\alpha - \beta > 0, \qquad 2\alpha + \beta > 0, \qquad \beta < 0. \tag{2-3}$$

Hereafter, we often omit the indices ω, c, α, and β for simplicity.

Lemma 2.6. *The following equality holds*:

$$\widetilde{\mathscr{G}}_{\omega,c} = \{e^{i\theta_0}e^{-(c/2)ix}\phi_{\omega,c}(\cdot - x_0) : \theta_0 \in [0, 2\pi),\ x_0 \in \mathbb{R}\}.$$

Proof. Since $e^{-(c/2)ix}\phi_{\omega,c}$ satisfies $\widetilde{S}'_{\omega,c}(e^{-(c/2)ix}\phi_{\omega,c}) = e^{-(c/2)ix}S'_{\omega,c}(\phi_{\omega,c}) = 0$, we have $\widetilde{\mathscr{G}}_{\omega,c} \supset \{e^{i\theta_0}e^{-(c/2)ix}\phi_{\omega,c}(\cdot - x_0) : \theta_0 \in [0, 2\pi), x_0 \in \mathbb{R}\}$. We prove $\widetilde{\mathscr{G}}_{\omega,c} \subset \{e^{i\theta_0}e^{-(c/2)ix}\phi_{\omega,c}(\cdot - x_0) : \theta_0 \in [0, 2\pi), x_0 \in \mathbb{R}\}$. Letting $\psi \in \widetilde{\mathscr{G}}_{\omega,c}$ and

$$\psi(x) = \Phi(x)\exp\left(-\frac{i}{2\sigma+2}\int_0^x |\Phi(y)|^{2\sigma}\,dy\right),$$

then Φ is a solution of

$$-\Phi'' + \tfrac{1}{2}c|\Phi|^{2\sigma}\Phi - \frac{2\sigma+1}{(2\sigma+2)^2}|\Phi|^{4\sigma}\Phi + \frac{\sigma}{\sigma+1}|\Phi|^{2\sigma-2}\operatorname{Im}(\bar{\Phi}\Phi')\Phi = 0.$$

Setting $A(\Phi) := \frac{1}{2}c|\Phi|^{2\sigma} - ((2\sigma+1)/(2\sigma+2)^2)|\Phi|^{4\sigma} + (\sigma/(\sigma+1))|\Phi|^{2\sigma-2}\operatorname{Im}(\bar{\Phi}\Phi')$, $f := \operatorname{Re}\Phi$, and $g := \operatorname{Im}\Phi$,

$$f'' = A(\Phi)f, \qquad g'' = A(\Phi)g.$$

Therefore,

$$(fg' - gf')' = fg'' - gf'' = fA(\Phi)g - gA(\Phi)f = A(\Phi)fg - A(\Phi)fg = 0.$$

Since $f, g \in \dot{H}^1(\mathbb{R}) \cap L^{2\sigma+2}(\mathbb{R})$, we obtain $fg' - gf' = 0$. On the other hand, $fg' - gf' = \operatorname{Re}\Phi \operatorname{Im}\Phi' - \operatorname{Im}\Phi \operatorname{Re}\Phi' = \operatorname{Im}(\bar{\Phi}\Phi')$. Thus, $\operatorname{Im}(\bar{\Phi}\Phi') = 0$ for any $x \in \mathbb{R}$. Therefore, Φ satisfies

$$-\Phi'' + \tfrac{1}{2}c|\Phi|^{2\sigma}\Phi - \frac{2\sigma+1}{(2\sigma+2)^2}|\Phi|^{4\sigma}\Phi = 0. \tag{2-4}$$

Therefore, there exist θ_0 and x_0 such that $\Phi = e^{i\theta_0}\Phi_{\omega,c}(\cdot - x_0)$ since $\Phi_{\omega,c}$ is the unique solution of (2-4) up to translation and phase (see the Appendix). This implies $\psi(x) = e^{i\theta}e^{-(c/2)ix}\phi_{\omega,c}(x - x_0)$. □

Remark. According to [Colin and Ohta 2006], it looks natural to take the integral on the infinite interval $(-\infty, x]$ in the gauge transformation as

$$\psi(x) = \Phi(x)\exp\left(-\frac{i}{2\sigma+2}\int_{-\infty}^{x}|\Phi(y)|^{2\sigma}\,dy\right).$$

However, in the massless case, it is not clear whether $\psi \in \widetilde{\mathscr{G}}_{\omega,c}$ belongs to $L^{2\sigma}(\mathbb{R})$. This is why we take the integral on the finite interval $[0, x]$ instead of $(-\infty, x]$.

Lemma 2.7. *We have* $\widetilde{\mathscr{G}}_{\omega,c} \supset \widetilde{\mathscr{M}}^{\alpha,\beta}_{\omega,c}$.

Proof. This is obvious if $\widetilde{\mathscr{M}} = \varnothing$. We consider the case of $\widetilde{\mathscr{M}} \neq \varnothing$. Let $\psi \in \widetilde{\mathscr{M}}$. Since ψ is a minimizer, there exists a Lagrange multiplier $\eta \in \mathbb{R}$ such that $\widetilde{S}'(\psi) = \eta\widetilde{K}'(\psi)$. Then

$$0 = \widetilde{K}(\psi) = \langle \widetilde{S}'(\psi), \partial_\lambda \psi_\lambda^{\alpha,\beta}|_{\lambda=0}\rangle = \eta\langle \widetilde{K}'(\psi), \partial_\lambda \psi_\lambda^{\alpha,\beta}|_{\lambda=0}\rangle = \eta\partial_\lambda \widetilde{K}(\psi_\lambda^{\alpha,\beta})|_{\lambda=0},$$

where we remark that this is justified by a density argument. By a direct calculation, we obtain

$$\begin{aligned}\partial_\lambda \widetilde{K}(\psi_\lambda^{\alpha,\beta})|_{\lambda=0} &= \frac{(2\alpha-\beta)^2}{2}\|\partial_x\psi\|_{L^2}^2 - \frac{\{(2\sigma+2)\alpha+\beta\}^2}{2(2\sigma+2)}\|\psi\|_{L^{2\sigma+2}}^{2\sigma+2} - \frac{\{(2\sigma+2)\alpha\}^2}{2\sigma+2}N(\psi)\\ &= \frac{-(2\alpha-\beta)(2\sigma\alpha+\beta)}{2}\|\partial_x\psi\|_{L^2}^2 + \frac{\{(2\sigma+2)\alpha+\beta\}\beta c}{2(2\sigma+2)}\|\psi\|_{L^{2\sigma+2}}^{2\sigma+2} + (2\sigma+2)\alpha\widetilde{K}(\psi)\\ &< 0,\end{aligned}$$

where in the last inequality we use

$$2\alpha - \beta > 0, \qquad 2\alpha + \beta > 0, \qquad \beta < 0, \qquad \widetilde{K}(\psi) = 0.$$

Therefore, $\eta = 0$. This implies $\widetilde{S}'_{\omega,c}(\psi) = 0$ and then $\psi \in \widetilde{\mathscr{G}}_{\omega,c}$. □

Lemma 2.8. *We have* $\widetilde{\mathscr{G}}_{\omega,c} \subset \widetilde{\mathscr{M}}^{\alpha,\beta}_{\omega,c}$ *if* $\widetilde{\mathscr{M}}^{\alpha,\beta}_{\omega,c} \neq \varnothing$.

Proof. Let $\psi \in \widetilde{\mathscr{G}}$. Then there exist $\theta_0 \in [0, 2\pi)$ and $x_0 \in \mathbb{R}$ such that $\psi = e^{i\theta_0}e^{-(c/2)ix}\phi_{\omega,c}(\cdot - x_0)$ by Lemma 2.6. If $\widetilde{\mathscr{M}} \neq \varnothing$, then we can take $\varphi \in \widetilde{\mathscr{M}}$. By Lemmas 2.6 and 2.7, there exist $\theta_1 \in [0, 2\pi)$ and $x_1 \in \mathbb{R}$ such that $\varphi = e^{i\theta_1}e^{-(c/2)ix}\phi_{\omega,c}(\cdot - x_1)$. Thus, $\widetilde{S}_{\omega,c}(\psi) = \widetilde{S}_{\omega,c}(\phi_{\omega,c}) = \widetilde{S}_{\omega,c}(\varphi) = \mu_{\omega,c}$. Moreover, we have $\widetilde{K}(\psi) = \langle \widetilde{S}'_{\omega,c}(\psi), \partial_\lambda \psi_\lambda^{\alpha,\beta}|_{\lambda=0}\rangle = 0$. □

Lemma 2.9. *We have* $\widetilde{\mathscr{M}}^{\alpha,\beta}_{\omega,c} \neq \varnothing$.

To prove this lemma, we show the following proposition.

Proposition 2.10. *Let $\{\psi_n\}_{n\in\mathbb{N}} \subset X_{\omega,c}$ satisfy*

$$\widetilde{S}_{\omega,c}(\psi_n) \to \mu^{\alpha,\beta}_{\omega,c} \quad \text{and} \quad \widetilde{K}^{\alpha,\beta}_{\omega,c}(\psi_n) \to 0.$$

Then there exist $\{y_n\} \subset \mathbb{R}$ and $\psi \in \mathscr{M}^{\alpha,\beta}_{\omega,c}$ such that $\{\psi_n(\cdot - y_n)\}$ has a subsequence which converges to ψ strongly in $X_{\omega,c}$.

To prove this proposition, first, we prove the following lemma.

Lemma 2.11. *We have $\mu^{\alpha,\beta}_{\omega,c} > 0$.*

Proof. We recall that $\mu^{\alpha,\beta}_{\omega,c} = \inf\{\widetilde{S}_{\omega,c}(\psi) : \psi \in X_{\omega,c} \setminus \{0\},\ \widetilde{K}^{\alpha,\beta}_{\omega,c}(\psi) = 0\}$. By (2-2), it is trivial that $\mu \geq 0$. We prove $\mu > 0$ by contradiction. We assume that $\mu = 0$. Taking the minimizing sequence $\{\psi_n\} \subset X_{\omega,c}$, i.e., $\widetilde{S}(\psi_n) \to \mu = 0$ and $\widetilde{K}(\psi_n) = 0$, we have $\|\partial_x \psi_n\|^2_{L^2} \to 0$ and $\|\psi_n\|^{2\sigma+2}_{L^{2\sigma+2}} \to 0$ by (2-2) and (2-3). Then, by using (2-1) with $p = (\sigma+2)/2$, we get $\|\psi_n\|_{L^\infty} \to 0$ as $n \to \infty$. By using

$$-N(\psi) = -\|\partial_x \psi\|^2_{L^2} - \tfrac{1}{4}\|\psi\|^{4\sigma+2}_{L^{4\sigma+2}} + \|\partial_x \psi + \tfrac{1}{2} i|\psi|^{2\sigma}\psi\|^2_{L^2},$$

we obtain

$$\begin{aligned}
\widetilde{K}(\psi_n) &= \frac{2\alpha-\beta}{2}\|\partial_x \psi_n\|^2_{L^2} + \frac{\{(2\sigma+2)\alpha+\beta\}c}{2(2\sigma+2)}\|\psi_n\|^{2\sigma+2}_{L^{2\sigma+2}} - \alpha N(\psi_n) \\
&= -\tfrac{1}{2}\beta\|\partial_x \psi_n\|^2_{L^2} + \frac{\{(2\sigma+2)\alpha+\beta\}c}{2(2\sigma+2)}\|\psi_n\|^{2\sigma+2}_{L^{2\sigma+2}} - \tfrac{1}{4}\alpha\|\psi_n\|^{4\sigma+2}_{L^{4\sigma+2}} + \alpha\|\partial_x \psi_n + \tfrac{1}{2} i|\psi_n|^{2\sigma}\psi_n\|^2_{L^2} \\
&\geq \frac{\{(2\sigma+2)\alpha+\beta\}c}{2(2\sigma+2)}\|\psi_n\|^{2\sigma+2}_{L^{2\sigma+2}} - \tfrac{1}{4}\alpha\|\psi_n\|^{4\sigma+2}_{L^{4\sigma+2}} \\
&\geq \frac{\{(2\sigma+2)\alpha+\beta\}c}{2(2\sigma+2)}\|\psi_n\|^{2\sigma+2}_{L^{2\sigma+2}} - \tfrac{1}{4}\alpha\|\psi_n\|^{2\sigma+2}_{L^{2\sigma+2}}\|\psi_n\|^{2\sigma}_{L^\infty} \\
&\geq \left(\frac{\{(2\sigma+2)\alpha+\beta\}c}{2(2\sigma+2)} - \tfrac{1}{4}\alpha\|\psi_n\|^{2\sigma}_{L^\infty}\right)\|\psi_n\|^{2\sigma+2}_{L^{2\sigma+2}} \\
&> 0,
\end{aligned}$$

for large $n \in \mathbb{N}$ since $\|\psi_n\|_{L^\infty} \to 0$ as $n \to \infty$. However, this contradicts $\widetilde{K}(\psi_n) = 0$ for all $n \in \mathbb{N}$. □

Proof of Proposition 2.10. We take $\{\psi_n\} \subset X_{\omega,c}$ such that $\widetilde{S}_{\omega,c}(\psi_n) \to \mu^{\alpha,\beta}_{\omega,c}$ and $\widetilde{K}^{\alpha,\beta}_{\omega,c}(\psi_n) \to 0$. Then, $\{\psi_n\}$ is a bounded sequence in $X_{\omega,c}$ by (2-2).

Step 1. We prove $\limsup_{n\to\infty}\|\psi_n\|_{L^{4\sigma+2}} > 0$ by contradiction. We suppose that $\limsup_{n\to\infty}\|\psi_n\|_{L^{4\sigma+2}} = 0$. Since

$$0 \leftarrow \widetilde{K}(\psi_n) \geq -\tfrac{1}{2}\beta\|\partial_x \psi_n\|^2_{L^2} + \frac{\{(2\sigma+2)\alpha+\beta\}c}{2(2\sigma+2)}\|\psi_n\|^{2\sigma+2}_{L^{2\sigma+2}} - \tfrac{1}{4}\alpha\|\psi_n\|^{4\sigma+2}_{L^{4\sigma+2}},$$

we obtain $\|\partial_x \psi_n\|^2_{L^2} \to 0$ and $\|\psi_n\|^{2\sigma+2}_{L^{2\sigma+2}} \to 0$ as $n \to \infty$. By (2-2), we get $\widetilde{S}(\psi_n) \to 0$. This contradicts $\mu > 0$.

Step 2. Since $\{\psi_n\}$ is bounded in $X_{\omega,c} = \dot{H}^1(\mathbb{R}) \cap L^{2\sigma+2}(\mathbb{R})$ and $\limsup_{n\to\infty}\|\psi_n\|_{L^{4\sigma+2}} > 0$, by applying Lemma 2.3 with $f_n = \psi_n$, $d = 1$, and $p = 2\sigma+2$, there exist $\{y_n\}$ and $v \in X_{\omega,c} \setminus \{0\}$ such that $\{\psi_n(\cdot - y_n)\}$ (we denote this by v_n) has a subsequence that converges to v weakly in $X_{\omega,c}$.

Step 3. We show

$$\widetilde{K}(v_n) - \widetilde{K}(v - v_n) - \widetilde{K}(v) \to 0 \quad \text{as } n \to \infty. \tag{2-5}$$

We note that

$$\widetilde{K}(\psi) = -\tfrac{1}{2}\beta\|\partial_x\psi\|_{L^2}^2 + \frac{\{(2\sigma+2)\alpha+\beta\}c}{2(2\sigma+2)}\|\psi\|_{L^{2\sigma+2}}^{2\sigma+2} - \tfrac{1}{4}\alpha\|\psi\|_{L^{4\sigma+2}}^{4\sigma+2} + \alpha\|\partial_x\psi + \tfrac{1}{2}i|\psi|^{2\sigma}\psi\|_{L^2}^2, \tag{2-6}$$

for any $\psi \in X_{\omega,c}$. Since v_n converges to v weakly in $X_{\omega,c}$, we have $v_n \to v$ a.e. in $\mathbb{R}$. Therefore, by Lemma 2.4, we have $\|v_n\|_{L^p}^p - \|v_n - v\|_{L^p}^p - \|v\|_{L^p}^p \to 0$ for $2\sigma + 2 \le p < \infty$. Moreover, setting

$$w_n := \partial_x v_n + \tfrac{1}{2}i|v_n|^{2\sigma}v_n \quad \text{and} \quad w = \partial_x v + \tfrac{1}{2}i|v|^{2\sigma}v,$$

w_n converges to w weakly in $L^2(\mathbb{R})$. Indeed, it is obvious that $\partial_x v_n \rightharpoonup \partial_x v$ in $L^2(\mathbb{R})$ and we have, for any $f \in C_0^\infty(\mathbb{R})$,

$$\begin{aligned}\left|\int_{\mathbb{R}} f(x)(|v_n(x)|^{2\sigma}v_n(x) - |v(x)|^{2\sigma}v(x))\,dx\right| &\lesssim \int_{\operatorname{supp} f} |f(x)|(|v_n(x)|^{2\sigma} + |v(x)|^{2\sigma})|v_n(x) - v(x)|\,dx \\ &\lesssim \int_{\operatorname{supp} f} |v_n(x) - v(x)|\,dx \to 0,\end{aligned}$$

where we use the Hölder inequality, the fact that $\{v_n\}$ is bounded in $L^\infty(\mathbb{R})$, and the compactness of the embedding $\dot{H}^1(\Omega) \cap L^{2\sigma+2}(\Omega) \hookrightarrow H^1(\Omega) \hookrightarrow L^p(\Omega)$ for a bounded domain $\Omega \subset \mathbb{R}$ and $1 \le p \le \infty$. Thus, w_n converges to w weakly in $L^2(\mathbb{R})$. Therefore, by (2-6), we get (2-5).

Step 4. We prove $\alpha(2\sigma+2)\mu < \tilde{J}(\psi)$ if $\widetilde{K}(\psi) < 0$. By the definition of μ,

$$\mu_{\omega,c}^{\alpha,\beta} = \frac{1}{\alpha(2\sigma+2)}\inf\{\tilde{J}_{\omega,c}^{\alpha,\beta}(\psi) : \psi \in X_{\omega,c}\setminus\{0\},\ \widetilde{K}_{\omega,c}^{\alpha,\beta}(\psi) = 0\}. \tag{2-7}$$

If $\psi \in X_{\omega,c}$ satisfies $\widetilde{K}(\psi) < 0$, then there exists $\lambda_0 \in (0,1)$ such that $\widetilde{K}(\lambda_0\psi) = 0$ since $\widetilde{K}(\lambda\psi) > 0$ for small $\lambda \in (0,1)$. Therefore, we have $\alpha(2\sigma+2)\mu \le \tilde{J}(\lambda_0\psi) < \tilde{J}(\psi)$.

Step 5. We prove $\widetilde{K}(v) \le 0$ by contradiction. We suppose $\widetilde{K}(v) > 0$. Since $\widetilde{K}(v_n) \to 0$ and (2-5) hold,

$$\widetilde{K}(v - v_n) \to -\widetilde{K}(v) < 0.$$

This implies that $\widetilde{K}(v - v_n) < 0$ for large $n \in \mathbb{N}$. Therefore, by Step 4, we get $\alpha(2\sigma+2)\mu < \tilde{J}(v - v_n)$ for large $n \in \mathbb{N}$. By the same argument as in Step 3,

$$\tilde{J}(v_n) - \tilde{J}(v - v_n) - \tilde{J}(v) \to 0 \quad \text{as } n \to \infty.$$

Therefore, we get $\tilde{J}(v) = \lim_{n\to\infty}(\tilde{J}(v_n) - \tilde{J}(v - v_n)) \le 0$ since we have $\tilde{J}(v_n) \to \alpha(2\sigma+2)\mu$ by the definition of $\tilde{J}$ and $\widetilde{K}(v_n) \to 0$. By Step 2, we have $v \ne 0$ and then $\tilde{J}(v) > 0$. This is a contradiction.

Step 6. We prove that v belongs to $\tilde{\mathscr{M}}$. By (2-7) and the weakly lower semicontinuity of $\tilde{J}$, we obtain

$$\alpha(2\sigma+2)\mu \le \tilde{J}(v) \le \liminf_{n\to\infty} \tilde{J}(v_n) = \alpha(2\sigma+2)\mu.$$

Thus, $\tilde{J}(v) = \alpha(2\sigma+2)\mu$ and v_n converges to v strongly in $X_{\omega,c}$. Therefore, we get $\widetilde{S}(v) = \mu$ and $\widetilde{K}(v) = 0$ by Steps 4 and 5. □

Therefore, we obtain Proposition 2.1 when $\omega = c^2/4$ and $c > 0$.

The case of $\omega > c^2/4$ is much easier. Indeed, we can obtain Proposition 2.1 by the same argument as in the case $\omega = c^2/4$ and $c > 0$ by using $L^2(\mathbb{R})$ instead of $L^{2\sigma+2}(\mathbb{R})$. See also [Colin and Ohta 2006], where the statement only for the Nehari functional $K^{1,0}_{\omega,c}$ is obtained. Thus, we omit the proof.

3. Global existence

In this section, we show Theorem 1.2.

Proof of Theorem 1.2. Let u_0 belong to $\mathcal{K}^{\alpha,\beta,+}_{\omega,c}$. First, we consider the case that $K^{\alpha,\beta}_{\omega,c}(u_0) = 0$. Then, $u_0 = 0$ or $u_0 = e^{i\theta_0}\phi_{\omega,c}(\cdot - x_0)$ by Theorem 1.1. By the uniqueness of solution to (gDNLS), we have $u(t) = 0$ or $u(t) = e^{i\theta_0}e^{i\omega t}\phi_{\omega,c}(x - ct - x_0)$, respectively. This implies that $K^{\alpha,\beta}_{\omega,c}(u(t)) = 0$ for all time. This means that $u(t) \in \mathcal{K}^{\alpha,\beta,+}_{\omega,c}$ for all time. Next, we consider the case that $K^{\alpha,\beta}_{\omega,c}(u_0) > 0$. We suppose that there exists a time t such that $K^{\alpha,\beta}_{\omega,c}(u(t)) \le 0$. Then there exists t_* such that $K^{\alpha,\beta}_{\omega,c}(u(t_*)) = 0$ by the continuity of the flow. By the above argument, $K^{\alpha,\beta}_{\omega,c}(u(t)) = 0$ for all time. This is a contradiction. Thus, $u(t)$ belongs to $\mathcal{K}^{\alpha,\beta,+}_{\omega,c}$ for all time. When u_0 belongs to $\mathcal{K}^{\alpha,\beta,-}_{\omega,c}$, the same argument implies that $u(t)$ belongs to $\mathcal{K}^{\alpha,\beta,-}_{\omega,c}$ for all time. Next, we prove that the solution is global if $u_0 \in \mathcal{K}^{\alpha,\beta,+}_{\omega,c}$. Then, since

$$\alpha(2\sigma+2)S_{\omega,c}(\varphi) = K^{\alpha,\beta}_{\omega,c}(\varphi) + \frac{2\sigma\alpha+\beta}{2}\|\partial_x\varphi - \tfrac{1}{2}ci\varphi\|^2_{L^2} + (\omega - \tfrac{1}{4}c^2)\frac{2\sigma\alpha-\beta}{2}\|\varphi\|^2_{L^2} - \frac{\beta c}{2(2\sigma+2)}\|\varphi\|^{2\sigma+2}_{L^{2\sigma+2}} \tag{3-1}$$

and $K^{\alpha,\beta}_{\omega,c}(u(t)) > 0$ for all time t, we have that $\|\partial_x u(t) - \frac{1}{2}ciu(t)\|^2_{L^2}$ is uniformly bounded. Therefore,

$$\|\partial_x u(t)\|_{L^2} \le \|\partial_x u(t) - \tfrac{1}{2}ciu(t)\|_{L^2} + \tfrac{1}{2}|c|\|u(t)\|_{L^2} < C + \tfrac{1}{2}|c|\|u_0\|_{L^2},$$

for some positive constant C independent of t. This boundedness and the conservation law of the L^2-norm imply that u is global in time. □

We give proofs of Corollaries 1.3, 1.4, and 1.5. Direct calculations imply the following lemma (see [Colin and Ohta 2006] for the details).

Lemma 3.1. *Let $\sigma = 1$ and (ω, c) satisfy* (1-5)*. Then, we have the relations*

$$M(\phi_{\omega,c}) = 8\tan^{-1}\sqrt{\frac{2\sqrt{\omega}+c}{2\sqrt{\omega}-c}},$$
$$P(\phi_{\omega,c}) = 2\sqrt{4\omega - c^2},$$
$$E(\phi_{\omega,c}) = -\tfrac{1}{2}c\sqrt{4\omega - c^2}.$$

In particular,

$$S_{\omega,c}(\phi_{\omega,c}) = 4\omega\tan^{-1}\sqrt{\frac{2\sqrt{\omega}+c}{2\sqrt{\omega}-c}} + \tfrac{1}{2}c\sqrt{4\omega - c^2}.$$

Remark. When $\sigma = 1$, we have $M(\phi_{c^2/4,c}) = 4\pi$, $P(\phi_{c^2/4,c}) = 0$, and $E(\phi_{c^2/4,c}) = 0$ for all $c > 0$ by Lemma 3.1. On the other hand, if $M(\phi) = 4\pi$, $P(\phi) = 0$, and $E(\phi) \le 0$, then $\phi(x) = e^{i\theta_0}\phi_{c_0^2/4,c_0}(x - x_0)$

for some $\theta_0 \in \mathbb{R}$, $x_0 \in \mathbb{R}$, and $c_0 > 0$. Indeed, $M(\phi) = 4\pi$, $P(\phi) = 0$, and $E(\phi) \le 0$ imply that

$$K^{\alpha,\beta}_{c^2/4,c}(\phi) \le -\frac{2\alpha+\beta}{2}\|\partial_x\phi\|_{L^2}^2 + \frac{2\alpha-\beta}{2}c^2\pi + \frac{\beta c}{8}\|\phi\|_{L^4}^4.$$

Since $K^{\alpha,\beta}_{c^2/4,c}(\phi) < 0$ for small $c > 0$ and $K^{\alpha,\beta}_{c^2/4,c}(\phi) \to +\infty$ as $c \to \infty$, there exists $c_0 > 0$ such that $K^{\alpha,\beta}_{c_0^2/4,c_0}(\phi) = 0$. Therefore, Theorem 1.1 implies that $\phi(x) = e^{i\theta_0}\phi_{c_0^2/4,c_0}(x - x_0)$. Note that this means that there is no function satisfying $M(\phi) = 4\pi$, $P(\phi) = 0$, and $E(\phi) < 0$.

First, we prove Corollary 1.3.

Proof of Corollary 1.3. Let u_0 satisfy $\|u_0\|_{L^2}^2 < 4\pi$. The statement is trivial if $u_0 = 0$. We assume that $u_0 \neq 0$. Since $\|u_0\|_{L^2}^2 < 4\pi$,

$$S_{c^2/4,c}(u_0) = E(u_0) + \tfrac{1}{8}c^2\|u_0\|_{L^2}^2 + \tfrac{1}{2}cP(u_0) < c^2\pi/2,$$

for sufficiently large $c > 0$. Moreover, since $\|u_0\|_{L^2}^2 \neq 0$,

$$\begin{aligned} K^{\alpha,\beta}_{c^2/4,c}(u_0) &= \frac{2\alpha-\beta}{2}\|\partial_x u_0\|_{L^2}^2 + \frac{2\alpha-\beta}{2}\frac{c^2}{4}\|u_0\|_{L^2}^2 + \frac{2\alpha-\beta}{2}cP(u_0) + \frac{\beta c}{8}\|u_0\|_{L^4}^4 - \alpha N(u_0) \\ &\to \infty \quad \text{as } c \to \infty, \end{aligned} \tag{3-2}$$

for any (α, β) satisfying (1-8). Thus, $K^{\alpha,\beta}_{c^2/4,c}(u_0) > 0$ for large $c > 0$. Thus, there exists $c > 0$ such that $K^{\alpha,\beta}_{c^2/4,c}(u_0) > 0$ and $S_{c^2/4,c}(u_0) < c^2\pi/2$, where we note that $\mu_{c^2/4,c} = c^2\pi/2$ by Lemma 3.1 when $\sigma = 1$. By Theorem 1.2, the solution u is global. □

Secondly, we give a proof of Corollary 1.4 by Theorem 1.2.

Proof of Corollary 1.4. Let u_0 satisfy $\|u_0\|_{L^2}^2 = 4\pi$ and $P(u_0) < 0$. We recall that $\mu_{c^2/4,c} = c^2\pi/2$ by Lemma 3.1 when $\sigma = 1$. Since $P(u_0) < 0$, we have, for large $c > 0$,

$$S_{c^2/4,c}(u_0) = E(u_0) + \tfrac{1}{2}c^2\pi + \tfrac{1}{2}cP(u_0) \le \mu_{c^2/4,c}.$$

On the other hand, because $2\alpha - \beta > 0$ and $\|u_0\|_{L^2}^2 \neq 0$, we obtain (3-2). Thus, $K^{\alpha,\beta}_{c^2/4,c}(u_0) > 0$ for large $c > 0$. This means that the assumption in Theorem 1.2 holds for sufficiently large c. This implies that u is global. □

We give another proof. This is due to [Guo and Wu 2017].

Another proof of Corollary 1.4. We have

$$P(u) \ge \tfrac{1}{4}\|u\|_{L^4}^4\left(1 - \frac{\|u\|_{L^2}}{2\sqrt{\pi}}\right) - \frac{8\sqrt{\pi}E(u)\|u\|_{L^2}}{\|u\|_{L^4}^4},$$

applying the gauge transformation $u = w\exp(-\frac{1}{4}i\int_{-\infty}^{x}|w(y)|^2\,dy)$ to (1-15). See [Guo and Wu 2017, Lemma 2.1] for the proof of (1-15). When $\|u_0\|_{L^2}^2 = 4\pi$ and $P(u_0) < 0$, we get

$$\|u(t)\|_{L^4}^4 \le \frac{8\sqrt{\pi}E(u_0)\|u_0\|_{L^2}}{|P(u_0)|}. \tag{3-3}$$

Therefore, by the Hölder inequality, the Gagliardo–Nirenberg inequality, and the Young inequality,

$$\begin{aligned}\|\partial_x u(t)\|_{L^2}^2 &= 2E(u_0) + \tfrac{1}{2}\,\mathrm{Re}\int_{\mathbb{R}} i|u(t,x)|^2\overline{u(t,x)}\partial_x u(t,x)\,dx\\ &\le 2E(u_0) + \tfrac{1}{2}\|u(t)\|_{L^6}^3\|\partial_x u(t)\|_{L^2}\\ &\le 2E(u_0) + C\|u(t)\|_{L^4}^{8/3}\|\partial_x u(t)\|_{L^2}^{4/3}\\ &\le 2E(u_0) + C\|u(t)\|_{L^4}^8 + \tfrac{1}{2}\|\partial_x u(t)\|_{L^2}^2.\end{aligned}$$

This inequality and (3-3) give an a priori estimate, and thus, the solution is global. □

At last, we prove Corollary 1.5.

Proof of Corollary 1.5. Let $\sigma \ge 1$. Since $u_{0,c} = e^{(c/2)ix}\psi$,

$$\begin{aligned}S_{c^2/4,c}(u_{0,c}) &= \widetilde{S}_{c^2/4,c}(\psi)\\ &= \tfrac{1}{2}\|\partial_x\psi\|_{L^2}^2 + \frac{c}{2(2\sigma+2)}\|\psi\|_{L^{2\sigma+2}}^{2\sigma+2} - \frac{1}{2\sigma+2}N(\psi)\\ &\le c^{1+1/\sigma}S_{1/4,1}(\phi_{1/4,1}) = S_{c^2/4,c}(\phi_{c^2/4,c}),\\ K_{c^2/4,c}^{\alpha,\beta}(u_{0,c}) &= \widetilde{K}_{c^2/4,c}^{\alpha,\beta}(\psi)\\ &= \frac{2\alpha-\beta}{2}\|\partial_x\psi\|_{L^2}^2 + \frac{\{(2\sigma+2)\alpha+\beta\}c}{2(2\sigma+2)}\|\psi\|_{L^{2\sigma+2}}^{2\sigma+2} - \alpha N(\psi)\\ &\ge 0,\end{aligned}$$

for large $c > 0$. By Theorem 1.2, therefore, the solution u_c with the initial data $u_{0,c}$ is global for large $c > 0$. □

Appendix: Uniqueness and nonexistence

We prove the uniqueness of the massless elliptic equation.

Proposition A.1. *Let $1 < p < q < \infty$, $a > 0$, and $b > 0$. Assume there exists a nontrivial solution in $\dot{H}^1(\mathbb{R}) \cap L^{p+1}(\mathbb{R})$ of the equation*

$$-\varphi'' + a|\varphi|^{p-1}\varphi - b|\varphi|^{q-1}\varphi = 0 \tag{A-1}$$

in the distribution sense. Then there exist $\theta_0 \in [0, 2\pi)$ and $x_0 \in \mathbb{R}$ such that $\varphi = e^{i\theta_0}\psi(\cdot - x_0)$, where ψ is the unique positive, even, and decreasing function which satisfies (A-1).

Proof. Since $a|\varphi|^{p-1}\varphi - b|\varphi|^{q-1}\varphi$ belongs to $L^2(\mathbb{R})$, we obtain $\varphi \in \dot{H}^2(\mathbb{R})$. A bootstrap argument gives us that $\varphi \in \dot{H}^3(\mathbb{R})$. By the Sobolev embedding, $\varphi \in C^2(\mathbb{R})$ and φ satisfies the equation in the classical sense. Multiplying the equation by φ' and integrating on $(-\infty, x)$, we obtain

$$-\tfrac{1}{2}|\varphi'(x)|^2 + \frac{a}{p+1}|\varphi(x)|^{p+1} - \frac{b}{q+1}|\varphi(x)|^{q+1} = 0. \tag{A-2}$$

We write $\varphi = \rho e^{i\theta}$, where $\rho > 0$ and $\rho, \theta \in C^2(\mathbb{R})$. It is easily seen that $\theta \equiv \theta_0$ for some $\theta_0 \in [0, 2\pi)$. Since $\rho \in L^{p+1}(\mathbb{R})$, there must exist $x_0 \in \mathbb{R}$ such that $\rho'(x_0) = 0$. By (A-2), $\rho(x_0) = c$, where $c^{q-p} =$

$(a(q+1))/(b(p+1))$. Let ψ be the real-valued solution of (A-1) such that $\psi(0) = c$ and $\psi'(0) = 0$. Using the uniqueness of the ordinary differential equation, we can deduce that $\varphi = e^{i\theta_0}\psi(\cdot - x_0)$. □

We prove the nonexistence of a nontrivial solution to the nonlinear elliptic equation (1-4) in the case $\omega < c^2/4$, or $\omega = c^2/4$ and $c \leq 0$. See [Berestycki and Lions 1983, Theorem 5] for the necessary and sufficient condition for the existence of nontrivial solutions to more general second-order ordinary differential equations.

Proposition A.2. *Let $1 < p, q < \infty$. If $\varphi \in H^1(\mathbb{R})$ satisfies*

$$-\varphi'' + \omega\varphi + a|\varphi|^{p-1}\varphi - b|\varphi|^{q-1}\varphi = 0 \quad \textit{in the distribution sense},$$

where $a, b \in \mathbb{R}$ and $\omega < 0$, then we have $\varphi = 0$.

Proof. By a usual bootstrap argument [Cazenave 2003, §8], we have $\varphi \in H^3(\mathbb{R})$. We get $\varphi \in C^2(\mathbb{R})$ by the Sobolev embedding. Therefore, $\varphi'(x) \to 0$ and $\varphi(x) \to 0$ as $|x| \to \infty$. Multiplying the equation by φ' and integrating on $(-\infty, x)$, we obtain

$$-\tfrac{1}{2}|\varphi'(x)|^2 + \tfrac{1}{2}\omega|\varphi(x)|^2 + \frac{a}{p+1}|\varphi(x)|^{p+1} - \frac{b}{q+1}|\varphi(x)|^{q+1} = 0. \tag{A-3}$$

Since $\varphi(x) \to 0$ as $|x| \to \infty$, we get

$$\tfrac{1}{2}\omega|\varphi(x)|^2 + \frac{a}{p+1}|\varphi(x)|^{p+1} - \frac{b}{q+1}|\varphi(x)|^{q+1} < 0 \quad \text{for some } x$$

or

$$|\varphi(x)| = 0 \quad \text{for some } x.$$

In the former case, we obtain $|\varphi'(x)| < 0$ by (A-3). This is a contradiction. In the latter case, we obtain $|\varphi'(x)| = 0$ by (A-3). By the uniqueness of the ordinary differential equation, we get $\varphi = 0$. □

By the same argument as in the proof of Proposition A.2, we obtain the nonexistence of a nontrivial solution to the nonlinear elliptic equation (1-4) when $\omega = c^2/4$ and $c \leq 0$ as follows.

Proposition A.3. *Let $1 < p, q < \infty$. If $\varphi \in \dot{H}^1(\mathbb{R}) \cap L^{p+1}(\mathbb{R})$ satisfies*

$$-\varphi'' - a|\varphi|^{p-1}\varphi - b|\varphi|^{q-1}\varphi = 0 \quad \textit{in the distribution sense},$$

where $a \geq 0$ and $b > 0$, then we have $\varphi = 0$.

Acknowledgements

The authors would like to express deep appreciation to Professor Kenji Nakanishi for constant encouragement, Professor Masahito Ohta for many useful suggestions, and Professor Tohru Ozawa for advice on notations. Inui is supported by Grant-in-Aid for JSPS Research Fellow 15J02570. The authors also would like to thank Guixiang Xu for introducing their works, Zihua Guo for a suggestion about Corollary 1.4, and the anonymous referee for his valuable comments.

References

[Bellazzini et al. 2014a] J. Bellazzini, R. L. Frank, and N. Visciglia, "Maximizers for Gagliardo–Nirenberg inequalities and related non-local problems", *Math. Ann.* **360**:3–4 (2014), 653–673.

[Bellazzini et al. 2014b] J. Bellazzini, M. Ghimenti, and S. Le Coz, "Multi-solitary waves for the nonlinear Klein–Gordon equation", *Comm. Partial Differential Equations* **39**:8 (2014), 1479–1522.

[Berestycki and Lions 1983] H. Berestycki and P.-L. Lions, "Nonlinear scalar field equations, I: Existence of a ground state", *Arch. Rational Mech. Anal.* **82**:4 (1983), 313–345.

[Biagioni and Linares 2001] H. A. Biagioni and F. Linares, "Ill-posedness for the derivative Schrödinger and generalized Benjamin–Ono equations", *Trans. Amer. Math. Soc.* **353**:9 (2001), 3649–3659.

[Brézis and Lieb 1983] H. m. Brézis and E. Lieb, "A relation between pointwise convergence of functions and convergence of functionals", *Proc. Amer. Math. Soc.* **88**:3 (1983), 486–490.

[Cazenave 2003] T. Cazenave, *Semilinear Schrödinger equations*, Courant Lecture Notes in Mathematics **10**, American Mathematical Society, Providence, RI, 2003.

[Colin and Ohta 2006] M. Colin and M. Ohta, "Stability of solitary waves for derivative nonlinear Schrödinger equation", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **23**:5 (2006), 753–764.

[Colliander et al. 2001] J. Colliander, M. Keel, G. Staffilani, H. Takaoka, and T. Tao, "Global well-posedness for Schrödinger equations with derivative", *SIAM J. Math. Anal.* **33**:3 (2001), 649–669.

[Colliander et al. 2002] J. Colliander, M. Keel, G. Staffilani, H. Takaoka, and T. Tao, "A refined global well-posedness result for Schrödinger equations with derivative", *SIAM J. Math. Anal.* **34**:1 (2002), 64–86.

[Fukaya 2016] N. Fukaya, "Instability of solitary waves for a generalized derivative nonlinear Schrödinger equation in a borderline case", preprint, 2016. To appear in *Kodai Math. J.* arXiv

[Grillakis et al. 1987] M. Grillakis, J. Shatah, and W. Strauss, "Stability theory of solitary waves in the presence of symmetry, I", *J. Funct. Anal.* **74**:1 (1987), 160–197.

[Grillakis et al. 1990] M. Grillakis, J. Shatah, and W. Strauss, "Stability theory of solitary waves in the presence of symmetry, II", *J. Funct. Anal.* **94**:2 (1990), 308–348.

[Guo and Wu 1995] B. L. Guo and Y. P. Wu, "Orbital stability of solitary waves for the nonlinear derivative Schrödinger equation", *J. Differential Equations* **123**:1 (1995), 35–55.

[Guo and Wu 2017] Z. Guo and Y. Wu, "Global well-posedness for the derivative nonlinear Schrödinger equation in $H^{\frac{1}{2}}(\mathbb{R})$", *Discrete Contin. Dyn. Syst.* **37**:1 (2017), 257–264.

[Hao 2007] C. Hao, "Well-posedness for one-dimensional derivative nonlinear Schrödinger equations", *Commun. Pure Appl. Anal.* **6**:4 (2007), 997–1021.

[Hayashi 1993] N. Hayashi, "The initial value problem for the derivative nonlinear Schrödinger equation in the energy space", *Nonlinear Anal.* **20**:7 (1993), 823–833.

[Hayashi and Ozawa 1992] N. Hayashi and T. Ozawa, "On the derivative nonlinear Schrödinger equation", *Phys. D* **55**:1–2 (1992), 14–36.

[Hayashi and Ozawa 1994a] N. Hayashi and T. Ozawa, "Finite energy solutions of nonlinear Schrödinger equations of derivative type", *SIAM J. Math. Anal.* **25**:6 (1994), 1488–1503.

[Hayashi and Ozawa 1994b] N. Hayashi and T. Ozawa, "Remarks on nonlinear Schrödinger equations in one space dimension", *Differential Integral Equations* **7**:2 (1994), 453–461.

[Hayashi and Ozawa 2016] M. Hayashi and T. Ozawa, "Well-posedness for a generalized derivative nonlinear Schrödinger equation", *J. Differential Equations* **261**:10 (2016), 5424–5445.

[Kwon and Wu 2016] S. Kwon and Y. Wu, "Orbital stability of solitary waves for derivative nonlinear Schrödinger equation", preprint, 2016. arXiv

[Le Coz and Wu 2016] S. Le Coz and Y. Wu, "Stability of multi-solitons for the derivative nonlinear Schrödinger equation", preprint, 2016. arXiv

[Lieb 1983] E. H. Lieb, "On the lowest eigenvalue of the Laplacian for the intersection of two domains", *Invent. Math.* **74**:3 (1983), 441–448.

[Liu et al. 2013] X. Liu, G. Simpson, and C. Sulem, "Stability of solitary waves for a generalized derivative nonlinear Schrödinger equation", *J. Nonlinear Sci.* **23**:4 (2013), 557–583.

[Miao et al. 2011] C. Miao, Y. Wu, and G. Xu, "Global well-posedness for Schrödinger equation with derivative in $H^{\frac{1}{2}}(\mathbb{R})$", *J. Differential Equations* **251**:8 (2011), 2164–2195.

[Miao et al. 2017a] C. Miao, X. Tang, and G. Xu, "Solitary waves for nonlinear Schrödinger equation with derivative", preprint, 2017. arXiv

[Miao et al. 2017b] C. Miao, X. Tang, and G. Xu, "Stability of the traveling waves for the derivative Schrödinger equation in the energy space", *Calc. Var. Partial Differential Equations* **56**:2 (2017), 56:45.

[Mio et al. 1976] K. Mio, T. Ogino, K. Minami, and S. Takeda, "Modified nonlinear Schrödinger equation for Alfvén waves propagating along the magnetic field in cold plasmas", *J. Phys. Soc. Japan* **41**:1 (1976), 265–271.

[Mjølhus 1976] E. Mjølhus, "On the modulational instability of hydromagnetic waves parallel to the magnetic field", *J. Plasma Phys.* **16**:3 (1976), 321–334.

[Moses et al. 2007] J. Moses, B. A. Malomed, and F. W. Wise, "Self-steepening of ultrashort optical pulses without self-phase-modulation", *Phys. Rev. A* **76**:2 (2007), 021802(R).

[Ozawa 1996] T. Ozawa, "On the nonlinear Schrödinger equations of derivative type", *Indiana Univ. Math. J.* **45**:1 (1996), 137–163.

[Payne and Sattinger 1975] L. E. Payne and D. H. Sattinger, "Saddle points and instability of nonlinear hyperbolic equations", *Israel J. Math.* **22**:3-4 (1975), 273–303.

[Santos 2015] G. d. N. Santos, "Existence and uniqueness of solution for a generalized nonlinear derivative Schrödinger equation", *J. Differential Equations* **259**:5 (2015), 2030–2060.

[Sattinger 1968] D. H. Sattinger, "On global solution of nonlinear hyperbolic equations", *Arch. Rational Mech. Anal.* **30** (1968), 148–172.

[Takaoka 1999] H. Takaoka, "Well-posedness for the one-dimensional nonlinear Schrödinger equation with the derivative nonlinearity", *Adv. Differential Equations* **4**:4 (1999), 561–580.

[Takaoka 2001] H. Takaoka, "Global well-posedness for Schrödinger equations with derivative in a nonlinear term and data in low-order Sobolev spaces", *Electron. J. Differential Equations* **2001**:42 (2001).

[Tang and Xu 2017] X. Tang and G. Xu, "Stability of the sum of two solitary waves for (gDNLS) in the energy space", preprint, 2017. arXiv

[Tsutsumi 1972] M. Tsutsumi, "On solutions of semilinear differential equations in a Hilbert space", *Math. Japon.* **17** (1972), 173–193.

[Tsutsumi and Fukuda 1980] M. Tsutsumi and I. Fukuda, "On solutions of the derivative nonlinear Schrödinger equation: existence and uniqueness theorem", *Funkcial. Ekvac.* **23**:3 (1980), 259–277.

[Weinstein 1982] M. I. Weinstein, "Nonlinear Schrödinger equations and sharp interpolation estimates", *Comm. Math. Phys.* **87**:4 (1982), 567–576.

[Wu 2013] Y. Wu, "Global well-posedness for the nonlinear Schrödinger equation with derivative in energy space", *Anal. PDE* **6**:8 (2013), 1989–2002.

[Wu 2015] Y. Wu, "Global well-posedness on the derivative nonlinear Schrödinger equation", *Anal. PDE* **8**:5 (2015), 1101–1112.

Received 8 Oct 2016. Revised 28 Feb 2017. Accepted 3 Apr 2017.

NORIYOSHI FUKAYA: 1116702@ed.tus.ac.jp
Department of Mathematics, Graduate School of Science, Tokyo University of Science, Shinjuku, Tokyo 162-8601, Japan

MASAYUKI HAYASHI: masayuki-884@fuji.waseda.jp
Department of Applied Physics, Waseda University, Shinjuku, Tokyo 169-8555, Japan

TAKAHISA INUI: inui@math.kyoto-u.ac.jp
Department of Mathematics, Graduate School of Science, Kyoto University, Kyoto City, Kyoto 606-8502, Japan

ANALYSIS AND PDE
Vol. 10, No. 5, 2017
dx.doi.org/10.2140/apde.2017.10.1169

LOCAL DENSITY APPROXIMATION FOR THE ALMOST-BOSONIC ANYON GAS

MICHELE CORREGGI, DOUGLAS LUNDHOLM AND NICOLAS ROUGERIE

We study the minimizers of an energy functional with a self-consistent magnetic field, which describes a quantum gas of almost-bosonic anyons in the average-field approximation. For the homogeneous gas we prove the existence of the thermodynamic limit of the energy at fixed effective statistics parameter, and the independence of such a limit from the shape of the domain. This result is then used in a local density approximation to derive an effective Thomas–Fermi-like model for the trapped anyon gas in the limit of a large effective statistics parameter (i.e., "less-bosonic" anyons).

1. Introduction

A convenient description of two-dimensional particles with exotic quantum statistics (different from Bose–Einstein and Fermi–Dirac) is via effective magnetic interactions. We are interested in a mean-field model for such particles, known as anyons. Indeed, in a certain scaling limit ("almost-bosonic anyons", see [Lundholm and Rougerie 2015]), a suitable magnetic nonlinear Schrödinger theory becomes appropriate. The corresponding energy functional is given by

$$\mathcal{E}_\beta^{\mathrm{af}}[u] := \int_{\mathbb{R}^2} \big(\big| \big(-i\nabla + \beta \boldsymbol{A}[|u|^2]\big) u \big|^2 + V|u|^2 \big), \tag{1-1}$$

acting on functions $u \in H^1(\mathbb{R}^2)$. Here $V\colon \mathbb{R}^2 \to \mathbb{R}^+$ is a trapping potential confining the particles, and the vector potential $\boldsymbol{A}[|u|^2]\colon \mathbb{R}^2 \to \mathbb{R}^2$ is defined through

$$\boldsymbol{A}[\varrho] := \nabla^\perp w_0 * \varrho, \quad w_0(\boldsymbol{x}) := \log|\boldsymbol{x}|, \tag{1-2}$$

for $\varrho = |u|^2 \in L^1(\mathbb{R}^2)$ and $\boldsymbol{x}^\perp = (x, y)^\perp := (-y, x)$. Thus, the self-consistent magnetic field, given by

$$\operatorname{curl} \boldsymbol{A}[\varrho](\boldsymbol{x}) = \Delta w_0 * \varrho(\boldsymbol{x}) = 2\pi\varrho(\boldsymbol{x}),$$

MSC2010: 35Q40, 81V70, 81S05, 46N50.
Keywords: mean-field energy, anyons, fractional statistics, Thomas–Fermi theory, magnetic Schrödinger operator.

is proportional to the particles' density. The parameter $\beta \in \mathbb{R}$ then regulates the strength of the magnetic self-interactions and, for reasons explained below, we will call it the *scaled statistics parameter*. By symmetry of (1-1) under complex conjugation $u \mapsto \bar{u}$ we may and shall assume

$$\beta \geqslant 0$$

in the following. We will study the ground-state problem for (1-1), namely the minimization under the mass constraint

$$\int_{\mathbb{R}^2} |u|^2 = 1. \tag{1-3}$$

The functional $\mathcal{E}^{\mathrm{af}}$ bears some similarity with other mean-field models such as the Gross–Pitaevskii energy functional

$$\mathcal{E}^{\mathrm{GP}}[u] := \int_{\mathbb{R}^2} \big(|-i\nabla u + \boldsymbol{A}u|^2 + V|u|^2 + g|u|^4\big), \tag{1-4}$$

with *fixed* vector potential $\boldsymbol{A}$. The above describes a gas of interacting bosons in a certain mean-field regime [Lieb et al. 2005; Lieb and Seiringer 2006; Nam et al. 2016; Rougerie 2014; 2015]: the quartic term originates from short-range pair interactions. The crucial difference between (1-1) and (1-4) is that, while the interactions of $\mathcal{E}^{\mathrm{GP}}$ are scalar (with interaction strength $g \in \mathbb{R}$), those of $\mathcal{E}^{\mathrm{af}}$ are purely magnetic and therefore involve mainly the phase of the function u. There is extensive literature dealing with (1-4) (see [Aftalion 2007; Correggi et al. 2011; 2012; Correggi and Rougerie 2013]) and with the related Ginzburg–Landau model of superconductivity [Bethuel et al. 1994; Fournais and Helffer 2010; Sandier and Serfaty 2007; Sigal 2015]. That the interactions are via the magnetic field in (1-1) poses however quite a few new difficulties in the asymptotic analysis of minimizers we initiate here. Note indeed (see the variational equation in Lemma A.2) that the nonlinearity consists in a quintic nonlocal semilinear term and a cubic quasilinear term (also nonlocal), both being critical when compared to the usual Laplacian.

The functional $\mathcal{E}^{\mathrm{af}}$ arises in a mean-field description[1] of a gas of particles whose many-body quantum wave function can change under particle exchange by a phase factor $e^{i\alpha\pi}$ (with $\alpha \in \mathbb{R}$ known as the statistics parameter). This is a generalization of the usual types of particles: bosons have $\alpha = 0$ (symmetric wave functions) and their mean-field description is via models of the form (1-4), and fermions have $\alpha = 1$ (antisymmetric wave functions) and appropriate models for them are Hartree–Fock functionals (see [Bach 1992; Lieb and Simon 1977; Lions 1987; 1988; Fournais et al. 2015]). For general α one speaks of anyons [Khare 2005; Myrheim 1999; Ouvry 2009; Wilczek 1990], which are believed to emerge as quasiparticle excitations of certain condensed-matter systems [Arovas et al. 1984; Haldane 1983; Halperin 1984; Zhang et al. 2014; Cooper and Simon 2015; Lundholm and Rougerie 2016].

Anyons can be modeled as bosons (respectively, fermions) but with a many-body magnetic interaction of coupling strength α (respectively, $\alpha - 1$). It was shown in [Lundholm and Rougerie 2015] that the ground-state energy per particle of such a system is correctly described by the minimum of (1-1) (and the ground states by the corresponding minimizers) in a limit where, as the number of particles N goes to ∞, one takes $\alpha = \beta/N \to 0$. We refer to this limit as that of *almost-bosonic* anyons, with β determining how far we are from usual bosons.

[1] Usually referred to as an *average*-field description in this context.

In the following we treat the anyon gas as fully described by a one-body wave function $u \in H^1(\mathbb{R}^2)$ minimizing (1-1) under the mass constraint (1-3). We shall consider asymptotic regimes for this minimization problem. The limit $\beta \to 0$ is trivial and leads to a linear theory for noninteracting bosons (see [Lundholm and Rougerie 2015, Appendix A]). The limit $\beta \to \infty$ is more interesting and more physically relevant: in a physical situation, the statistics parameter α is fixed and finite and N large, so that taking $\beta \to \infty$ is the relevant regime, at least if one is allowed to exchange the two limits.

In an approximation that has been used frequently in the physics literature [Chitra and Sen 1992; Iengo and Lechner 1992; Li et al. 1992; Trugenberger 1992a; 1992b; Wen and Zee 1990; Westerberg 1993], the ground-state energy per particle of the N-particle anyon gas with statistics parameter α is given by

$$\frac{E_0(N)}{N} \approx \int_{\mathbb{R}^2} (2\pi|\alpha| N\varrho^2 + V\varrho). \tag{1-5}$$

This relies on assuming that each particle sees the others by their approximately constant *average* magnetic field $B(\boldsymbol{x}) \approx 2\pi\alpha N\varrho(\boldsymbol{x})$, with $\varrho(\boldsymbol{x}) \geqslant 0$ the local particle density (normalized to $\int_{\mathbb{R}^2} \varrho = 1$). In the ground state of this magnetic field (the lowest Landau level) this leads to a magnetic energy $|B| \approx 2\pi|\alpha|N\varrho$ per particle.[2]

In this work we prove that, for large β, the behavior of the functional (1-1) is captured at leading order by a Thomas–Fermi-type [Catto et al. 1998; Lieb 1981] energy functional of a form similar to the right-hand side of (1-5) with $|\alpha|N = \beta$. The coupling constant appearing in this functional is defined via the large-volume limit of the homogeneous anyon gas energy (i.e., the infimum of (1-1) confined to a bounded domain with $V = 0$). In particular we prove that this limit exists and is bounded from below by the value 2π predicted by (1-5). We do not know the exact value, but there are good reasons to believe that it is *not* equal to 2π, thus refining the simple approximations leading to (1-5).

We state our main theorems in Section 2 and present their proofs in Sections 3 and 4. The Appendix recalls a few facts concerning the minimizers of (1-1). In particular, although we do not need it for the proof of our main results, we derive the associated variational equation.

2. Main results

We now proceed to state our main theorems. We first discuss the large-volume limit for the homogeneous gas in Section 2A and then state our results about the trapped anyons functional (1-1) in Section 2B.

2A. ***Thermodynamic limit for the homogeneous gas.*** Let $\Omega \subset \mathbb{R}^2$ be a fixed bounded domain in $\mathbb{R}^2$, with the associated energy for almost-bosonic anyons confined to it:

$$\mathcal{E}^{\mathrm{af}}_{\Omega}[u] = \mathcal{E}^{\mathrm{af}}_{\Omega,\beta}[u] := \int_{\Omega} \left|\left(-i\nabla + \beta \boldsymbol{A}[|u|^2]\right)u\right|^2, \tag{2-1}$$

with

$$\boldsymbol{A}[|u|^2](\boldsymbol{x}) = \int_{\Omega} \nabla^{\perp} w_0(\boldsymbol{x} - \boldsymbol{y})|u(\boldsymbol{y})|^2 \,\mathrm{d}\boldsymbol{y}. \tag{2-2}$$

[2] Because of the periodicity of the exchange phase $e^{i\alpha\pi}$, it is known that such an approximation can only be valid for certain values of α and ϱ. See [Larson and Lundholm 2016; Lundholm 2016; Trugenberger 1992b] for further discussion.

We define two energies, with homogeneous Dirichlet boundary conditions

$$E_0(\Omega, \beta, M) := \inf\{\mathcal{E}^{\mathrm{af}}_{\Omega,\beta}[u] : u \in H^1_0(\Omega),\ \textstyle\int_\Omega |u|^2 = M\}, \tag{2-3}$$

and without boundary conditions,

$$E(\Omega, \beta, M) := \inf\{\mathcal{E}^{\mathrm{af}}_{\Omega,\beta}[u] : u \in H^1(\Omega),\ \textstyle\int_\Omega |u|^2 = M\}. \tag{2-4}$$

Of course, the last minimization leads to a magnetic Neumann boundary condition for the solutions. We are interested in the thermodynamic limit of these quantities, i.e., the scaling limit in which the size of the domain tends to ∞ with fixed density $\rho := M/|\Omega|$ and the normalization changes accordingly.

Theorem 2.1 (Thermodynamic limit for the homogeneous anyon gas).
Let $\Omega \subset \mathbb{R}^2$ be a bounded simply connected domain with Lipschitz boundary, and let $\beta \geqslant 0$ and $\rho \geqslant 0$ be fixed parameters. Then, the limits

$$e(\beta, \rho) := \lim_{L\to\infty} \frac{E(L\Omega, \beta, \rho L^2|\Omega|)}{L^2|\Omega|} = \lim_{L\to\infty} \frac{E_0(L\Omega, \beta, \rho L^2|\Omega|)}{L^2|\Omega|} \tag{2-5}$$

exist, coincide and are independent of Ω. Moreover,

$$e(\beta, \rho) = \beta\rho^2 e(1, 1). \tag{2-6}$$

Remark 2.2 (Error estimate).
A close inspection of the proof reveals that we also have an estimate of the error appearing in (2-5), which coincides with the error appearing in the estimate of the difference between the Neumann and Dirichlet energies in a box (Lemma 3.8). Such a quantity is expected to be of the order of the box's side length L, which is subleading if compared to the total energy of order L^2. Our error estimate $O(L^{12/7+\varepsilon})$ (see (3-26)) is however much larger and far from being optimal. ⋄

The above result defines the thermodynamic energy per unit area at scaled statistics parameter β and density ρ, denoted $e(\beta, \rho)$, and shows that it has a nice scaling property. The latter is responsible for the occurrence of a Thomas–Fermi-type functional in the trapped anyons case. The fact that $e(\beta, \rho)$ does not depend on boundary conditions is a crucial technical ingredient in our study of the trapped case. This is very different from the usual Schrödinger energy in a fixed external magnetic field, for example, a constant one, for which the type of boundary conditions do matter (see, e.g., [Fournais and Helffer 2010, Chapter 5]).

The constant $e(1, 1)$ will be used to define a corresponding coupling parameter below. One may observe that (see Lemma 3.7)

$$e(1, 1) \geqslant 2\pi, \tag{2-7}$$

and we conjecture that this inequality is actually *strict*, contrary to what might be expected when comparing to the coupling constant of the conventional (constant-field) average-field approximation (1-5). The reason for this is that the self-interaction encoded by the functional $\mathcal{E}^{\mathrm{af}}$ has not been fully incorporated in (1-5). In fact, the lower bound (2-7) is based on a magnetic L^4-bound (Lemma 3.2) which is saturated only for constant functions, and hence for constant densities, which certainly is compatible with (1-5) in the case

of homogeneous traps. On the other hand, in order to minimize the magnetic energy in (2-1) for large β, the function has to have a large phase circulation and therefore also a large vorticity. This suggests the formation of an approximately homogeneous vortex lattice, in some analogy to the Abrikosov lattice that arises in superconductivity and in rotating bosonic gases [Aftalion 2007; Correggi and Yngvason 2008; Sandier and Serfaty 2007]. Such a picture has already been hinted at in [Chen et al. 1989, p. 1012] for the almost-bosonic gas. However the implication that the actual coupling constant may then be larger than the one expected from (1-5) seems not to have been observed in the literature before.

One should note here that there is a certain abuse of language in using the term "thermodynamic limit". Indeed, we consider the large-volume behavior of a mean-field energy functional, and there is no guarantee that this rigorously approximates the true thermodynamic energy of the underlying many-body system.

2B. ***Local density approximation for the trapped gas.*** We now return to (1-1) and discuss the ground state problem

$$E_\beta^{\mathrm{af}} := \min\bigl\{\mathcal{E}_\beta^{\mathrm{af}}[u] : u \in H^1(\mathbb{R}^2),\ V|u|^2 \in L^1(\mathbb{R}^2),\ \textstyle\int_{\mathbb{R}^2} |u|^2 = 1\bigr\}. \tag{2-8}$$

We denote by u^{af} any associated minimizer. We refer to the Appendix (see also [Lundholm and Rougerie 2015, Appendix A]) for a discussion of the minimization domain as well as the existence of a minimizer. In the limit $\beta \to \infty$, the simpler Thomas–Fermi-like functional

$$\mathcal{E}^{\mathrm{TF}}[\varrho] = \mathcal{E}_\beta^{\mathrm{TF}}[\varrho] := \int_{\mathbb{R}^2} (\beta e(1,1)\varrho^2 + V\varrho) \tag{2-9}$$

emerges, whose ground-state energy we denote by

$$E_\beta^{\mathrm{TF}} := \min\bigl\{\mathcal{E}_\beta^{\mathrm{TF}}[\varrho] : \varrho \in L^2(\mathbb{R}^2;\mathbb{R}^+),\ V\varrho \in L^1(\mathbb{R}^2),\ \textstyle\int_{\mathbb{R}^2} \varrho = 1\bigr\}, \tag{2-10}$$

with associated (unique) minimizer $\varrho_\beta^{\mathrm{TF}}$. Here $e(1,1)$ is the fixed, universal constant defined by Theorem 2.1.

A typical potential one could have in mind for physical relevance is a harmonic trap, $V(\boldsymbol{x}) = c|\boldsymbol{x}|^2$, or an asymmetric trap, $V(x,y) = c_1x^2 + c_2y^2$. We shall work under the assumption that V is homogeneous of degree s and smooth:

$$V(\lambda\boldsymbol{x}) = \lambda^s V(\boldsymbol{x}), \quad V \in C^\infty(\mathbb{R}^2). \tag{2-11}$$

These conditions can be relaxed significantly; in particular we could extend the approach to asymptotically homogeneous potentials as defined in [Lieb et al. 2001, Definition 1.1]. We refrain from doing so to avoid lengthy technical discussions in the proofs. We shall always impose that V is trapping in the sense that it grows superlinearly at infinity, i.e., $s > 1$ and

$$\min_{|\boldsymbol{x}|\geqslant R} V(\boldsymbol{x}) \to \infty \quad \text{as } R \to \infty. \tag{2-12}$$

The Thomas–Fermi (TF) problem (2-10) has the merit of being exactly soluble. We obtain by scaling

$$E_\beta^{\mathrm{TF}} = \beta^{s/(s+2)} E_1^{\mathrm{TF}}, \quad \varrho_\beta^{\mathrm{TF}}(\boldsymbol{x}) = \beta^{-2/(2+s)} \varrho_1^{\mathrm{TF}}(\beta^{-1/(s+2)}\boldsymbol{x}), \tag{2-13}$$

and by an explicit computation

$$\varrho_1^{\mathrm{TF}}(\boldsymbol{x}) = \frac{1}{2e(1,1)}(\lambda_1^{\mathrm{TF}} - V(\boldsymbol{x}))_+, \tag{2-14}$$

with the chemical potential

$$\lambda_1^{\mathrm{TF}} = E_1^{\mathrm{TF}} + e(1,1)\int_{\mathbb{R}^2}(\varrho_1^{\mathrm{TF}})^2. \tag{2-15}$$

Clearly the above considerations imply

$$\mathrm{supp}(\varrho_\beta^{\mathrm{TF}}) \subset B_{C\beta^{1/(2+s)}}(0), \tag{2-16}$$

where $B_R(\boldsymbol{x})$ stands for a ball (disk) of radius R centered at $\boldsymbol{x}$, and the estimates

$$\|\varrho_\beta^{\mathrm{TF}}\|_{L^\infty(\mathbb{R}^2)} \leqslant C\beta^{-2/(2+s)}, \quad \|\nabla\varrho_\beta^{\mathrm{TF}}\|_{L^\infty(\mathbb{R}^2)} \leqslant C\beta^{-3/(2+s)} \tag{2-17}$$

for some fixed constant $C > 0$. Noticing that ϱ_1^{TF} vanishes along a level curve of the smooth homogeneous potential V, we also have the nondegeneracy

$$|\partial_{\boldsymbol{n}} V| \neq 0 \quad \text{a.e. on } \partial\,\mathrm{supp}(\varrho_1^{\mathrm{TF}}), \tag{2-18}$$

where $\boldsymbol{n}$ denotes the (say outward) normal vector to $\partial\,\mathrm{supp}(\varrho_1^{\mathrm{TF}})$.

We have the following result showing the accuracy of TF theory to determine the leading order of the minimization problem (2-8):

Theorem 2.3 (Local density approximation for the anyon gas).
Let V satisfy (2-11) *and* (2-12). *In the limit $\beta \to \infty$ we have the energy convergence*

$$\lim_{\beta\to+\infty} \frac{E_\beta^{\mathrm{af}}}{E_\beta^{\mathrm{TF}}} = 1. \tag{2-19}$$

Moreover, for any function u^{af} achieving the infimum (2-8), *with $\varrho^{\mathrm{af}} := |u^{\mathrm{af}}|^2$, we have for any $R > 0$*

$$\left\| \beta^{2/(2+s)}\varrho^{\mathrm{af}}(\beta^{1/(2+s)}\,\cdot\,) - \varrho_1^{\mathrm{TF}} \right\|_{W^{-1,1}(B_R(0))} \to 0 \quad \textit{as } \beta \to 0, \tag{2-20}$$

where $W^{-1,1}(B_R(0))$ is the dual space of Lipschitz functions on the ball $B_R(0)$.

Remark 2.4 (Extension to more general potentials).
The result can be straightforwardly extended to asymptotically homogeneous potentials, i.e., functions $V(\boldsymbol{x})$ that satisfy the following property [Lieb et al. 2001, Definition 1.1]: there exists another function $\widetilde{V}$, nonvanishing for $\boldsymbol{x} \neq 0$, such that, for some $s > 0$,

$$\lim_{\lambda\to\infty} \frac{\lambda^{-s}V(\lambda\boldsymbol{x}) - \widetilde{V}(\boldsymbol{x})}{1 + |\widetilde{V}(\boldsymbol{x})|} = 0 \tag{2-21}$$

uniformly in $\boldsymbol{x} \in \mathbb{R}^2$. The function $\widetilde{V}$ is necessarily homogeneous of degree $s > 0$ and, if we denote by $\widetilde{\mathcal{E}}_\beta$ the TF functional (2-9) with $\widetilde{V}$ in place of V, we have

$$E_\beta^{\mathrm{TF}} = (1+o(1))\widetilde{E}_\beta^{\mathrm{TF}} \quad \text{and} \quad \widetilde{E}_\beta^{\mathrm{TF}} = \beta^{s/(s+2)}\widetilde{E}_1^{\mathrm{TF}} \quad \text{as } \beta \to \infty. \qquad \diamond$$

Remark 2.5 (Density approximation on finer length scales).
We conjecture that the estimate (2-20) can be improved to show that ϱ^{af} is close to $\varrho_\beta^{\mathrm{TF}}$ on finer scales. Namely (2-20) implies that they are close on length scales of order $\beta^{1/(2+s)}$, which is the extent of the support of $\varrho_\beta^{\mathrm{TF}}$, but we expect them to be close on scales $\gg \beta^{-s/(2(s+2))}$, which is the smallest length scale appearing in our proofs. We however believe that the density convergence *cannot* hold on scales smaller than $\beta^{-s/(2(s+2))}$, for we expect the latter to be the length scale of a vortex lattice developed by minimizers. ◇

Remark 2.6 (Large β limit for the homogeneous gas on bounded domains).
We can think of the homogeneous gas by formally taking the limit $s \to \infty$ of the homogeneous potentials we have considered so far, which naturally leads to the restriction of the functional $\mathcal{E}^{\mathrm{af}}$ in (1-1) to bounded domains Ω with $V = 0$ and Dirichlet boundary conditions, that is, (2-1)–(2-3). In fact, we have by the scaling laws discussed in Section 3B,

$$\lim_{\beta\to+\infty} \frac{E(\Omega, \beta, 1)}{\beta} = \lim_{\beta\to+\infty} \frac{E_0(\Omega, \beta, 1)}{\beta} = |\Omega|^{-1} e(1, 1) \tag{2-22}$$

for any bounded and simply connected Ω with Lipschitz boundary. Convergence of the density to the TF minimizer ϱ_1^{TF} holds true in the same form as in (2-20). In this case ϱ_1^{TF} is simply the constant function on the domain (confirming that the gas is indeed homogeneous). The shortest length scale on which we expect (but cannot prove) the density convergence is $\beta^{-1/2}$, which should be the typical length scale of the vortex structure. ◇

3. Proofs for the homogeneous gas

The basic ingredient of the proof for the inhomogeneous case is the understanding of the thermodynamic limit of the model where the trap is replaced by a finite domain with sharp walls. We discuss this here, proving Theorem 2.1 and defining the constant $e(1, 1)$ appearing in the TF functional (2-9). For the sake of concreteness we first set

$$e(\beta, \rho) := \liminf_{L\to\infty} \frac{E_0(L\Omega, \beta, \rho L^2|\Omega|)}{L^2|\Omega|} \tag{3-1}$$

for Ω equal to a unit square and observe that such a quantity certainly exists and is nonnegative. At this stage it might as well be infinite but we are going to prove that actually the limit exists, is finite, and furthermore is independent of the domain shape.

We briefly outline here the plan for the proof: Section 3A contains basic technical estimates that we are going to use throughout the paper. Section 3B contains the proof of a crucial scaling property of the energy in the homogeneous case. In Section 3C we prove the existence of the thermodynamic limit for the case of squares, and then extend the result to general domains.

3A. ***Toolbox.*** Let us gather a few lemmas that will be used repeatedly in the sequel. We start with a variational a priori upper bound confirming that the energy scales like the area. The idea of the proof, relying deeply on the magnetic nature of the interaction, will be employed again several times.

Lemma 3.1 (Trial upper bound).
For any fixed bounded domain Ω, and β, $\rho \geqslant 0$, there exists a constant $C > 0$ such that

$$\frac{E(L\Omega, \beta, \rho L^2|\Omega|)}{L^2} \leqslant \frac{E_0(L\Omega, \beta, \rho L^2|\Omega|)}{L^2} \leqslant C, \quad \textit{for all } L \geqslant 1.$$

Proof. Since $H_0^1(\Omega) \subseteq H^1(\Omega)$, it is trivial that the Dirichlet energy is an upper bound to the Neumann energy. Let us then prove the second inequality.

We fill the domain $L\Omega$ with $N \sim L^2$ subdomains on which we use fixed trial states with Dirichlet boundary conditions. The crucial observation is that the magnetic interactions between subdomains can be canceled by a suitable choice of phase (local gauge transformation). For concreteness we here take disks as our subdomains.

Let $f \in C_c^\infty(B_1(0); \mathbb{R}^+)$ be a radial function with $\int_{B_1(0)} |f|^2 = 1$, and let

$$u_j(\boldsymbol{x}) := \sqrt{\omega_N} f(\boldsymbol{x} - \boldsymbol{x}_j) \in C_c^\infty(B_j), \quad \omega_N := \rho L^2|\Omega|/N.$$

Here the points $\boldsymbol{x}_j$, $j = 1, \dots, N$, are distributed in $L\Omega$ in such a way that the disks $B_j := B_1(\boldsymbol{x}_j)$ are contained in $L\Omega$ and disjoint, with $N \sim c|L\Omega|$ as $L \to \infty$ for some $c > 0$. Hence

$$\lim_{N\to\infty} \omega_N = \rho/c.$$

Take then the trial state

$$u(\boldsymbol{x}) := \sum_{j=1}^N u_j(\boldsymbol{x}) e^{-i\beta\omega_N \sum_{k\neq j} \arg(\boldsymbol{x}-\boldsymbol{x}_k)} \in C_c^\infty(L\Omega).$$

Note that its phase is smooth on each piece B_j of its support and that

$$|u(\boldsymbol{x})|^2 = \sum_{j=1}^N |u_j(\boldsymbol{x})|^2 = \begin{cases} |u_j(\boldsymbol{x})|^2 & \text{for } \boldsymbol{x} \in B_j, \\ 0 & \text{otherwise,} \end{cases}$$

and hence

$$\int_{L\Omega} |u|^2 = N\omega_N = \rho L^2|\Omega|.$$

Then

$$\begin{aligned}
\mathcal{E}^{\mathrm{af}}_{\Omega,\beta}[u] &= \sum_{j=1}^N \int_{B_j} \left|\left(-i\nabla + \beta \textstyle\sum_{k=1}^N \boldsymbol{A}[|u_k|^2]\right) e^{-i\beta\omega_N \sum_{k\neq j} \arg(\boldsymbol{x}-\boldsymbol{x}_k)} u_j(\boldsymbol{x})\right|^2 \mathrm{d}\boldsymbol{x} \\
&= \sum_{j=1}^N \int_{B_j} \left|\left(-i\nabla + \beta \boldsymbol{A}[|u_j|^2] + \textstyle\sum_{k\neq j} \left(\beta \boldsymbol{A}[|u_k|^2] - \beta\omega_N \nabla \arg(\boldsymbol{x} - \boldsymbol{x}_k)\right)\right) u_j(\boldsymbol{x})\right|^2 \mathrm{d}\boldsymbol{x} \\
&= \sum_{j=1}^N \int_{B_j} \left|\left(-i\nabla + \beta \boldsymbol{A}[|u_j|^2]\right) u_j\right|^2 = N\omega_N \int_{B_1(0)} \left|\left(-i\nabla + \beta\omega_N \boldsymbol{A}[|f|^2]\right) f\right|^2,
\end{aligned}$$

where we used that by Newton's theorem [Lieb and Loss 2001, Theorem 9.7]

$$\boldsymbol{A}[|u_k|^2](\boldsymbol{x}) = \nabla^\perp \int_{B_k} \ln|\boldsymbol{x} - \boldsymbol{y}|\, |u_k(\boldsymbol{y})|^2 \,\mathrm{d}\boldsymbol{y} = \nabla^\perp \ln|\boldsymbol{x} - \boldsymbol{x}_k| \int_{B_k} |u_k|^2 \,\mathrm{d}\boldsymbol{y} = \omega_N \nabla \arg(\boldsymbol{x} - \boldsymbol{x}_k)$$

for $\boldsymbol{x} \notin B_k$. It then follows that

$$E_0(L\Omega, \beta, \rho L^2|\Omega|) \leqslant \mathcal{E}^{\mathrm{af}}_{\Omega,\beta}[u] \leqslant N\omega_N\big(\|\nabla f\|_{L^2} + \beta\omega_N \big\| \boldsymbol{A}[|f|^2]f\big\|_{L^2}\big)^2 \leqslant CL^2$$

for some large enough constant $C > 0$ independent of N or L (but possibly depending on β, ρ and Ω). □

The following well-known inequalities provide useful a priori bounds on the functional's minimizers:

Lemma 3.2 (Elementary magnetic inequalities).
Diamagnetic inequality: *for any $\beta \in \mathbb{R}$ and $u \in H^1(\Omega)$,*

$$\int_\Omega \big|\big(\nabla + i\beta \boldsymbol{A}[|u|^2]\big)u\big|^2 \geqslant \int_\Omega \big|\nabla|u|\big|^2. \tag{3-2}$$

Magnetic L^4 bound: *for any $\beta \in \mathbb{R}$ and $u \in H^1_0(\Omega)$,*

$$\int_\Omega \big|\big(\nabla + i\beta \boldsymbol{A}[|u|^2]\big)u\big|^2 \geqslant 2\pi|\beta| \int_\Omega |u|^4. \tag{3-3}$$

Proof. The diamagnetic inequality is, e.g., given in [Lieb and Loss 2001, Theorem 7.21], while the L^4 bound follows immediately from the well-known inequality

$$\int_\Omega \big|(\nabla + i\boldsymbol{A})u\big|^2 \geqslant \pm \int_\Omega \operatorname{curl} \boldsymbol{A}\, |u|^2, \quad u \in H^1_0(\Omega); \tag{3-4}$$

see, e.g., [Fournais and Helffer 2010, Lemma 1.4.1].

A proof of (3-4) is to integrate the identity

$$|(\nabla + i\boldsymbol{A})u|^2 = \big|((\partial_1 + iA_1) \pm i(\partial_2 + iA_2))u\big|^2 \pm \operatorname{curl}\, \boldsymbol{J}[u] \pm \boldsymbol{A}\cdot\nabla^\perp|u|^2,$$

with

$$\boldsymbol{J}[u] := \frac{i}{2}(u\nabla\bar{u} - \bar{u}\nabla u).$$

Thanks to the Dirichlet boundary conditions, the integral of the next-to-last term vanishes, while the last one can be integrated by parts yielding

$$\mp \int_\Omega \operatorname{curl} \boldsymbol{A}\, |u|^2.$$

Again, no boundary terms are present because of the vanishing of u on $\partial\Omega$. Dirichlet boundary conditions are necessary since the bound (3-4) (resp. (3-3)) is otherwise invalid as $\boldsymbol{A} \to 0$ (resp. $\beta \to 0$), as can be seen by taking the trial state $u \equiv 1$. □

In order to perform energy localizations we shall also need an IMS-type inequality,[3] i.e., a suitable generalization of the well-known localization formula [Cycon et al. 1987, Theorem 3.2]:

$$|\nabla u|^2 = |\nabla(\chi u)|^2 + |\nabla(\eta u)|^2 - \big(|\nabla\chi|^2 + |\nabla\eta|^2\big)|u|^2, \tag{3-5}$$

where χ^2, η^2 form a partition of unity.

[3]The initials IMS may refer either to Israel Michael Sigal or to Ismagilov–Morgan–Simon.

Lemma 3.3 (IMS formula).
Let $\Omega \subseteq \mathbb{R}^2$ be a domain with Lipschitz boundary and $\chi^2 + \eta^2 = 1$ be a partition of unity such that $\chi \in C_c^\infty(\Omega)$ and $\operatorname{supp}\chi$ *is simply connected. Then, for any $u \in H^1(\Omega)$ and $\beta \in \mathbb{R}$,*

$$\mathcal{E}^{\mathrm{af}}_{\Omega,\beta}[u] = \int_\Omega \big|\big(\nabla + i\beta \boldsymbol{A}[|u|^2]\big)(\chi u)\big|^2 + \int_\Omega \big|\big(\nabla + i\beta \boldsymbol{A}[|u|^2]\big)(\eta u)\big|^2 - \int_\Omega \big(|\nabla\chi|^2 + |\nabla\eta|^2\big)|u|^2, \quad (3\text{-}6)$$

where

$$\int_\Omega \big|\big(\nabla + i\beta \boldsymbol{A}[|u|^2]\big)(\eta u)\big|^2 \geqslant \int_\Omega \big|\nabla|\eta u|\big|^2 \quad (3\text{-}7)$$

and

$$\int_\Omega \big|\big(\nabla + i\beta \boldsymbol{A}[|u|^2]\big)(\chi u)\big|^2 \geqslant \begin{cases} \displaystyle\int_\Omega \big|\nabla|\chi u|\big|^2, \\ \displaystyle 2\pi|\beta| \int_\Omega \chi^2 |u|^4, \\ \displaystyle (1-\varepsilon)\mathcal{E}^{\mathrm{af}}_{\Omega,\beta}[\psi] - (\varepsilon^{-1}-1)\beta^2 \int_\Omega \big|\boldsymbol{A}\big[|\eta u|^2 \mathbb{1}_K\big]\big|^2 |\chi u|^2, \end{cases} \quad (3\text{-}8)$$

with $\varepsilon \in (0,1)$ arbitrary, $K := \operatorname{supp}\chi \cap \operatorname{supp}\eta$, and $\psi = e^{i\beta\phi}\chi u \in H^1_0(\operatorname{supp}\chi)$ for some harmonic function $\phi \in C^2(\operatorname{supp}\chi)$.

Proof. We expand

$$\mathcal{E}^{\mathrm{af}}_{\Omega,\beta}[u] = \int_\Omega |\nabla u|^2 + 2\beta \int_\Omega \boldsymbol{A}[|u|^2]\cdot \boldsymbol{J}[u] + \beta^2 \int_\Omega \big|\boldsymbol{A}[|u|^2]\big|^2 |u|^2.$$

For the first term we use the standard IMS formula (3-5), while for the term involving $\boldsymbol{J}$ we have

$$\begin{aligned} \frac{2}{i}\big(\boldsymbol{J}[\chi u] + \boldsymbol{J}[\eta u]\big) &= u\chi\nabla(\chi\bar u) + u\eta\nabla(\eta\bar u) - \bar u\chi\nabla(\chi u) - \bar u\eta\nabla(\eta u) \\ &= u(\chi^2+\eta^2)\nabla\bar u - \bar u(\chi^2+\eta^2)\nabla u = \frac{2}{i}\boldsymbol{J}[u]. \end{aligned}$$

We can then recollect the terms to obtain (3-6). Equation (3-7) and the first version of (3-8) follow from the diamagnetic inequality (3-2), while the second version of (3-8) follows from the magnetic bound (3-3) with Dirichlet boundary conditions. For the third version we write

$$\int_\Omega \big|\big(\nabla + i\beta \boldsymbol{A}[|u|^2]\big)(\chi u)\big|^2 = \int_\Omega \big|\big(\nabla + i\beta \boldsymbol{A}[|\chi u|^2] + i\beta \boldsymbol{A}[|\eta u|^2 \mathbb{1}_K] + i\beta\big(\boldsymbol{A}[|\eta u|^2 \mathbb{1}_{K^c}] - \nabla\phi\big)\big)(e^{i\beta\phi}\chi u)\big|^2,$$

where the last magnetic term vanishes by taking the gauge choice

$$\phi(\boldsymbol{x}) := \int_{K^c} \arg(\boldsymbol{x}-\boldsymbol{y})|\eta u(\boldsymbol{y})|^2 \,\mathrm{d}\boldsymbol{y}, \quad \boldsymbol{x} \in \operatorname{supp}\chi.$$

Thus, noting that $|\chi u|^2 = |\psi|^2$,

$$\int_\Omega \big|\big(\nabla + i\beta \boldsymbol{A}[|u|^2]\big)(\chi u)\big|^2 = \int_\Omega \big|\big(\nabla + i\beta \boldsymbol{A}[|\psi|^2]\big)\psi + i\beta \boldsymbol{A}[|\eta u|^2 \mathbb{1}_K]\psi\big|^2,$$

and we can conclude by expanding the square and bounding the cross-term using Cauchy–Schwarz. □

3B. ***Scaling laws.*** In fact the large β and large volume limits are equivalent, as follows from the simple observation:

Lemma 3.4 (Scaling laws for the homogeneous gas).
For any domain $\Omega \subset \mathbb{R}^2$ and $\lambda, \mu > 0$ we have

$$E(\Omega, \beta, M) = \frac{1}{\lambda^2} E\left(\mu\Omega, x\frac{\beta}{\lambda^2\mu^2}, \lambda^2\mu^2 M\right), \tag{3-9}$$

and an identical scaling relation holds true for $E_0(\Omega, \beta, M)$.

Proof. Given any $u \in H^1(\Omega)$ we may set

$$u_{\lambda,\mu}(\boldsymbol{x}) := \lambda u(\boldsymbol{x}/\mu), \tag{3-10}$$

and observe that $u_{\lambda,\mu} \in H^1(\mu\Omega)$,

$$\int_{\mu\Omega} |u_{\lambda,\mu}|^2 = \lambda^2\mu^2 \int_\Omega |u|^2 \quad \text{and} \quad \mathcal{E}^{\mathrm{af}}_{\mu\Omega,\beta}[u_{\lambda,\mu}] = \lambda^2 \mathcal{E}^{\mathrm{af}}_{\Omega,\beta\lambda^2\mu^2}[u].$$

Namely, using $\nabla^\perp w_0(\boldsymbol{x}) = \boldsymbol{x}^{-\perp} := \boldsymbol{x}^\perp/|\boldsymbol{x}|^2$ and

$$\begin{aligned}\boldsymbol{A}_{\mu\Omega}[|u_{\lambda,\mu}|^2](\boldsymbol{x}) &= \int_{\mu\Omega} (\boldsymbol{x}-\boldsymbol{y})^{-\perp} |u_{\lambda,\mu}(\boldsymbol{y})|^2 \,\mathrm{d}\boldsymbol{y} = \lambda^2 \int_{\mu\Omega} (\boldsymbol{x}-\boldsymbol{y})^{-\perp} |u(\boldsymbol{y}/\mu)|^2 \,\mathrm{d}\boldsymbol{y} \\ &= \lambda^2\mu \int_\Omega (\boldsymbol{x}/\mu - z)^{-\perp} |u(z)|^2 \,\mathrm{d}z = \lambda^2\mu \boldsymbol{A}_\Omega[|u|^2](\boldsymbol{x}/\mu),\end{aligned}$$

we have

$$\begin{aligned}\mathcal{E}^{\mathrm{af}}_{\mu\Omega,\beta}[u_{\lambda,\mu}] &= \int_{\mu\Omega} \left|\nabla u_{\lambda,\mu}(\boldsymbol{x}) + i\beta \boldsymbol{A}_{\mu\Omega}[|u_{\lambda,\mu}|^2](\boldsymbol{x}) u_{\lambda,\mu}(\boldsymbol{x})\right|^2 \mathrm{d}\boldsymbol{x} \\ &= \int_{\mu\Omega} \left|\lambda\mu^{-1}(\nabla u)(\boldsymbol{x}/\mu) + i\beta\lambda^3\mu \boldsymbol{A}_\Omega[|u|^2](\boldsymbol{x}/\mu) u(\boldsymbol{x}/\mu)\right|^2 \mathrm{d}\boldsymbol{x} \\ &= \lambda^2\mu^{-2} \int_{\mu\Omega} \left|(\nabla u)(\boldsymbol{x}/\mu) + i\beta\lambda^2\mu^2 \boldsymbol{A}_\Omega[|u|^2](\boldsymbol{x}/\mu) u(\boldsymbol{x}/\mu)\right|^2 \mathrm{d}\boldsymbol{x} \\ &= \lambda^2 \int_\Omega \left|\nabla u(z) + i\beta\lambda^2\mu^2 \boldsymbol{A}_\Omega[|u|^2](z) u(z)\right|^2 \mathrm{d}z \;= \lambda^2 \mathcal{E}^{\mathrm{af}}_{\Omega,\beta\lambda^2\mu^2}[u].\end{aligned}$$

Hence, we may take as a trial state for $\mathcal{E}^{\mathrm{af}}_{\mu\Omega,\beta\lambda^{-2}\mu^{-2}}$ the function $u_{\lambda,\mu}$, where u is the minimizer (or minimizing sequence) of $\mathcal{E}^{\mathrm{af}}_{\Omega,\beta}$, and vice versa. Moreover, if $u \in H^1_0$ then so is $u_{\lambda,\mu}$. □

It follows immediately from the above that the thermodynamic energy has a very simple dependence on its parameters, which justifies (2-6) and the way it appears in (2-9).

Corollary 3.5 (Scaling laws for $e(\beta, \rho)$).
For any $\rho \geqslant 0$ and bounded $\Omega \subset \mathbb{R}^2$, with $e(\beta, \rho)$ defined as in (3-1), *we have*

$$e(1, \rho) = |\Omega| \liminf_{\beta\to\infty} \frac{E_0(\Omega, \beta, \rho)}{\beta}, \tag{3-11}$$

and for any $\beta, \rho \geqslant 0$,

$$e(\beta, \rho) = \beta\rho^2 e(1, 1). \tag{3-12}$$

Remark 3.6. At the moment each shape of the domain Ω may give rise to a different limit $e(\beta, \rho)$ in (3-1), and this corollary and proof apply in such a situation. However, it will be shown below in the case of Lipschitz regular domains that the limit is independent of the shape, and one may therefore without loss of generality take the unit square $\Omega = Q$ as a reference domain.

Proof. A first consequence of the scaling property (3-9) is that taking the thermodynamic limit as described in (2-5) or (3-1) is equivalent to taking the limit $\beta \to \infty$ at a fixed size of the domain, i.e.,

$$e(c, \rho) = \liminf_{L\to\infty} \frac{E_0(L\Omega, c, \rho|\Omega|L^2)}{L^2|\Omega|} = \liminf_{L\to\infty} \frac{E_0(\Omega, cL^2|\Omega|, \rho)}{L^2},$$

where we have applied (3-9) with $\mu = L$, $\lambda = |\Omega|^{1/2}$ and $M = \rho$. Now if, for any $c > 0$, we set $\beta = cL^2|\Omega| \to \infty$, the above expression becomes

$$e(c, \rho) = c|\Omega| \liminf_{\beta\to\infty} \frac{E_0(\Omega, \beta, \rho)}{\beta}, \tag{3-13}$$

which proves the first claim, and also implies

$$e(c, \rho) = c\, e(1, \rho). \tag{3-14}$$

Next we take $\mu = 1$ in (3-9) and obtain

$$E_0(\Omega, \beta, M) = \lambda^{-2} E_0(\Omega, \beta\lambda^{-2}, \lambda^2 M).$$

Taking $M = |\Omega|$, dividing by $|\Omega|$ and taking the limit $|\Omega| \to \infty$, we deduce

$$e(\beta, 1) = \lambda^{-2} e(\beta\lambda^{-2}, \lambda^2) = \lambda^{-4} e(\beta, \lambda^2),$$

where we used (3-14) in the last equality. This yields

$$e(\beta, \rho) = \rho^2 e(\beta, 1) \tag{3-15}$$

for all $\beta, \rho \geqslant 0$. Combining (3-14) and (3-15) yields the result (3-12). □

3C. ***Proof of Theorem 2.1.*** We split the proof in three lemmas:

Lemma 3.7 (Thermodynamic limit for the Dirichlet energy in a square).
Let Q be a unit square, and $\rho \geqslant 0$ and $\beta \geqslant 0$ be fixed constants. The limit

$$e(\beta, \rho) = \lim_{L\to+\infty} \frac{E_0(LQ, \beta, \rho L^2)}{L^2}$$

exists, is finite, and satisfies $e(\beta, \rho) \geqslant 2\pi\beta\rho^2$.

Lemma 3.8 (Neumann–Dirichlet comparison).
Let Ω be a bounded simply connected domain with Lipschitz boundary. Then for any fixed ρ and β positive, as $L \to \infty$

$$\frac{E_0(L\Omega, \beta, \rho L^2|\Omega|)}{L^2|\Omega|} \geqslant \frac{E(L\Omega, \beta, \rho L^2|\Omega|)}{L^2|\Omega|} \geqslant \frac{E_0(L\Omega, \beta, \rho L^2|\Omega|)}{L^2|\Omega|} - o(1).$$

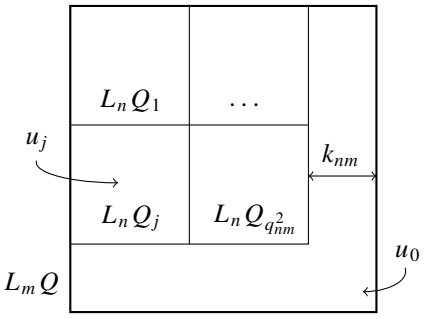

Figure 1. Filling the square L_mQ with smaller squares L_nQ_j.

Lemma 3.9 (Thermodynamic limit for the Dirichlet energy in a general domain). *Let $\Omega \subset \mathbb{R}^2$ be a bounded simply connected domain with Lipschitz boundary, then*

$$\lim_{L\to+\infty} \frac{E_0(L\Omega, \beta, \rho L^2|\Omega|)}{L^2|\Omega|} = e(\beta, \rho). \tag{3-16}$$

Theorem 2.1 immediately follows from these three results: combining Lemma 3.7 with Lemma 3.8 one obtains the existence of the thermodynamic limit for squares. In order to derive the result for general domains, one then uses Lemma 3.9 together with Lemma 3.8. Notice that the proof of Lemma 3.9 requires only Lemmas 3.7 and 3.8 for squares as key ingredients.

Proof of Lemma 3.7. From Lemma 3.1 we know that the sequence of energies per unit area has both an upper and lower limit. We denote by $(L_n)_{n\in\mathbb{N}}$ and $(L_m)_{m\in\mathbb{N}}$ two increasing sequences of positive real numbers such that $L_n \to \infty$, $L_m \to \infty$ and

$$\frac{E_0(L_nQ, \beta, \rho L_n^2)}{L_n^2} \to \liminf_{L\to\infty} \frac{E_0(LQ, \beta, \rho L^2)}{L^2} \quad \text{as } n \to \infty,$$
$$\frac{E_0(L_mQ, \beta, \rho L_m^2)}{L_m^2} \to \limsup_{L\to\infty} \frac{E_0(LQ, \beta, \rho L^2)}{L^2} \quad \text{as } m \to \infty.$$

For each n, there must exist a sequence of integers

$$q_{nm} \to +\infty \quad \text{as } m \to \infty$$

such that, for m large enough, e.g., $m \gg n$,

$$L_m = q_{nm}L_n + k_{nm}, \quad 0 \leqslant k_{nm} < L_n.$$

We then build a trial state for $E_0(L_mQ, \beta, \rho L_m^2)$ as follows (see Figure 1). The square L_mQ must contain q_{nm}^2 disjoint squares of side length L_n that we denote by L_nQ_j, $j = 1, \ldots, q_{nm}^2$. Then we pick u_j a minimizer of $E_0(L_nQ_j, \beta, \rho L_n^2)$ and remark that by definition,

$$\sum_{k=1,k\neq j}^{q_{nm}^2} \operatorname{curl} \boldsymbol{A}[|u_k|^2] = 0 \quad \text{in } L_nQ_j.$$

Thus there exists a gauge phase ϕ_j on the simply connected domain $L_n Q_j$ such that

$$\sum_{k=1,k\neq j}^{q_{nm}^2} \boldsymbol{A}[|u_k|^2] = \nabla\phi_j \quad \text{in } L_n Q_j.$$

Similarly, there exists ϕ_0 on the remaining part of the domain (which can be arranged to be simply connected as well, as in Figure 1) such that

$$\sum_{k=1}^{q_{nm}^2} \boldsymbol{A}[|u_k|^2] = \nabla\phi_0 \quad \text{on } L_m Q \setminus \bigcup_{j=1}^{q_{nm}^2} L_n Q_j.$$

We define the trial state as (see the proof of Lemma 3.1)

$$u := \sum_{j=1}^{q_{nm}^2} u_j e^{-i\beta\phi_j} + u_0 e^{-i\beta\phi_0},$$

where u_0 is a function with compact support in $L_m Q \setminus \bigcup_{j=1}^{q_{nm}^2} L_n Q_j$ satisfying

$$\int_{L_m Q} |u_0|^2 = \rho L_m^2 - q_{nm}^2 \rho L_n^2.$$

By Lemma 3.1, we can construct u_0 such that

$$\int_{L_m Q} \left|\left(\nabla + i\beta \boldsymbol{A}[|u_0|^2]\right)u_0\right|^2 \leqslant C(L_m^2 - q_{nm}^2 L_n^2) \leqslant 2C L_m k_{nm}$$

(where $C > 0$ may depend on β and ρ). The function u is an admissible trial state on $L_m Q$ because it is in H^1 on each subdomain, and continuous across boundaries due to the Dirichlet boundary conditions satisfied by each u_j. Computing the energy, we have

$$\begin{aligned}\mathcal{E}^{\mathrm{af}}_{L_m Q,\beta}[u] &= \sum_{j=0}^{q_{nm}^2} \int_{L_m Q} \left|e^{-i\phi_j}\left(\nabla + i\beta \boldsymbol{A}[|u|^2] - i\beta\nabla\phi_j\right)u_j\right|^2 = \sum_{j=0}^{q_{nm}^2} \int_{L_m Q} \left|\left(\nabla + i\beta \boldsymbol{A}[|u_j|^2]\right)u_j\right|^2 \\ &= \sum_{j=1}^{q_{nm}^2} \mathcal{E}^{\mathrm{af}}_{L_n Q,\beta}[u_j] + \int_{L_m Q} \left|\left(\nabla + i\beta \boldsymbol{A}[|u_0|^2]\right)u_0\right|^2 = q_{nm}^2 E_0(L_n Q, \beta, \rho L_n^2) + O(L_m k_{nm}),\end{aligned}$$

with

$$q_{nm}^2 = \frac{L_m^2}{L_n^2}\left(1 - \frac{k_{nm}}{L_m}\right)^2.$$

Since u has by definition mass ρL_m^2, it follows from the variational principle that

$$\frac{E_0(L_m Q, \beta, \rho L_m^2)}{L_m^2} \leqslant \frac{E_0(L_n Q, \beta, \rho L_n^2)}{L_n^2}\left(1 + O\left(\frac{k_{nm}}{L_m}\right)\right) + O\left(\frac{k_{nm}}{L_m}\right).$$

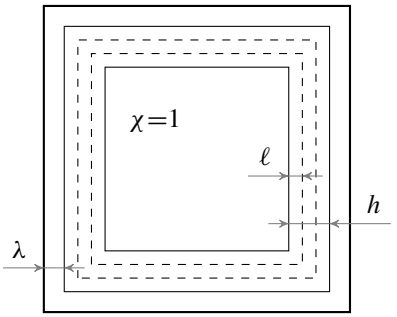

Figure 2. Localizing on thin shells for the square LQ.

Passing to the limit $m \to \infty$ first and then $n \to \infty$ yields

$$\limsup_{L\to\infty} \frac{E_0(LQ, \beta, \rho L^2)}{L^2} \leqslant \liminf_{L\to\infty} \frac{E_0(LQ, \beta, \rho L^2)}{L^2},$$

and thus the limit exists.

Additionally, we have by the bound (3-3),

$$\frac{1}{L^2}\mathcal{E}^{\mathrm{af}}_{LQ,\beta}[u] \geqslant \frac{2\pi\beta}{L^2}\int_{LQ}|u|^4 \geqslant \frac{2\pi\beta}{L^4}\left(\int_{LQ}|u|^2\right)^2$$

for any $u \in H^1_0(LQ)$, proving that $e(\beta, \rho) \geqslant 2\pi\beta\rho^2$. □

Proof of Lemma 3.8. Since $H^1_0(\Omega) \subseteq H^1(\Omega)$, we obviously have

$$E_0(\Omega, \beta, M) \geqslant E(\Omega, \beta, M).$$

Only the second inequality in the statement requires some work. Let $u \in H^1(L\Omega)$ denote the minimizer of $\mathcal{E}^{\mathrm{af}}_{L\Omega,\beta}[u]$ (see Proposition A.1 of the Appendix) with mass

$$\int_{L\Omega}|u|^2 = \rho L^2|\Omega|$$

and no further constraint (thus satisfying Neumann boundary conditions). In the sequel we take $\beta = 1$ and $|\Omega| = 1$ to simplify the notation.

We will need to make an IMS localization on a small enough region, and therefore consider a division of $L\Omega$ into a bulk region surrounded by thin shells close to the boundary, where we will be using several different length scales $L^{-1/3} \lesssim \lambda \ll 1 \ll L$ and $L^{-1} \ll \ell \ll h \ll L$ (see Figure 2 for the case of $\Omega = Q$ a square).

We shall use Lemma 3.3 a first time at distance λ from the boundary to deduce some useful a priori bounds. Next, using a mean-value argument we show that, within a window of thickness h further from the boundary, there must exist one particular shell of thickness ℓ where we have a good control on the mass and energy. Finally we perform a second IMS localization with the truncation located in this particular shell. This yields a lower bound in terms of the Dirichlet energy in the bulk region, plus error terms that

we can control using the a priori bounds and in particular the good control on mass and energy in the second localization shell.

Step 1: a priori bounds. Let $\delta_\Omega(\boldsymbol{x}) := \operatorname{dist}(\boldsymbol{x}, \partial(L\Omega))$ denote the distance function to the boundary, which is Lipschitz and satisfies $|\nabla\delta_\Omega| \leqslant 1$ a.e. We make a first partition of unity

$$\tilde{\chi}^2 + \tilde{\eta}^2 = 1$$

such that $\tilde{\chi}$ varies smoothly from 1 to 0 on a shell K_λ of width λ closest to the boundary of $L\Omega$, i.e., $K_\lambda := \{\boldsymbol{x} \in L\Omega : \delta_\Omega(\boldsymbol{x}) < \lambda\}$. One may note that it is possible to construct these functions so as to satisfy

$$|\nabla\tilde{\chi}| \leqslant c\lambda^{-1}\tilde{\chi}^{1-\mu}, \quad |\nabla\tilde{\eta}| \leqslant c\lambda^{-1}\tilde{\chi}^{1-\mu}$$

for some arbitrarily small $\mu > 0$, independent of λ, e.g., by taking $\tilde{\chi} = f^a$ and $\tilde{\eta} = \sqrt{1-\tilde{\chi}^2}$ in $\operatorname{supp}\tilde{\chi} \cap \operatorname{supp}\tilde{\eta}$ for a large and some smooth function $0 \leqslant f \leqslant 1$ varying on the right length scale and reflection symmetric. Then, by Lemmas 3.1 and 3.3,

$$\begin{aligned} CL^2 \geqslant \mathcal{E}^{\mathrm{af}}_{L\Omega,1}[u] &\geqslant \int_{L\Omega} \left(2\pi\,\tilde{\chi}^2|u|^4 + \left|\nabla|\tilde{\eta}u|\right|^2 - (|\nabla\tilde{\chi}|^2 + |\nabla\tilde{\eta}|^2)|u|^2\right) \\ &\geqslant \int_{L\Omega} \left(2\pi\,\tilde{\chi}^2|u|^4 + \left|\nabla|\tilde{\eta}u|\right|^2 - C\lambda^{-2}\mathbb{1}_{K_\lambda}\tilde{\chi}^{2-2\mu}|u|^2\right). \end{aligned} \tag{3-17}$$

We bound the unwanted negative term as follows:

$$\begin{aligned} \lambda^{-2}\int_{L\Omega} \mathbb{1}_{K_\lambda}\tilde{\chi}^{2-2\mu}|u|^2 &\leqslant \lambda^{-2}\left(\int_{K_\lambda}\tilde{\chi}^{2-4\mu}\right)^{1/2}\left(\int_{K_\lambda}\tilde{\chi}^2|u|^4\right)^{1/2} \\ &\leqslant C\lambda^{-3/2}L^{1/2}\left(\int_{K_\lambda}\tilde{\chi}^2|u|^4\right)^{1/2} \leqslant C\delta L\lambda^{-3} + C\delta^{-1}\int_{L\Omega}\tilde{\chi}^2|u|^4, \end{aligned}$$

with δ a fixed, large enough, constant. Combining with (3-17) we deduce

$$\int_{L\Omega}\left(2\pi\,\tilde{\chi}^2|u|^4 + \left|\nabla|\tilde{\eta}u|\right|^2\right) \leqslant CL^2 + CL\lambda^{-3} \leqslant CL^2 \tag{3-18}$$

since we have chosen $\lambda \gtrsim L^{-1/3}$. We note that this bound implies for the mass in a shell K_ℓ of thickness ℓ in $L\Omega \setminus K_\lambda$

$$\int_{K_\ell}|u|^2 \leqslant |K_\ell|^{1/2}\left(\int_{K_\ell}\tilde{\chi}^2|u|^4\right)^{1/2} \lesssim \ell^{1/2}L^{3/2}. \tag{3-19}$$

Step 2: finding a good shell. We now select a region where the bounds (3-18) and (3-19) can be improved. Consider dividing $L\Omega \setminus K_\lambda$ into shells of thickness ℓ that form a layer closest to the shell K_λ of total thickness $h \sim L^{1-\varepsilon} \gg \ell$ (again, see Figure 2). Hence, we have

$$N_s := h/\ell \gg 1$$

such shells in the layer. Denote by N_M the number of such shells K_ℓ with $\int_{K_\ell}|u|^4 \geqslant M$. If $N_M < N_s$, there must exist a shell K_ℓ with $\int_{K_\ell}|u|^4 \leqslant M$. But, using (3-18) and the fact that all the shells are included

in the region where $\tilde{\chi} = 1$, we have

$$MN_M \leqslant \int_{L\Omega} \tilde{\chi}^2 |u|^4 \leqslant CL^2.$$

We can thus ensure that $N_M < N_s$ by setting

$$N_s = h/\ell \sim L^{1-\varepsilon}\ell^{-1} \sim L^2/M,$$

i.e., taking $M \sim \ell L^{1+\varepsilon}$. Hence we have found a shell K_ℓ with

$$\int_{K_\ell} |u|^4 \leqslant C\ell L^{1+\varepsilon}, \tag{3-20}$$

and thus

$$\int_{K_\ell} |u|^2 \leqslant C(\ell L)^{1/2}(\ell L^{1+\varepsilon})^{1/2} = C\ell L^{1+\varepsilon/2}, \tag{3-21}$$

improving (3-19).

Step 3: IMS localization in the good shell. We now perform a new magnetic localization on this K_ℓ. We pick a partition $\chi^2 + \eta^2 = 1$, such that χ varies smoothly from 1 to 0 outwards on K_ℓ, so that $\chi = 1$ (resp. $\eta = 1$) on the inner (resp. outer) component of K_ℓ^c. Then, using Lemma 3.3, we have

$$\mathcal{E}^{\mathrm{af}}_{L\Omega,1}[u] \geqslant (1-\delta)\mathcal{E}^{\mathrm{af}}_{L\Omega,1}[\psi] - (\delta^{-1}-1)\int_{L\Omega} \left|\boldsymbol{A}[|\eta u|^2 \mathbb{1}_K]\right|^2 |\chi u|^2 - \int_{K_\ell} (|\nabla\chi|^2 + |\nabla\eta|^2)|u|^2 \tag{3-22}$$

for any $\delta \in (0,1)$, where we let $\psi = \chi e^{i\phi} u$ and $K = \operatorname{supp}\chi \cap \operatorname{supp}\eta \subseteq K_\ell$. Since ψ is compactly supported in $L\Omega$, we have for the first term

$$\mathcal{E}^{\mathrm{af}}_{L\Omega,1}[\psi] \geqslant E_0(L\Omega, 1, \|\psi\|^2_{L^2(L\Omega)}) = E_0(L\Omega, 1, \|\chi u\|^2_{L^2(L\Omega)}).$$

Recalling the scaling relation (3-9) (taking $\mu = \lambda^{-1} = \widetilde{L}/L$) and defining

$$M = \int_{L\Omega} \chi^2 |u|^2, \quad \widetilde{L} = \sqrt{M/\rho},$$

we have

$$E_0(L\Omega, 1, M) = \frac{M}{\rho L^2} E_0(\widetilde{L}\Omega, 1, \rho\widetilde{L}^2). \tag{3-23}$$

We need to estimate the deviation of the mass M of $\chi^2|u|^2$ from $\rho L^2 = \int_{L\Omega} |u|^2$:

$$\begin{aligned}\left|\rho L^2 - \int_{L\Omega} \chi^2|u|^2\right| = \int_{L\Omega} \eta^2|u|^2 &= \int_{L\Omega} \tilde{\eta}^2|u|^2 + \int_{L\Omega} \tilde{\chi}^2\eta^2|u|^2 \\ &\leqslant C\lambda^2 \int_{K_\lambda} \left|\nabla|\tilde{\eta}u|\right|^2 + \left(\int_{L\Omega} \eta^2\tilde{\chi}^2\right)^{1/2}\left(\int_{L\Omega} \tilde{\chi}^2|u|^4\right)^{1/2} \\ &\leqslant C\lambda^2 L^2 + Ch^{1/2}L^{3/2} \ll L^2. \end{aligned} \tag{3-24}$$

Here we have used a Poincaré inequality to control the $\tilde{\eta}^2|u|^2$ term, making use of the fact that this function vanishes at the inner boundary of K_λ. It is not difficult (see the proof methods of [Evans 1998, Theorems 1 and 2 in Section 5.8.1] and [Lieb and Loss 2001, Theorem 8.11]) to realize that the constant

involved in this inequality applied on the set K_λ can be taken to be proportional to λ^2. Note that $\widetilde{L} \to \infty$, if $L \to \infty$, thanks to (3-24). Hence, inserting the above estimate in (3-23), we get

$$\frac{\mathcal{E}^{\mathrm{af}}_{L\Omega,1}[\psi]}{L^2} \geqslant \frac{E_0(L\Omega,1,M)}{L^2} = \frac{M^2}{(\rho L^2)^2}\frac{E_0(\widetilde{L}\Omega,1,\rho\widetilde{L}^2)}{\widetilde{L}^2} = (1+o(1))\frac{E_0(\widetilde{L}\Omega,1,\rho\widetilde{L}^2)}{\widetilde{L}^2}. \tag{3-25}$$

Then, there only remains to control the error terms in (3-22): Using the Hölder and generalized Young inequalities ($\|\cdot\|_{p,w}$ denotes the weak-L^p norm [Lieb and Loss 2001, Theorem 4.3, Remarks]),

$$\begin{aligned}\int_{L\Omega} \left|\boldsymbol{A}[|\eta u|^2 \mathbb{1}_K]\right|^2 |\chi u|^2 &\leqslant \left\|\nabla w_0 * |\eta u|^2 \mathbb{1}_K\right\|_{2p}^2 \|\chi u\|_{2q}^2 \leqslant c\|\nabla w_0\|_{2,w}^2 \|\eta u \mathbb{1}_K\|_{2r}^4 \|\chi u\|_{2q}^2 \\ &\leqslant C\left(\int_{K_\ell} |\eta u|^{4q/(2q-1)}\right)^{(2q-1)/q}\left(\int_{L\Omega}|\chi u|^{2q}\right)^{1/q},\end{aligned}$$

where

$$\frac{1}{p}+\frac{1}{q}=1 \quad \text{and} \quad 1+\frac{1}{2p}=\frac{1}{2}+\frac{1}{r},$$

that is,

$$r=\frac{2q}{2q-1}\in(1,2) \quad \text{with } q\in(1,\infty).$$

We can take $q=2$ and insert (3-18)–(3-20) to obtain

$$\left(\int_{K_\ell}|\eta u|^{8/3}\right)^{3/2}\left(\int_{L\Omega}|\chi u|^4\right)^{1/2} \leqslant |K_\ell|^{1/2}\int_{K_\ell}|\eta u|^4\left(\int_{L\Omega}|\chi u|^4\right)^{1/2} \lesssim (\ell L)^{1/2}\ell L^{1+\varepsilon}(L^2)^{1/2} = \ell^{3/2}L^{5/2+\varepsilon}.$$

The last term in (3-22) is, using (3-21), bounded by

$$c\ell^{-2}\int_{K_\ell}|u|^2 \lesssim \ell^{-1}L^{1+\varepsilon/2}.$$

There only remains to optimize the error terms in (3-22):

$$\delta E_0(L\Omega,1,\|\psi\|^2_{L^2(L\Omega)}) + c_1(\delta^{-1}-1)\ell^{3/2}L^{5/2+\varepsilon} + c_2\ell^{-1}L^{1+\varepsilon/2} \leqslant c_3\delta L^2 + c_4\delta^{-2/5}L^{8/5+7\varepsilon/10},$$

where we have picked $\ell = L^{-3/5-\varepsilon/5}\delta^{2/5}$, assuming that $\delta \ll 1$, as it will be. Thus, optimizing now over δ, i.e., taking $\delta \sim L^{-2/7+\varepsilon/2}$, we have the bounds

$$\frac{E_0(L\Omega,1,\rho L^2)}{L^2} \geqslant \frac{E(L\Omega,1,\rho L^2)}{L^2} \geqslant \frac{E_0(L\Omega,1,\|\psi\|^2_{L^2(L\Omega)})}{L^2} - cL^{-2/7+\varepsilon/2}. \tag{3-26}$$

Combining with (3-25) and passing to the liminf completes the proof. □

Proof of Lemma 3.9. The result is proven as usual by comparing suitable upper and lower bounds to the energy.

Step 1: upper bound. We first cover $L\Omega$ with squares Q_j, $j=1,\dots,N_\ell$, of side length $\ell = L^\eta$, $0<\eta<1$, retaining only the squares Q_j completely contained in $L\Omega$. One can estimate the area not covered by

such squares as

$$\left|\Omega \setminus \left(\bigcup_{j=1}^{N_\ell} Q_j\right)\right| \leqslant C\ell L = o(L^2). \tag{3-27}$$

Then we define the trial state

$$u(\boldsymbol{x}) := \sum_{j=1}^{N_\ell} u_j e^{-i\beta\phi_j}, \tag{3-28}$$

where

$$u_j(\boldsymbol{x}) := u_0(\boldsymbol{x} - \boldsymbol{x}_j)\mathbb{1}_{Q_j}, \tag{3-29}$$

with u_0 a minimizer of the Dirichlet problem with mass $\rho L^2|\Omega|/N_\ell$ in a square Q with side length ℓ centered at the origin, and $\boldsymbol{x}_j$ the center point of Q_j. The phases ϕ_j are chosen in such a way that (see the proof of Lemma 3.1 again)

$$\sum_{k=1,k\neq j}^{N_\ell} \boldsymbol{A}[|u_k|^2] = \nabla\phi_j \quad \text{in } Q_j.$$

The existence of such phases is indeed guaranteed by the fact that

$$\sum_{k=1,k\neq j}^{N_\ell} \operatorname{curl} \boldsymbol{A}[|u_k|^2] = 0 \quad \text{in } Q_j.$$

Hence

$$\mathcal{E}^{\mathrm{af}}_{L\Omega,\beta}[u] = \sum_{j=1}^{N_\ell} \mathcal{E}^{\mathrm{af}}_{Q_j,\beta}[u_j] = \sum_{j=1}^{N_\ell} E_0(\ell Q, \beta, \rho L^2|\Omega|N_\ell^{-1}),$$

which implies

$$\begin{aligned}\frac{E_0(L\Omega,\beta,\rho L^2)}{L^2|\Omega|} &\leqslant \frac{1}{L^2|\Omega|}\sum_{j=1}^{N_\ell} E_0(\ell Q,\beta,\rho L^2|\Omega|N_\ell^{-1}) \\ &= \frac{\ell^2}{L^2|\Omega|}\sum_{j=1}^{N_\ell} E_0(\ell Q,\beta,(1+o(1))\rho\ell^2)/\ell^2 = (1+o(1))e(\beta,\rho),\end{aligned} \tag{3-30}$$

where we have estimated

$$N_\ell = \frac{\left|\bigcup_j Q_j\right|}{|Q_j|} = \frac{(1+o(1))L^2|\Omega|}{\ell^2}, \tag{3-31}$$

and used Lemma 3.7. Notice that, thanks to the assumption on η, we have $\ell \to \infty$, which is crucial in order to apply Lemma 3.7.

Step 2: lower bound. We again cover $L\Omega$ with squares Q_j, $j = 1, \ldots, N_\ell$, this time keeping the full covering but still having $\ell^2 N_\ell/|L\Omega| \to 1$ as $L \to \infty$. We pick a minimizer $u^{\mathrm{af}} = u^{\mathrm{af}}_L \in H^1_0(L\Omega)$ of $\mathcal{E}^{\mathrm{af}}_{L\Omega,\beta}$, with mass $\rho L^2|\Omega|$, and set

$$u^{\mathrm{af}}_j := u^{\mathrm{af}}\mathbb{1}_{Q_j}, \quad \rho_j := \fint_{Q_j} |u^{\mathrm{af}}(\boldsymbol{x})|^2 \,\mathrm{d}\boldsymbol{x}. \tag{3-32}$$

The idea of the proof is reminiscent of that in the upper bound part: we gauge away the magnetic interaction between the cells, and this leads to a lower bound in terms of the Neumann energy of the cells.

Note that $u_j^{\mathrm{af}} \in H^1(Q_j)$ for each j, and

$$\sum_{j=1}^{N_\ell} \rho_j \ell^2 = \rho L^2 |\Omega|.$$

Before estimating the energy, we need to distinguish between squares with sufficient mass and squares which will not contribute to the energy to leading order. We thus set

$$\mathcal{Q}_L := \left\{ Q_j,\, j \in \{1, \dots, N_\ell\} \,:\, \rho_j \geqslant L^{-2\eta+\delta} \right\} \tag{3-33}$$

for some $0 < \delta < 2\eta$. Note that the mass concentrated outside cells $\mathcal{Q}_L$ is relatively small:

$$\sum_{Q_j \notin \mathcal{Q}_L} \rho_j \ell^2 \leqslant C \ell^2 N_\ell L^{-2\eta+\delta} = o(L^2). \tag{3-34}$$

We can now estimate, using the gauge covariance of the functional on each Q_j,

$$\begin{aligned} E_0(L\Omega, \beta, \rho L^2|\Omega|) = \mathcal{E}^{\mathrm{af}}_{L\Omega,\beta}[u^{\mathrm{af}}] &\geqslant \sum_{j=1}^{N_\ell} \int_{Q_j} \left| \left(-i\nabla + \beta \boldsymbol{A}[|u^{\mathrm{af}}|^2]\right) u^{\mathrm{af}} \right|^2 \\ &= \sum_{j=1}^{N_\ell} \int_{Q_j} \left| \left(-i\nabla + \beta \boldsymbol{A}[|u_j^{\mathrm{af}} e^{i\beta\phi_j}|^2]\right) u_j^{\mathrm{af}} e^{i\beta\phi_j} \right|^2 \\ &\geqslant \sum_{j=1}^{N_\ell} \rho_j \ell^2 \frac{E(\ell Q, \beta, \rho_j \ell^2)}{\rho_j \ell^2} \geqslant \sum_{j : Q_j \in \mathcal{Q}_L} \rho_j^2 \ell^2 \frac{E(\ell_j Q, \beta, \ell_j^2)}{\ell_j^2}, \end{aligned} \tag{3-35}$$

where ϕ_j satisfies (observe that the left-hand side is curl-free on Q_j)

$$\sum_{k=1, k\neq j}^{N_\ell} \boldsymbol{A}[|u_k^{\mathrm{af}}|^2] = \nabla \phi_j \quad \text{in } Q_j,$$

and in the last step we used the scaling law (3-9) with $\mu = 1/\lambda = \sqrt{\rho_j}$. Also,

$$\ell_j := \sqrt{\rho_j}\ell \geqslant L^{\delta/2} \to +\infty \quad \text{as } L \to \infty$$

uniformly in j for cells $Q_j \in \mathcal{Q}_L$, and we thus conclude by Lemmas 3.7 and 3.8 that

$$\frac{1}{L^2|\Omega|} E_0(L\Omega, \beta, \rho L^2|\Omega|) \geqslant (1 - o(1)) \frac{e(\beta, 1)}{L^2|\Omega|} \sum_{j : Q_j \in \mathcal{Q}_L} \rho_j^2 \ell^2 = (1 - o(1)) \frac{e(\beta, 1)}{L^2|\Omega|} \int_{\mathcal{Q}} \bar{\varrho}^2, \tag{3-36}$$

where we consider here the step function $\bar{\varrho} := \sum_{j : Q_j \in \mathcal{Q}_L} \rho_j \mathbb{1}_{Q_j}$ and denote by $\mathcal{Q}$ the union of the cells $\mathcal{Q}_L$. It remains then to observe that the constrained minimum

$$B = \min \left\{ \int_{\mathcal{Q}} \varrho^2 \,:\, 0 \leqslant \varrho \in L^2(\mathcal{Q}),\ \int_{\mathcal{Q}} \varrho = (1 - o(1)) \rho L^2 |\Omega| \right\}$$

is achieved by ϱ constant and thus

$$\int_{\mathcal{Q}} \bar{\varrho}^2 \geqslant B = \big((1-o(1))\rho L^2|\Omega|\big)^2 |\mathcal{Q}|^{-1} \geqslant (1-o(1))\rho^2 L^2 |\Omega|.$$

Inserting this in (3-36) and using $\rho^2 e(\beta, 1) = e(\beta, \rho)$ leads to the desired energy lower bound. □

4. Proofs for the trapped gas

4A. ***Local density approximation: energy upper bound.*** Here we prove the upper bound corresponding to (2-19):

$$E_\beta^{\mathrm{af}} \leqslant E_\beta^{\mathrm{TF}}(1+o(1)) \quad \text{as } \beta \to \infty. \tag{4-1}$$

We start by covering the support of $\varrho_\beta^{\mathrm{TF}}$ with squares Q_j, $j = 1, \dots, N_\beta$, centered at points $\boldsymbol{x}_j$ and of side length L with

$$L = \beta^\eta, \quad -\frac{s}{2(s+2)} < \eta < \frac{1}{s+2}. \tag{4-2}$$

We choose the tiling in such a way that for any $j = 1, \dots, N_\beta$, we have $Q_j \cap \operatorname{supp}(\varrho_\beta^{\mathrm{TF}}) \neq \varnothing$. The upper bound on L indicates that the length scale of the tiling is much smaller than the size of the TF support. The lower bound ensures that it is much larger than the scale on which we expect the fine structure of the minimizer to live.

Our trial state is defined much as in the proof of Lemma 3.9:

$$u^{\mathrm{test}} := \sum_{j=1}^{N_\beta} u_j e^{-i\beta\phi_j}, \tag{4-3}$$

where u_j realizes the Dirichlet infimum

$$E_0(Q_j, \beta, M_j) := \min\big\{\mathcal{E}_j^{\mathrm{af}}[u] : u \in H_0^1(Q_j),\ \textstyle\int_{Q_j} |u|^2 = M_j\big\},$$

where of course

$$\mathcal{E}_j^{\mathrm{af}}[u] = \mathcal{E}_{Q_j,\beta}^{\mathrm{af}}[u] = \int_{Q_j} \big|\big(-i\nabla + \beta \boldsymbol{A}[|u|^2]\big)u\big|^2$$

and we set

$$M_j = \int_{Q_j} |u_j|^2 := \int_{Q_j} \varrho_\beta^{\mathrm{TF}}, \quad \rho_j := M_j/L^2 = \fint_{Q_j} \varrho_\beta^{\mathrm{TF}}. \tag{4-4}$$

The phase factors in (4-3) are again defined so as to gauge away the interaction between cells, i.e.,

$$\sum_{k=1, k\neq j}^{N_\beta} \boldsymbol{A}[|u_k|^2] = \nabla\phi_j \quad \text{in } Q_j.$$

This construction yields an admissible trial state since u^{test} is locally in H^1, continuous across cells by being zero on the boundaries, and clearly

$$\int_{\mathbb{R}^2} |u^{\mathrm{test}}|^2 = \sum_{j=1}^{N_\beta} \int_{Q_j} |u_j|^2 = \sum_{j=1}^{N_\beta} \int_{Q_j} \varrho_\beta^{\mathrm{TF}} = 1.$$

Much as in the proofs of Lemmas 3.1 and 3.9 we thus obtain

$$E_\beta^{\mathrm{af}} \leqslant \mathcal{E}_\beta^{\mathrm{af}}[u^{\mathrm{test}}] = \sum_{j=1}^{N_\beta} \mathcal{E}_j^{\mathrm{af}}[u_j] + \int_{\mathbb{R}^2} V|u^{\mathrm{test}}|^2 = \sum_{j=1}^{N_\beta} E_0(Q_j, \beta, M_j) + \int_{\mathbb{R}^2} V|u^{\mathrm{test}}|^2. \tag{4-5}$$

Our task is then to estimate the right-hand side.

We denote, for some $\varepsilon > 0$ small enough

$$S_\varepsilon = \left\{ \boldsymbol{x} \in \mathrm{supp}(\varrho_\beta^{\mathrm{TF}}) : \varrho_\beta^{\mathrm{TF}}(\boldsymbol{x}) \geqslant \beta^{-2/(s+2)-\varepsilon} \right\}$$

and split the above sum into two parts, distinguishing between cells fully included in S_ε and the others. Using (2-13), it is clear that

$$\left| \mathrm{supp}(\varrho_\beta^{\mathrm{TF}}) \setminus S_\varepsilon \right| \leqslant C\beta^{1/(s+2)} \cdot \beta^{1/(s+2)-\varepsilon},$$

where the first factor comes from the dilation transforming ϱ_1^{TF} into $\varrho_\beta^{\mathrm{TF}}$ and the second one is an estimate of the thickness of S_ε based on (2-16)–(2-18).

By a simple estimate of the potential V in the vicinity of S_ε, we obtain

$$\sum_{j : Q_j \not\subseteq S_\varepsilon} \int_{Q_j} V|u_j|^2 \leqslant C\beta^{s/(s+2)} \cdot \beta^{2/(s+2)-\varepsilon} \cdot \beta^{-2/(s+2)-\varepsilon} = C\beta^{s/(s+2)-2\varepsilon} \ll E_\beta^{\mathrm{TF}},$$

where the factor $\beta^{s/(s+2)}$ accounts for the supremum of V, the factor $\beta^{2/(s+2)-\varepsilon}$ for the volume of the integration domain and the factor $\beta^{-2/(s+2)-\varepsilon}$ for the typical value of $|u_j|^2$ on this domain. Also, using in addition Lemmas 3.4 and 3.1, we deduce

$$\sum_{j : Q_j \not\subseteq S_\varepsilon} E_0(Q_j, \beta, M_j) = \sum_{j : Q_j \not\subseteq S_\varepsilon} E_0(\beta^\eta Q, \beta, \beta^{2\eta}\rho_j) \ll E_\beta^{\mathrm{TF}}.$$

For the main part of the sum in (4-5) we use the scaling law (take $\lambda = \sqrt{\rho_j}$ and $\mu = \sqrt{\beta\rho_j}$ in Lemma 3.4) to write

$$E_0(Q_j, \beta, M_j) = \rho_j E_0(L\sqrt{\beta\rho_j}\, Q, 1, L^2\beta\rho_j),$$

with Q the unit square. Then

$$\sum_{j : Q_j \subseteq S_\varepsilon} E_0(Q_j, \beta, M_j) = \sum_{j : Q_j \subseteq S_\varepsilon} L^2\beta\rho_j^2 e(1,1) + \sum_{j : Q_j \subseteq S_\varepsilon} L^2\beta\rho_j^2 \left(\frac{E_0(L_j Q, 1, L_j^2)}{L_j^2} - e(1,1) \right)$$

with, provided ε is suitably small and in view of the lower bound in (4-2) and the fact that we sum over squares included in S_ε,

$$L_j := L\sqrt{\beta\rho_j} \geqslant \beta^{\eta + s/(2(s+2)) - \varepsilon/2} \to +\infty, \quad \text{uniformly with respect to } j = 1, \dots, N_\beta.$$

We thus obtain (recall the definition of the thermodynamic energy in (2-5))

$$\frac{E_0(L_j Q, 1, L_j^2)}{L_j^2} \to e(1,1) \quad \text{as } L_j \to \infty$$

uniformly in j, and deduce that

$$\sum_{j:Q_j \subset S_\varepsilon} E_0(Q_j, \beta, M_j) = (1+o(1))\beta e(1,1) \sum_{j:Q_j \subset S_\varepsilon} \rho_j^2 L^2.$$

Recalling that

$$\rho_j = \fint_{Q_j} \varrho_\beta^{\mathrm{TF}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x},$$

we recognize a Riemann sum in the above. Using (2-17) and the upper bound in (4-2) we may approximate $\varrho_\beta^{\mathrm{TF}}$ by a constant in each square (this is most easily seen by rescaling to ϱ_1^{TF} and observing that the size of squares then tends to zero), and bound the part of the integral located in the complement of S_ε in the same way as above to conclude that

$$\sum_{j:Q_j \subset S_\varepsilon} E_0(Q_j, \beta, M_j) = (1+o(1))\beta e(1,1) \int_{\mathbb{R}^2} (\varrho_\beta^{\mathrm{TF}})^2.$$

Using (2-11) and (2-16) we obtain

$$|\nabla V(\boldsymbol{x})| \leqslant C\beta^{(s-1)/(s+2)}$$

for any $\boldsymbol{x} \in S_\varepsilon$. Combining with (4-2) we deduce as above that

$$\sum_{j:Q_j \subset S_\varepsilon} \int_{Q_j} V|u_j|^2 = (1+o(1)) \int_{\mathbb{R}^2} V \varrho_\beta^{\mathrm{TF}}$$

and this completes the proof of (4-1).

4B. ***Local density approximation: energy lower bound.*** Let us now complement (4-1) by proving the lower bound

$$E_\beta^{\mathrm{af}} \geqslant E_\beta^{\mathrm{TF}}(1+o(1)), \tag{4-6}$$

thus completing the proof of (2-19). We again tile the plane with squares Q_j, $j = 1, \dots, N_\beta$, of side length

$$L = \beta^\eta$$

satisfying (4-2), and taken to cover the finite disk $B_{\beta^t}(0)$ with

$$t := \frac{1}{2+s} + \varepsilon$$

for some $\varepsilon > 0$ to be chosen small enough. We also define

$$\mathcal{Q}_\beta := \left\{ Q_j \subset B_{\beta^t}(0) : L\sqrt{\rho_j \beta} \geqslant \beta^\mu \right\}, \tag{4-7}$$

where $u^{\mathrm{af}} = u^{\mathrm{af}}_\beta$ is a minimizer for $\mathcal{E}^{\mathrm{af}}_\beta$ with unit mass and

$$\rho_j := \fint_{Q_j} |u^{\mathrm{af}}(\boldsymbol{x})|^2 \, \mathrm{d}\boldsymbol{x}.$$

Define the piecewise constant function

$$\bar{\varrho}^{\mathrm{af}}(\boldsymbol{x}) := \sum_{Q_j \in \mathcal{Q}_\beta} \rho_j \mathbb{1}_{Q_j}(\boldsymbol{x}). \tag{4-8}$$

We claim that one may find some $\mu > 0$ in (4-7) such that

$$M := \int_{\mathbb{R}^2} \bar{\varrho}^{\mathrm{af}} \to 1 \quad \text{as } \beta \to \infty. \tag{4-9}$$

Indeed, using (2-11) and (2-12) we get that for any $\boldsymbol{x} \in B^c_{\beta^t}(0)$

$$V(\boldsymbol{x}) \geqslant C\beta^{st} \min_{B^c_1(0)} V \geqslant C\beta^{st}$$

for β large enough. Thus, using the energy upper bound (4-1) and dropping some positive terms we obtain

$$\beta^{st} \int_{B^c_{\beta^t}(0)} |u^{\mathrm{af}}|^2 \leqslant \int_{\mathbb{R}^2} V |u^{\mathrm{af}}|^2 \leqslant \mathcal{E}^{\mathrm{af}}_\beta[u^{\mathrm{af}}] \leqslant C\beta^{s/(s+2)}$$

and thus

$$\int_{B^c_{\beta^t}(0)} |u^{\mathrm{af}}|^2 \leqslant C\beta^{-s\varepsilon}. \tag{4-10}$$

On the other hand, by the definition of $\mathcal{Q}_\beta$,

$$\sum_{Q_j \notin \mathcal{Q}_\beta} \int_{Q_j} |u^{\mathrm{af}}|^2 \leqslant N_\beta \beta^{2\mu - 1},$$

where N_β is the total number of squares needed to tile $B_{\beta^t}(0)$. Clearly, we may estimate $N_\beta \leqslant C\beta^{2t} L^{-2} = C\beta^{2(t-\eta)}$ and then

$$\sum_{Q_j \notin \mathcal{Q}_\beta} \int_{Q_j} |u^{\mathrm{af}}|^2 \leqslant C\beta^{2t - 2\eta + 2\mu - 1} \ll 1 \tag{4-11}$$

because of (4-2), which implies $-s/(s+2) - 2\eta < 0$, and provided we take ε and μ positive and small enough, e.g. (recall that $L = \beta^\eta$ is the side length of the tiling squares),

$$0 < \varepsilon \leqslant \frac{1}{4}\left(\frac{s}{s+2} + 2\eta\right), \quad 0 < \mu \leqslant \varepsilon. \tag{4-12}$$

Combining (4-10) and (4-11) and recalling that u^{af} is L^2-normalized proves (4-9).

With this in hand we turn to the energy lower bound per se. Let us again set

$$u^{\mathrm{af}}_j = u^{\mathrm{af}} \mathbb{1}_{Q_j}, \quad M_j = \rho_j L^2 = \int_{Q_j} |u^{\mathrm{af}}|^2.$$

Dropping some positive terms we get

$$\begin{aligned} E_\beta^{\mathrm{af}} = \mathcal{E}_\beta^{\mathrm{af}}[u^{\mathrm{af}}] &\geqslant \sum_{Q_j\in\mathcal{Q}_\beta} \int_{Q_j} \left\{ \left| \left(-i\nabla + \beta \boldsymbol{A}[|u^{\mathrm{af}}|^2]\right) u^{\mathrm{af}} \right|^2 + V|u^{\mathrm{af}}|^2 \right\} \\ &= \sum_{Q_j\in\mathcal{Q}_\beta} \int_{Q_j} \left\{ \left| \left(-i\nabla + \beta \boldsymbol{A}[|u_j^{\mathrm{af}} e^{i\beta\phi_j}|^2]\right) u_j^{\mathrm{af}} e^{i\beta\phi_j} \right|^2 + V|u_j^{\mathrm{af}}|^2 \right\} \\ &\geqslant \sum_{Q_j\in\mathcal{Q}_\beta} \left\{ E(Q_j, \beta, M_j) + \int_{Q_j} V|u_j^{\mathrm{af}}|^2 \right\} \\ &\geqslant \sum_{Q_j\in\mathcal{Q}_\beta} \left\{ \rho_j E\big(L\sqrt{\beta\rho_j}\, Q, 1, (L\sqrt{\beta\rho_j})^2\big) + \int_{Q_j} V|u_j^{\mathrm{af}}|^2 \right\}, \end{aligned} \tag{4-13}$$

where the local gauge phase factors are defined as in previous arguments by demanding that (this is again possible because the left-hand side is curl-free in the simply connected domain Q_j)

$$\sum_{k=1,k\neq j}^{N_\beta} \boldsymbol{A}[|u_k^{\mathrm{af}}|^2] = \nabla\phi_j \quad \text{in } Q_j.$$

The minimum (Neumann) energy $E(Q_j, \beta, M_j)$ in the square Q_j is defined as in (2-4) and we used the scaling laws following from Lemma 3.4 as previously to obtain

$$E(Q_j, \beta, M_j) = \rho_j E\big(L\sqrt{\beta\rho_j}\, Q, 1, (L\sqrt{\beta\rho_j})^2\big),$$

with Q the unit square. Next, we note that (4-2) and (4-7) imply, using (4-12),

$$L_j = L\sqrt{\beta\rho_j} \geqslant \beta^\mu \to \infty$$

uniformly in j for all j such that $Q_j \in \mathcal{Q}_\beta$. Then, by Theorem 2.1,

$$\begin{aligned} \sum_{Q_j\in\mathcal{Q}_\beta} \rho_j E\big(L\sqrt{\beta\rho_j}\, Q, 1, (L\sqrt{\beta\rho_j})^2\big) &= \sum_{Q_j\in\mathcal{Q}_\beta} \beta L^2\rho_j^2\, E(L_j Q, 1, L_j^2)/L_j^2 \\ &= (1+o(1))\beta e(1,1) \sum_{Q_j\in\mathcal{Q}_\beta} L^2\rho_j^2 = (1+o(1))\beta e(1,1)\int_{\mathbb{R}^2} (\bar{\varrho}^{\mathrm{af}})^2. \end{aligned}$$

On the other hand, it follows from (2-11) that, on all the squares of $\mathcal{Q}_\beta$,

$$|\nabla V| \leqslant C\beta^{(s-1)/(s+2)+\varepsilon(s-1)},$$

and thus if

$$\widetilde{V}(\boldsymbol{x}) := \sum_{Q_j\in\mathcal{Q}_\beta} V(\boldsymbol{x}_j)\mathbb{1}_{Q_j}(\boldsymbol{x}), \tag{4-14}$$

we have

$$|V(\boldsymbol{x}) - \widetilde{V}(\boldsymbol{x})| \leqslant CL\beta^{(s-1)/(s+2)+\varepsilon(s-1)} = o(E_\beta^{\mathrm{TF}}) \quad \text{for any } \boldsymbol{x}\in\mathcal{Q}_\beta.$$

Recalling (4-8) and (4-9) we then have

$$\sum_{Q_j\in\mathcal{Q}_\beta}\int_{Q_j} V\,|u_j^{\mathrm{af}}|^2 = \int_{\mathbb{R}^2}\widetilde{V}\,\bar{\varrho}^{\mathrm{af}} + O(L\beta^{(s-1)/(s+2)+\varepsilon(s-1)}) = \int_{\mathbb{R}^2}\widetilde{V}\,\bar{\varrho}^{\mathrm{af}} + o(E_\beta^{\mathrm{TF}}). \tag{4-15}$$

The last assertion follows from (2-13) and (4-2), provided we take ε small enough; e.g., for $s>1$ (recall that the tiling squares have side length $L=\beta^\eta$),

$$\varepsilon \leqslant \frac{1}{2(s-1)}\left(\frac{s-1}{s+2}+\eta\right). \tag{4-16}$$

In the very same way however we can put back V in place of $\widetilde{V}$, obtaining

$$\sum_{Q_j\in\mathcal{Q}_\beta}\int_{Q_j} V\,|u_j^{\mathrm{af}}|^2 = \int_{\mathbb{R}^2}\widetilde{V}\,\bar{\varrho}^{\mathrm{af}} + o(E_\beta^{\mathrm{TF}}) = \int_{\mathbb{R}^2} V\,\bar{\varrho}^{\mathrm{af}} + o(E_\beta^{\mathrm{TF}}). \tag{4-17}$$

Combining (4-13), (4-15) and (4-17) yields

$$\begin{aligned} E_\beta^{\mathrm{af}} &\geqslant \int_{\mathbb{R}^2} V\,\bar{\varrho}^{\mathrm{af}} + (1+o(1))\beta e(1,1)\int_{\mathbb{R}^2}(\bar{\varrho}^{\mathrm{af}})^2 + o(E_\beta^{\mathrm{TF}}) \\ &\geqslant (1+o(1))\mathcal{E}_\beta^{\mathrm{TF}}[\bar{\varrho}^{\mathrm{af}}] + o(E_\beta^{\mathrm{TF}}) \geqslant (1+o(1))E_\beta^{\mathrm{TF}}(M) + o(E_\beta^{\mathrm{TF}}), \end{aligned} \tag{4-18}$$

where the latter energy denotes the ground state energy of the TF functional (2-9) minimized under the constraint that the L^1-norm be equal to M. Inserting (4-9) and using explicit expressions as in (2-13) and (2-14), one obtains

$$E_\beta^{\mathrm{TF}}(M) = (1+o(1))E_\beta^{\mathrm{TF}}$$

in the limit $\beta\to\infty$, thus completing the proof of (4-6).

4C. ***Density convergence.*** The lower bound in (4-6) together with the energy upper bound (4-1) implies that $\bar{\varrho}^{\mathrm{af}}$, the piecewise constant approximation of ϱ^{af} on scale $L=\beta^\eta$, is close in strong L^2 sense to $\varrho_\beta^{\mathrm{TF}}$. We will deduce (2-20) from the following.

Lemma 4.1 (Convergence of the piecewise approximation).
Let $\bar{\varrho}^{\mathrm{af}}$ be defined as in (4-8) *and $\varrho_\beta^{\mathrm{TF}}$ be the minimizer of* (2-9)*. Then*

$$\|\bar{\varrho}^{\mathrm{af}} - \varrho_\beta^{\mathrm{TF}}\|_{L^2(\mathbb{R}^2)} = o(\beta^{-1/(s+2)}) \tag{4-19}$$

in the limit $\beta\to\infty$.

Proof. Combining (4-1) and (4-18) we have

$$\mathcal{E}_\beta^{\mathrm{TF}}[\bar{\varrho}^{\mathrm{af}}] \leqslant E_\beta^{\mathrm{af}} + o(1)\beta^{s/(s+2)} \leqslant E_\beta^{\mathrm{TF}} + o(1)\beta^{s/(s+2)}. \tag{4-20}$$

The variational equation for $\varrho_\beta^{\mathrm{TF}}$ takes the form

$$2\beta e(1,1)\varrho_\beta^{\mathrm{TF}} + V = \lambda_\beta^{\mathrm{TF}} = E_\beta^{\mathrm{TF}} + \beta e(1,1)\int_{\mathbb{R}^2}(\varrho_\beta^{\mathrm{TF}})^2$$

on the support of $\varrho_\beta^{\mathrm{TF}}$ (recall (2-14) and (2-15)). Thus,

$$\begin{aligned}\int_{\mathbb{R}^2}(\bar\varrho^{\mathrm{af}}-\varrho_\beta^{\mathrm{TF}})^2 &= \int_{\mathbb{R}^2}\big((\bar\varrho^{\mathrm{af}})^2+(\varrho_\beta^{\mathrm{TF}})^2\big)-2\int_{\mathbb{R}^2}\bar\varrho^{\mathrm{af}}\varrho_\beta^{\mathrm{TF}}\\ &= \int_{\mathbb{R}^2}(\bar\varrho^{\mathrm{af}})^2+\int_{\mathbb{R}^2}(\varrho_\beta^{\mathrm{TF}})^2-\frac{1}{\beta e(1,1)}\int_{\mathbb{R}^2}\bar\varrho^{\mathrm{af}}(\lambda_\beta^{\mathrm{TF}}-V)_+\\ &\leqslant \frac{1}{\beta e(1,1)}\left[\mathcal{E}_\beta^{\mathrm{TF}}[\bar\varrho^{\mathrm{af}}]-\lambda_\beta^{\mathrm{TF}}+\beta e(1,1)\int_{\mathbb{R}^2}(\varrho_\beta^{\mathrm{TF}})^2\right]\\ &= \frac{1}{\beta e(1,1)}[\mathcal{E}_\beta^{\mathrm{TF}}[\bar\varrho^{\mathrm{af}}]-E_\beta^{\mathrm{TF}}]=o(\beta^{-2/(s+2)}),\end{aligned}$$

where we used (4-20) in the last step. □

By the definition (4-8) of $\bar\varrho^{\mathrm{af}}$ we also have, for any Lipschitz function ϕ with compact support,

$$\begin{aligned}\int_{\mathbb{R}^2}\phi(\beta^{-1/(s+2)}\boldsymbol{x})\bar\varrho^{\mathrm{af}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} &= \sum_{j=1}^{N_\beta}\int_{Q_j}\phi(\beta^{-1/(s+2)}\boldsymbol{x})\,\bar\varrho^{\mathrm{af}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\ &= \sum_{j=1}^{N_\beta}\phi(\beta^{-1/(s+2)}\boldsymbol{x}_j)\int_{Q_j}\varrho^{\mathrm{af}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}+O(\beta^{\eta-1/(s+2)}\|\phi\|_{\mathrm{Lip}})\\ &= \int_{\mathbb{R}^2}\phi(\beta^{-1/(s+2)}\boldsymbol{x})\,\varrho^{\mathrm{af}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}+O(\beta^{\eta-1/(s+2)}\|\phi\|_{\mathrm{Lip}}),\end{aligned}$$

using the normalization of ϱ^{af}. Furthermore, by Cauchy–Schwarz and Lemma 4.1 we obtain

$$\int_{\mathbb{R}^2}\phi(\beta^{-1/(s+2)}\boldsymbol{x})\big(\bar\varrho^{\mathrm{af}}(\boldsymbol{x})-\varrho_\beta^{\mathrm{TF}}(\boldsymbol{x})\big)\,\mathrm{d}\boldsymbol{x}=o(1)\|\phi\|_{L^2(\mathbb{R}^2)}.$$

Since the above estimates are uniform with respect to the Lipschitz norm of ϕ, we can take $\eta<1/(s+2)$, change scales in the above and recall (2-13) to deduce

$$\sup_{\substack{\phi\in C_0(B_R(0))\\ \|\phi\|_{\mathrm{Lip}}\leqslant 1}}\left|\int_{\mathbb{R}^2}\phi(\boldsymbol{x})\big(\beta^{2/(s+2)}\varrho^{\mathrm{af}}(\beta^{1/(s+2)}\boldsymbol{x})-\varrho_1^{\mathrm{TF}}(\boldsymbol{x})\big)\,\mathrm{d}\boldsymbol{x}\right|=o(1),\quad \beta\to\infty,$$

for fixed $R>0$, and hence (2-20).

Appendix: Properties of minimizers

In this appendix we recall a few fundamental properties of the average-field functional (1-1) in a trap V, respectively (2-1) on a domain Ω, as well as their minimizers.

As discussed in [Lundholm and Rougerie 2015, Appendix], the natural, maximal domain of $\mathcal{E}^{\mathrm{af}}$ is

$$\mathscr{D}^{\mathrm{af}}:=\big\{u\in H^1(\Omega):\textstyle\int_{\mathbb{R}^2}V|u|^2<\infty\big\},$$

and one may also use that the space $C_c^\infty(\mathbb{R}^2)$ is dense in this form domain with respect to $\mathcal{E}^{\mathrm{af}}$. Furthermore, [Lundholm and Rougerie 2015, Appendix: Proposition 3.7] ensures the existence of a minimizer $u^{\mathrm{af}}\in\mathscr{D}^{\mathrm{af}}$

of $\mathcal{E}^{\mathrm{af}}_\beta$ for any value of $\beta \in \mathbb{R}$ for confining potentials V, and by a similar proof and the compactness of the embedding $H^1_0(\Omega) \subset H^1(\Omega) \hookrightarrow L^p(\Omega)$, $1 \leqslant p < \infty$, the same holds for $\mathcal{E}^{\mathrm{af}}_\Omega$ for any bounded Ω with Lipschitz boundary:

Proposition A.1 (Existence of minimizers).
Let $\beta \in \mathbb{R}$ be arbitrary. Given any $V : \mathbb{R}^2 \to \mathbb{R}^+$ such that $-\Delta + V$ has compact resolvent, there exists $u^{\mathrm{af}} \in \mathscr{D}^{\mathrm{af}}$ with $\int_{\mathbb{R}^2} |u^{\mathrm{af}}|^2 = 1$ and $\mathcal{E}^{\mathrm{af}}_\beta[u^{\mathrm{af}}] = E^{\mathrm{af}}$. Moreover, if $M \geqslant 0$ and $\Omega \subset \mathbb{R}^2$ is bounded with Lipschitz boundary then there exists $u^{\mathrm{af}} \in H^1_{(0)}(\Omega)$ with $\int_\Omega |u^{\mathrm{af}}|^2 = M$ and $\mathcal{E}^{\mathrm{af}}_{\Omega,\beta}[u^{\mathrm{af}}] = E_{(0)}(\Omega, \beta, M)$.

Proof. The first part is [Lundholm and Rougerie 2015, Appendix: Proposition 3.7]. For $\Omega \subset \mathbb{R}^2$ we have by the Hölder, weak Young, and Sobolev inequalities, as well as Lemma 3.2, that

$$\begin{aligned}\big\| \boldsymbol{A}[|u|^2]u \big\|_{L^2(\Omega)} &\leqslant \big\| \boldsymbol{A}[|u|^2] \big\|_{L^4(\Omega)} \|u\|_{L^4(\Omega)} \\ &\leqslant C \big\| |u|^2 \big\|_{L^{4/3}(\Omega)} \|\nabla w_0\|_{L^{2,w}(\mathbb{R}^2)} \|u\|_{L^4(\Omega)} \leqslant C' \big\| |u| \big\|^3_{H^1(\Omega)} \leqslant C'(M + \mathcal{E}^{\mathrm{af}}_\Omega[u])^{3/2},\end{aligned}$$

and therefore

$$\|\nabla u\|_{L^2(\Omega)} = \big\| \nabla u + i\beta \boldsymbol{A}[|u|^2]u - i\beta \boldsymbol{A}[|u|^2]u \big\|_{L^2(\Omega)} \leqslant \mathcal{E}^{\mathrm{af}}[u]^{1/2} + C'|\beta|(M + \mathcal{E}^{\mathrm{af}}_\Omega[u])^{3/2}.$$

Hence, given a minimizing sequence

$$(u_n)_{n\to\infty} \subset H^1_{(0)}(\Omega), \quad \|u_n\|^2_{L^2(\Omega)} = M, \quad \lim_{n\to\infty} \mathcal{E}^{\mathrm{af}}_\Omega[u_n] = E_{(0)}(\Omega, \beta, M),$$

by uniform boundedness and the Rellich–Kondrachov theorem (see, for example, [Lieb and Loss 2001, Theorem 8.9]) there exists a convergent subsequence (again denoted u_n) and a limit $u^{\mathrm{af}} \in H^1_{(0)}(\Omega)$ such that

$$u_n \to u^{\mathrm{af}} \quad \text{in } L^p(\Omega),\ 1 \leqslant p < \infty, \qquad \nabla u_n \rightharpoonup \nabla u^{\mathrm{af}} \quad \text{in } L^2(\Omega).$$

Furthermore, by estimating

$$\big\| \boldsymbol{A}[|u_n|^2]u_n - \boldsymbol{A}[|u^{\mathrm{af}}|^2]u^{\mathrm{af}} \big\|_2 \leqslant \big\| \boldsymbol{A}[|u_n|^2 - |u^{\mathrm{af}}|^2]u_n \big\|_2 + \big\| \boldsymbol{A}[|u^{\mathrm{af}}|^2](u_n - u) \big\|_2$$

as above and using the strong convergence in $L^p(\Omega)$ for any $1 \leqslant p < \infty$, we have

$$\boldsymbol{A}[|u_n|^2]u_n \to \boldsymbol{A}[|u^{\mathrm{af}}|^2]u^{\mathrm{af}} \text{ in } L^2(\Omega).$$

Hence,

$$\begin{aligned}\big\| \big(\nabla + i\beta \boldsymbol{A}[|u^{\mathrm{af}}|^2]\big)u^{\mathrm{af}} \big\|_2 &= \sup_{\|v\|=1} \big| \langle \nabla u^{\mathrm{af}} + i\beta \boldsymbol{A}[|u^{\mathrm{af}}|^2]u^{\mathrm{af}}, v \rangle \big| = \sup_{\|v\|=1} \lim_{n\to\infty} \big| \langle \nabla u_n + i\beta \boldsymbol{A}[|u_n|^2]u_n, v \rangle \big| \\ &\leqslant \liminf_{n\to\infty} \sup_{\|v\|=1} \big| \langle \nabla u_n + i\beta \boldsymbol{A}[|u_n|^2]u_n, v \rangle \big| = \liminf_{n\to\infty} \big\| \big(\nabla + i\beta \boldsymbol{A}[|u_n|^2]\big)u_n \big\|_2,\end{aligned}$$

that is, $E_{(0)}(\Omega, \beta, M) \leqslant \mathcal{E}^{\mathrm{af}}_\Omega[u^{\mathrm{af}}] \leqslant \liminf_{n\to\infty} \mathcal{E}^{\mathrm{af}}_\Omega[u_n] = E_{(0)}(\Omega, \beta, M)$, and furthermore $\int_\Omega |u^{\mathrm{af}}|^2 = \lim_{n\to\infty} \int_\Omega |u_n|^2 = M$. □

For completeness, we finish with a derivation of the variational equation associated to the minimization of the energy functional (1-1). Let us define

$$\boldsymbol{J}[u] := \frac{i}{2}(u\nabla\bar{u} - \bar{u}\nabla u)$$

and for two vector functions $\boldsymbol{F}, \boldsymbol{G} : \mathbb{R}^2 \to \mathbb{R}^2$, their convolution

$$\boldsymbol{F} * \boldsymbol{G}(\boldsymbol{x}) := \int_{\mathbb{R}^2} \boldsymbol{F}(\boldsymbol{x}-\boldsymbol{y}) \cdot \boldsymbol{G}(\boldsymbol{y})\, \mathrm{d}\boldsymbol{y}.$$

Lemma A.2 (Variational equation).
Let $u = u^{\mathrm{af}}$ be a solution to (2-8)*. Then*

$$\left[\left(-i\nabla + \beta \boldsymbol{A}[|u|^2]\right)^2 + V - 2\beta \nabla^\perp w_0 * \left(\beta \boldsymbol{A}[|u|^2]|u|^2 + \boldsymbol{J}[u]\right)\right] u = \lambda u, \tag{A-1}$$

where

$$\begin{aligned}\lambda &= \mathcal{E}^{\mathrm{af}}[u] + \int_{\mathbb{R}^2} \left(2\beta \boldsymbol{A}[|u|^2] \cdot \boldsymbol{J}[u] + 2\beta^2 \left|\boldsymbol{A}[|u|^2]\right|^2 |u|^2\right) \\ &= \int_{\mathbb{R}^2} \left(1(|\nabla u|^2 + V|u|^2) + 2 \cdot 2\beta \boldsymbol{A}[|u|^2] \cdot \boldsymbol{J}[u] + 3\beta^2 \left|\boldsymbol{A}[|u|^2]\right|^2 |u|^2\right). \end{aligned} \tag{A-2}$$

(*Note that the factors* 1, 2, *and* 3 *correspond to the total degree of* $|u|^2$ *in each term.*)

Proof. Let

$$\begin{aligned}\mathcal{F}[u, \bar{u}, \lambda] &:= \mathcal{E}^{\mathrm{af}}[u, \bar{u}] + \lambda(1 - \textstyle\int |u|^2) \\ &= \int \left(|\nabla u|^2 + (V - \lambda)|u|^2 + \beta^2 \left|\boldsymbol{A}[|u|^2]\right|^2 |u|^2 + 2\beta \boldsymbol{A}[|u|^2] \cdot \boldsymbol{J}[u]\right) + \lambda, \\ \mathcal{E}_1[u, \bar{u}] &:= \int |\boldsymbol{A}[u\bar{u}]|^2 u\bar{u} = \iiint \nabla^\perp w_0(\boldsymbol{x}-\boldsymbol{y}) \cdot \nabla^\perp w_0(\boldsymbol{x}-\boldsymbol{z})\, u\bar{u}(\boldsymbol{x})\, u\bar{u}(\boldsymbol{y})\, u\bar{u}(\boldsymbol{z})\, \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{y}\, \mathrm{d}\boldsymbol{z}, \\ \mathcal{E}_2[u, \bar{u}] &:= \int \boldsymbol{A}[u\bar{u}] \cdot i(u\nabla\bar{u} - \bar{u}\nabla u) = \iint \nabla^\perp w_0(\boldsymbol{x}-\boldsymbol{y})\, u\bar{u}(\boldsymbol{y}) \cdot i(u\nabla\bar{u} - \bar{u}\nabla u)(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{y}. \end{aligned}$$

We have

$$\begin{aligned}\mathcal{E}_1[u, \bar{u}+\varepsilon v] = \mathcal{E}_1[u, \bar{u}] + \varepsilon \iiint \Big(&\nabla^\perp w_0(\boldsymbol{x}-\boldsymbol{y}) \cdot \nabla^\perp w_0(\boldsymbol{x}-\boldsymbol{z}) \big(v(\boldsymbol{x})u(\boldsymbol{x})|u(\boldsymbol{y})|^2 |u(\boldsymbol{z})|^2 \\ &+ |u(\boldsymbol{x})|^2 u(\boldsymbol{y}) v(\boldsymbol{y}) |u(\boldsymbol{z})|^2 + |u(\boldsymbol{x})|^2 |u(\boldsymbol{y})|^2 u(\boldsymbol{z}) v(\boldsymbol{z})\big)\Big)\, \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{y}\, \mathrm{d}\boldsymbol{z} + O(\varepsilon^2). \end{aligned}$$

Hence at $O(\varepsilon)$,

$$\begin{aligned}&\int_{\boldsymbol{x}} v(\boldsymbol{x})u(\boldsymbol{x})\boldsymbol{A}[|u|^2]^2\, \mathrm{d}\boldsymbol{x} - \int_{\boldsymbol{y}} v(\boldsymbol{y})u(\boldsymbol{y}) \int_{\boldsymbol{x}} \nabla^\perp w_0(\boldsymbol{y}-\boldsymbol{x})|u(\boldsymbol{x})|^2 \cdot \int_{\boldsymbol{z}} \nabla^\perp w_0(\boldsymbol{x}-\boldsymbol{z})|u(\boldsymbol{z})|^2\, \mathrm{d}\boldsymbol{z}\, \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{y} \\ &\qquad - \int_{\boldsymbol{z}} v(\boldsymbol{z})u(\boldsymbol{z}) \int_{\boldsymbol{x}} \nabla^\perp w_0(\boldsymbol{z}-\boldsymbol{x})|u(\boldsymbol{x})|^2 \cdot \int_{\boldsymbol{y}} \nabla^\perp w_0(\boldsymbol{x}-\boldsymbol{y})|u(\boldsymbol{y})|^2\, \mathrm{d}\boldsymbol{y}\, \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{z} \\ &= \int vu\boldsymbol{A}[|u|^2]^2 - 2\int vu\nabla^\perp w_0 * |u|^2 \boldsymbol{A}[|u|^2]. \end{aligned}$$

Also

$$\begin{aligned}\mathcal{E}_2[u, \bar{u}+\varepsilon v] = \mathcal{E}_2[u, \bar{u}] + \varepsilon \iint \Big(&\nabla^\perp w_0(\boldsymbol{x}-\boldsymbol{y})u(\boldsymbol{y})v(\boldsymbol{y}) \cdot i(u\nabla\bar{u} - \bar{u}\nabla u)(\boldsymbol{x}) \\ &+ \nabla^\perp w_0(\boldsymbol{x}-\boldsymbol{y})|u(\boldsymbol{y})|^2 \cdot i\big(u(\boldsymbol{x})\nabla v(\boldsymbol{x}) - v(\boldsymbol{x})\nabla u(\boldsymbol{x})\big)\Big)\, \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{y} + O(\varepsilon^2), \end{aligned}$$

hence at $O(\varepsilon)$ and using $\nabla \cdot \boldsymbol{A} = 0$,

$$-\int_y v(y)u(y)\int_x \nabla^\perp w_0(y-x)\cdot 2\boldsymbol{J}[u](x)\,\mathrm{d}x\,\mathrm{d}y - i\int v(x)\nabla u(x)\cdot \boldsymbol{A}[|u|^2](x)\,\mathrm{d}x$$
$$+\underbrace{i\int u(x)\boldsymbol{A}[|u|^2](x)\cdot\nabla v(x)\,\mathrm{d}x}_{=\{PI\}=-i\int \nabla u\cdot \boldsymbol{A}v - i\int u(\nabla\cdot\boldsymbol{A})v} = -2\int vu\nabla^\perp w_0 * \boldsymbol{J}[u] - 2i\int v\nabla u\cdot \boldsymbol{A}[|u|^2].$$

Thus

$$\mathcal{F}[u,\bar u+\varepsilon v,\lambda] = \mathcal{F}[u,\bar u,\lambda] + \varepsilon\int v\Big[(-\Delta+V-\lambda)u + \beta^2\big|\boldsymbol{A}[|u|^2]\big|^2 u - 2\beta^2\nabla^\perp w_0 * |u|^2\boldsymbol{A}[|u|^2]u$$
$$-2\beta\nabla^\perp w_0 * \boldsymbol{J}[u]u - 2i\beta\boldsymbol{A}[|u|^2]\cdot\nabla u\Big] + O(\varepsilon^2),$$

and using

$$\big(-i\nabla+\beta\boldsymbol{A}[|u|^2]\big)^2 u = -\Delta u - 2i\beta\boldsymbol{A}[|u|^2]\cdot\nabla u + \beta^2\boldsymbol{A}[|u|^2]^2 u,$$

we arrive at (A-1).

For (A-2) we use $\int |u|^2 = 1$ by multiplying (A-1) with $\bar u$ and integrating:

$$\lambda = \mathcal{E}^{\mathrm{af}}[u] - 2\beta\int |u|^2\nabla^\perp w_0 * \big(\beta\boldsymbol{A}[|u|^2]|u|^2 + \boldsymbol{J}[u]\big).$$

We then use that

$$\int |u|^2\nabla^\perp w_0 * \boldsymbol{A}[|u|^2]|u|^2 = \iiint |u(x)|^2\nabla^\perp w_0(x-y)\cdot\nabla^\perp w_0(y-z)|u(z)|^2|u(y)|^2\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$$
$$= -\iiint \nabla^\perp w_0(y-x)\cdot\nabla^\perp w_0(y-z)|u(x)|^2|u(z)|^2|u(y)|^2\,\mathrm{d}x\,\mathrm{d}z\,\mathrm{d}y$$
$$= -\int \boldsymbol{A}[|u|^2]^2|u|^2$$

and

$$2\int |u|^2\nabla^\perp w_0 * \boldsymbol{J}[u] = \iint |u(x)|^2\nabla^\perp w_0(x-y)\cdot i\big(u(y)\nabla\bar u(y) - \bar u(y)\nabla u(y)\big)\,\mathrm{d}x\,\mathrm{d}y$$
$$= -\int_y i(u\nabla\bar u - \bar u\nabla u)(y)\cdot\int_x \nabla^\perp w_0(y-x)|u(x)|^2\,\mathrm{d}x\,\mathrm{d}y = -2\int \boldsymbol{J}[u]\cdot\boldsymbol{A}[|u|^2]$$

to arrive at (A-2). □

Acknowledgments

This work is supported by MIUR through the FIR grant 2013 "Condensed Matter in Mathematical Physics (Cond-Math)" (code RBFR13WAET), the Swedish Research Council (grant no. 2013-4734) and the ANR (Project MaThoStaQ ANR-13-JS01-0005-01). We thank Jan Philip Solovej for insightful suggestions and Romain Duboscq for inspiring numerical simulations. D. Lundholm also thanks Simon Larson for discussions.

References

[Aftalion 2007] A. Aftalion, "Vortex patterns in Bose Einstein condensates", pp. 1–18 in *Perspectives in nonlinear partial differential equations*, edited by H. Berestycki et al., Contemp. Math. **446**, Amer. Math. Soc., Providence, RI, 2007. MR Zbl

[Arovas et al. 1984] D. Arovas, J. Schrieffer, and F. Wilczek, "Fractional statistics and the quantum Hall effect", *Phys. Rev. Lett.* **53**:7 (1984), 722–723.

[Bach 1992] V. Bach, "Error bound for the Hartree–Fock energy of atoms and molecules", *Comm. Math. Phys.* **147**:3 (1992), 527–548. MR Zbl

[Bethuel et al. 1994] F. Bethuel, H. Brezis, and F. Hélein, *Ginzburg–Landau vortices*, Progress in Nonlinear Differential Equations and their Applications **13**, Birkhäuser, Boston, 1994. MR Zbl

[Catto et al. 1998] I. Catto, C. Le Bris, and P.-L. Lions, *The mathematical theory of thermodynamic limits: Thomas–Fermi type models*, Clarendon/Oxford University Press, New York, 1998. MR Zbl

[Chen et al. 1989] Y.-H. Chen, F. Wilczek, E. Witten, and B. I. Halperin, "On anyon superconductivity", *Internat. J. Modern Phys. B* **3**:7 (1989), 1001–1067. MR

[Chitra and Sen 1992] R. Chitra and D. Sen, "Ground state of many anyons in a harmonic potential", *Phys. Rev. B* **46**:17 (1992), 10923–10930.

[Cooper and Simon 2015] N. R. Cooper and S. H. Simon, "Signatures of fractional exclusion statistics in the spectroscopy of quantum Hall droplets", *Phys. Rev. Lett.* **114**:10 (2015), art. id. 106802.

[Correggi and Rougerie 2013] M. Correggi and N. Rougerie, "Inhomogeneous vortex patterns in rotating Bose–Einstein condensates", *Comm. Math. Phys.* **321**:3 (2013), 817–860. MR Zbl

[Correggi and Yngvason 2008] M. Correggi and J. Yngvason, "Energy and vorticity in fast rotating Bose–Einstein condensates", *J. Phys. A* **41**:44 (2008), art. id. 45002. MR Zbl

[Correggi et al. 2011] M. Correggi, N. Rougerie, and J. Yngvason, "The transition to a giant vortex phase in a fast rotating Bose–Einstein condensate", *Comm. Math. Phys.* **303**:2 (2011), 451–508. MR Zbl

[Correggi et al. 2012] M. Correggi, F. Pinsker, N. Rougerie, and J. Yngvason, "Critical rotational speeds for superfluids in homogeneous traps", *J. Math. Phys.* **53**:9 (2012), art. id. 095203. MR Zbl

[Cycon et al. 1987] H. L. Cycon, R. G. Froese, W. Kirsch, and B. Simon, *Schrödinger operators with application to quantum mechanics and global geometry*, Springer, 1987. MR Zbl

[Evans 1998] L. C. Evans, *Partial differential equations*, Graduate Studies in Mathematics **19**, Amer. Math. Soc., Providence, RI, 1998. MR Zbl

[Fournais and Helffer 2010] S. Fournais and B. Helffer, *Spectral methods in surface superconductivity*, Progress in Nonlinear Differential Equations and their Applications **77**, Birkhäuser, Boston, 2010. MR Zbl

[Fournais et al. 2015] S. Fournais, M. Lewin, and J. P. Solovej, "The semi-classical limit of large fermionic systems", 2015. arXiv

[Haldane 1983] F. D. M. Haldane, "Fractional quantization of the Hall effect: a hierarchy of incompressible quantum fluid states", *Phys. Rev. Lett.* **51**:7 (1983), 605–608. MR

[Halperin 1984] B. I. Halperin, "Statistics of quasiparticles and the hierarchy of fractional quantized Hall states", *Phys. Rev. Lett.* **52**:18 (1984), 1583–1586.

[Iengo and Lechner 1992] R. Iengo and K. Lechner, "Anyon quantum mechanics and Chern–Simons theory", *Phys. Rep.* **213**:4 (1992), 179–269. MR

[Khare 2005] A. Khare, *Fractional statistics and quantum theory*, 2nd ed., World Scientific, Singapore, 2005. Zbl

[Larson and Lundholm 2016] S. Larson and D. Lundholm, "Exclusion bounds for extended anyons", preprint, 2016. arXiv

[Li et al. 1992] S. Li, R. K. Bhaduri, and M. V. N. Murthy, "Thomas–Fermi approximation for confined anyons", *Phys. Rev. B* **46**:2 (1992), 1228–1231.

[Lieb 1981] E. H. Lieb, "Thomas–Fermi and related theories of atoms and molecules", *Rev. Modern Phys.* **53**:4 (1981), 603–641. MR Zbl

[Lieb and Loss 2001] E. H. Lieb and M. Loss, *Analysis*, 2nd ed., Graduate Studies in Mathematics **14**, Amer. Math. Soc., Providence, RI, 2001. MR Zbl

[Lieb and Seiringer 2006] E. H. Lieb and R. Seiringer, "Derivation of the Gross–Pitaevskii equation for rotating Bose gases", *Comm. Math. Phys.* **264**:2 (2006), 505–537. MR Zbl

[Lieb and Simon 1977] E. H. Lieb and B. Simon, "The Hartree–Fock theory for Coulomb systems", *Comm. Math. Phys.* **53**:3 (1977), 185–194. MR

[Lieb et al. 2001] E. H. Lieb, R. Seiringer, and J. Yngvason, "A rigorous derivation of the Gross–Pitaevskii energy functional for a two-dimensional Bose gas", *Comm. Math. Phys.* **224**:1 (2001), 17–31. MR Zbl

[Lieb et al. 2005] E. H. Lieb, R. Seiringer, J. P. Solovej, and J. Yngvason, *The mathematics of the Bose gas and its condensation*, Oberwolfach Seminars **34**, Birkhäuser, Basel, 2005. MR Zbl

[Lions 1987] P.-L. Lions, "Solutions of Hartree–Fock equations for Coulomb systems", *Comm. Math. Phys.* **109**:1 (1987), 33–97. MR Zbl

[Lions 1988] P.-L. Lions, "Hartree–Fock and related equations", pp. 304–333 in *Nonlinear partial differential equations and their applications* (Paris, 1985–1986), edited by J.-L. Lions, Pitman Res. Notes Math. Ser. **181**, Longman Sci. Tech., Harlow, 1988. MR Zbl

[Lundholm 2016] D. Lundholm, "Many-anyon trial states", preprint, 2016. To appear in *Phys. Rev. A*. arXiv

[Lundholm and Rougerie 2015] D. Lundholm and N. Rougerie, "The average field approximation for almost bosonic extended anyons", *J. Stat. Phys.* **161**:5 (2015), 1236–1267. MR Zbl

[Lundholm and Rougerie 2016] D. Lundholm and N. Rougerie, "Emergence of fractional statistics for tracer particles in a Laughlin liquid", *Phys. Rev. Lett.* **116**:17 (2016), art. id. 170401.

[Myrheim 1999] J. Myrheim, "Anyons", pp. 265–413 in *Aspects topologiques de la physique en basse dimension* (Les Houches, 1998), edited by A. Comtet et al., EDP Sci., Les Ulis, 1999. MR Zbl

[Nam et al. 2016] P. T. Nam, N. Rougerie, and R. Seiringer, "Ground states of large bosonic systems: the Gross–Pitaevskii limit revisited", *Anal. PDE* **9**:2 (2016), 459–485. MR Zbl

[Ouvry 2009] S. Ouvry, "Anyons and lowest Landau level anyons", pp. 71–103 in *The spin: Poincaré Seminar 2007*, edited by B. Duplantier et al., Progress in Mathematical Physics **55**, Birkhäuser, Basel, 2009. Zbl

[Rougerie 2014] N. Rougerie, "Théorèmes de de Finetti, limites de champ moyen et condensation de Bose–Einstein", lecture notes, 2014. arXiv

[Rougerie 2015] N. Rougerie, "De finetti theorems, mean-field limits and Bose–Einstein condensation", LMU lecture notes, 2015. arXiv

[Sandier and Serfaty 2007] E. Sandier and S. Serfaty, *Vortices in the magnetic Ginzburg–Landau model*, Progress in Nonlinear Differential Equations and their Applications **70**, Birkhäuser, Boston, 2007. MR Zbl

[Sigal 2015] I. M. Sigal, "Magnetic vortices, Abrikosov lattices, and automorphic functions", pp. 19–58 in *Mathematical and computational modeling*, edited by R. Melnik, Wiley, Hoboken, NJ, 2015. MR

[Trugenberger 1992a] C. A. Trugenberger, "The anyon fluid in the Bogoliubov approximation", *Phys. Rev. D* **45**:10 (1992), 3807–3817.

[Trugenberger 1992b] C. A. Trugenberger, "Ground state and collective excitations of extended anyons", *Phys. Lett. B* **288**:1-2 (1992), 121–128.

[Wen and Zee 1990] X. G. Wen and A. Zee, "Compressibility and superfluidity in the fractional-statistics liquid", *Phys. Rev. B* **41**:1 (1990), 240–253.

[Westerberg 1993] E. Westerberg, "Mean field approximation for anyons in a magnetic field", *Int. J. Mod. Phys. B* **7**:11 (1993), 2177–2199.

[Wilczek 1990] F. Wilczek, *Fractional statistics and anyon superconductivity*, World Scientific, Teaneck, NJ, 1990. MR Zbl

[Zhang et al. 2014] Y. Zhang, N. D. Sreejith, N. D. Gemelke, and J. K. Jain, "Fractional angular momentum in cold-atom systems", *Phys. Rev. Lett.* **113**:16-17 (2014), art. id. 160404.

Received 3 Nov 2016. Revised 3 Mar 2017. Accepted 3 Apr 2017.

MICHELE CORREGGI: michele.correggi@gmail.com
Dipartimento di Matematica "G. Castelnuovo", Università degli Studi di Roma "La Sapienza", P. le Aldo Moro, 5, 00185 Rome, Italy

DOUGLAS LUNDHOLM: dogge@math.kth.se
KTH Royal Institute of Technology, Department of Mathematics, SE-100 44 Stockholm, Sweden

NICOLAS ROUGERIE: nicolas.rougerie@lpmmc.cnrs.fr
CNRS & Université Grenoble Alpes, LPMMC (UMR 5493), B.P. 166, 38042 Grenoble, France

ANALYSIS AND PDE
Vol. 10, No. 5, 2017
dx.doi.org/10.2140/apde.2017.10.1201

REGULARITY OF VELOCITY AVERAGES FOR TRANSPORT EQUATIONS ON RANDOM DISCRETE VELOCITY GRIDS

NATHALIE AYI AND THIERRY GOUDON

We go back to the question of the regularity of the "velocity average" $\int f(x,v)\psi(v)\,\mathrm{d}\mu(v)$ when f and $v\cdot\nabla_x f$ both belong to L^2, and the variable v lies in a discrete subset of $\mathbb{R}^D$. First of all, we provide a rate, depending on the number of velocities, for the defect of $H^{1/2}$ regularity which is reached when v ranges over a continuous set. Second of all, we show that the $H^{1/2}$ regularity holds in expectation when the set of velocities is chosen randomly. We apply this statement to investigate the consistency with the diffusion asymptotics of a Monte Carlo-like discrete velocity model.

1. Introduction

The averaging lemma is now a classical tool for the analysis of kinetic equations. Roughly speaking it can be explained as follows. Let $\mathscr{V}\subset\mathbb{R}^D$, endowed with a measure $\mathrm{d}\mu$. We consider a sequence of functions $f_n:\mathbb{R}^D\times\mathscr{V}\to\mathbb{R}$. We assume that

(a) $(f_n)_{n\in\mathbb{N}}$ is bounded in $L^2(\mathbb{R}^D\times\mathscr{V})$,

(b) $(v\cdot\nabla_x f_n)_{n\in\mathbb{N}}$ is bounded in $L^2(\mathbb{R}^D\times\mathscr{V})$.

Given $\psi\in C_c^\infty(\mathbb{R}^D)$, we are interested in the *velocity average*

$$\rho_n[\psi](x)=\int_{\mathscr{V}} f_n(x,v)\psi(v)\,\mathrm{d}\mu(v).$$

Of course, (a) already tells us that $(\rho_n[\psi])_{n\in\mathbb{N}}$ is bounded in $L^2(\mathbb{R}^D)$. We wish to obtain further regularity or compactness properties, as a consequence of the additional assumption (b), and the fact that we are averaging with respect to the variable v. The first result in that direction dates back to [Bardos et al. 1988] (see also [Agoshkov 1984]); it asserts that $(\rho_n[\psi])_{n\in\mathbb{N}}$ is bounded in the Sobolev space $H^{1/2}(\mathbb{R}^D)$ and it is thus relatively compact in $L^2_{\mathrm{loc}}(\mathbb{R}^D)$, by virtue of the standard Rellich's theorem. This basic result has been improved in many directions: L^2 can be replaced by the L^p framework, at least with $1<p<\infty$, and we can relax (b) by allowing derivatives with respect to v and certain loss of regularity with respect to x; see, among others, [DiPerna et al. 1991; Golse et al. 1988; Perthame and Souganidis 1998]. Time-derivative or force terms can be considered as well; see, in addition to the above-mentioned references, [Berthelin and Junca 2010]. Such an argument plays a crucial role in the stunning theory of "renormalized solutions" of the Boltzmann equation [DiPerna and Lions 1989b], and more generally in proving the existence of solutions to nonlinear kinetic models like in [DiPerna and

MSC2010: primary 35B65; secondary 35F05, 35Q20, 82C40.
Keywords: average lemma, discrete velocity models, random velocity grids, hydrodynamic limits.

Lions 1989a]. It is equally a crucial ingredient for the analysis of hydrodynamic regimes, which establish the connection between microscopic models and fluid mechanics systems, and for the asymptotic of the Boltzmann equation to the incompressible Navier–Stokes system, which needs a suitable L^1 version of the average lemma [Golse and Saint-Raymond 2002]; we refer the reader to [Golse and Saint-Raymond 2004; Saint-Raymond 2009; Villani 2002]. Finally, it is worth pointing out that the averaging lemma can be used to investigate the regularizing effects of certain PDEs (convection-diffusion and elliptic equations, nonlinear conservation laws, etc.) [Tadmor and Tao 2007].

In order to illustrate our purpose, let us consider the following simple model which can be motivated from radiative transfer theory:

$$\varepsilon\,\partial_t f_\varepsilon + v\cdot\nabla_x f_\varepsilon = \frac{1}{\varepsilon}\sigma(\rho_\varepsilon)(\rho_\varepsilon - f_\varepsilon), \tag{1-1}$$

where

$$\rho_\varepsilon(t,x) = \int_{\mathscr{V}} f_\varepsilon(t,x,v)\,\mathrm{d}\mu(v),$$

and $\sigma : [0,\infty) \to [0,\infty)$ is a given smooth function. The parameter $0 < \varepsilon \ll 1$ is defined from physical quantities. As it tends to 0, both $f_\varepsilon(t,x,v)$ and $\rho_\varepsilon(t,x)$ converge to $\rho(t,x)$, which satisfies the nonlinear diffusion equation

$$\partial_t \rho = \nabla_x\cdot(A\nabla_x F(\rho)), \quad A = \int_{\mathscr{V}} v\otimes v\,\mathrm{d}\mu(v), \quad F(\rho) = \int_0^\rho \frac{\mathrm{d}z}{\sigma(z)}. \tag{1-2}$$

The averaging lemma is an efficient tool to deal with the nonlinearity of such a problem, as discussed in [Bardos et al. 1988].

However the discussion above hides the fact that we need some assumptions on the measured set of velocities $(\mathscr{V}, \mathrm{d}\mu)$ in order to obtain the regularization property of the velocity averaging. Roughly speaking, we need "enough" directions v when we consider the derivatives in (b). More technically, the compactness statement holds provided for any $0 < R < \infty$ we can find $C_R > 0,\ \delta_0 > 0,\ \gamma > 0$ such that for $0 < \delta < \delta_0$ and $\xi \in \mathbb{S}^{N-1}$, we have

$$\mathrm{meas}\big(\{v\in\mathscr{V}\cap B(0,R) : |v\cdot\xi|\le\delta\}\big) \le C_R\delta^\gamma.$$

This assumption appears in many statements about regularity of the velocity averages; when we are only interested in the compactness issue, it can be replaced by a more intuitive assumption (see, e.g., [Golse 2000, Theorem 1 in Lecture 3]): for any $\xi \in \mathbb{S}^{N-1}$ we have

$$\mathrm{meas}\big(\{v\in\mathscr{V}\cap B(0,R) : v\cdot\xi = 0\}\big) = 0. \tag{1-3}$$

Clearly these assumptions are satisfied when the measure $\mathrm{d}\mu$ is absolutely continuous with respect to the Lebesgue measure (with, for the sake of concreteness, $\mathscr{V} = \mathbb{R}^D$ or $\mathscr{V} = \mathbb{S}^{D-1}$). However, they fail for models based on a discrete set of velocities. For instance let $\mathscr{V} = \{v_1,\dots,v_N\}$, with $v_j \in \mathbb{R}^D$, and $\mathrm{d}\mu(v) = \frac{1}{N}\sum_{j=1}^N \delta(v=v_j)$; it suffices to pick $\xi \in \mathbb{S}^{N-1}$ orthogonal to one of the v_j to contradict (1-3). (Note that alternative proofs based on compensated compactness techniques have been proposed to justify the asymptotic regime from (1-1) to (1-2) that apply to certain discrete velocity models; see [Degond et al.

2000; Goudon and Poupaud 2001; Lions and Toscani 1997].) Nevertheless, when the discrete velocities come from a discretization grid of the whole space, the averaging lemma can be recovered asymptotically letting the mesh step go to 0, as shown in [Mischler 1997], motivated by the convergence analysis of numerical schemes for the Boltzmann equation.

This paper aims at investigating further these issues. To be more specific, in Section 2 we revisit the averaging lemma for discrete velocities in two directions. First of all, we make more precise the analysis of [Mischler 1997], obtaining a rate on the defect to the $H^{1/2}$ regularity of the velocity average, depending on the mesh size. Second of all, we establish a stochastic version of the averaging lemma. We are still working with a finite number of velocities on bounded sets; however, choosing the velocities randomly, the "compactifying" property of assumption (b) can be restored by dealing with the expectation of $\rho_n[\psi]$. This is a natural way to involve "enough velocities", by looking at a large set of realizations of the discrete velocity grid. The analysis is completed in Section 3 by going back to the asymptotic problem $\varepsilon \to 0$ in (1-1), with a random discretization of the velocity variable, in the spirit of the Monte Carlo approach.

2. Discrete velocity averaging lemmas

Deterministic case: evaluation of the defect. As mentioned above, it is a well-known fact that, in the deterministic context, the averaging lemma fails for discrete velocity models. However, as established by S. Mischler [1997], the compactness of velocity averages is recovered asymptotically when we refine a velocity grid in order to recover a continuous velocity model. Here, we wish to quantify the defect of compactness when the number of velocities is finite and fixed. This is the aim of the following claim which shows that the macroscopic density $\rho[\psi]$ "belongs to $H^{1/2}(\mathbb{R}^D) + O(1/\sqrt{N})L^2(\mathbb{R}^D)$".

Proposition 2.1. *Let $N \in \mathbb{N} \setminus \{0\}$ and define*

$$A_N = \Big(\frac{1}{N}\mathbb{Z}\Big)^D \cap [-0.5, 0.5]^D.$$

Let $f, g \in L^2(\mathbb{R}^D \times A_N)$ satisfy, for all $k \in \mathbb{Z}^D$,

$$v_k \cdot \nabla_x f(x, v_k) = g(x, v_k). \tag{2-1}$$

We suppose that the L^2 norm of f and g is bounded uniformly with respect to N. Then, for all $\psi \in C_c^\infty(\mathbb{R}^D)$, the macroscopic quantity

$$\rho[\psi](x) = \frac{1}{(N+1)^D} \sum_k f(x, v_k)\psi(v_k)$$

can be split as $\rho[\psi](x) = \Theta[\psi](x) + (1/\sqrt{N})\widetilde{\Delta[\psi]}(x)$, where $\Theta[\psi]$ and $\widetilde{\Delta[\psi]}$ are bounded uniformly with respect to N in $H^{1/2}(\mathbb{R}^D)$ and $L^2(\mathbb{R}^D)$ respectively.

Remark 2.2. Note that in this statement N is the number of grid points per axis. Accordingly, there are $\mathcal{N} = (N+1)^D$ velocities in the set A_N. Therefore the defect of $H^{1/2}$ regularity decays like $\mathcal{N}^{1/2D}$, depending on the dimension.

Proof. As usual, we start by applying the Fourier transform to (2-1). Then for all $k \in \mathbb{Z}$ and $\xi \in \mathbb{R}^D$, we get

$$\xi \cdot v_k \, \hat{f}(\xi, v_k) = (-i)\hat{g}(\xi, v_k).$$

Let us set

$$F(\xi) := \left(\frac{1}{(N+1)^D} \sum_k |\hat{f}(\xi, v_k)|^2 \right)^{1/2}, \quad G(\xi) := \left(\frac{1}{(N+1)^D} \sum_k |\hat{g}(\xi, v_k)|^2 \right)^{1/2}.$$

By assumption, we have $F, G \in L^2_\xi$. Still following the standard arguments, we pick $\delta > 0$ and we split

$$\begin{aligned} \hat{\rho}[\psi](\xi) &= \frac{1}{(N+1)^D} \sum_k \hat{f}(\xi, v_k)\psi(v_k) \\ &= \frac{1}{(N+1)^D} \sum_{|\xi \cdot v_k| < \delta|\xi|} \hat{f}(\xi, v_k)\psi(v_k) + \frac{1}{(N+1)^D} \sum_{|\xi \cdot v_k| \geq \delta|\xi|} \hat{f}(\xi, v_k)\psi(v_k). \end{aligned}$$

The Cauchy–Schwarz inequality permits us to dominate the first term:

$$\left| \frac{1}{(N+1)^D} \sum_{|\xi \cdot v_k| < \delta|\xi|} \hat{f}(\xi, v_k)\psi(v_k) \right| \leq \|\psi\|_\infty \left(\frac{1}{(N+1)^D} \sum_k |\hat{f}(\xi, v_k)|^2 \right)^{1/2} \left(\frac{1}{(N+1)^D} \sum_{|\xi \cdot v_k| < \delta|\xi|} 1 \right)^{1/2}. \quad (2\text{-}2)$$

For the second term, we use the information in (2-1); it yields

$$\begin{aligned} &\left| \frac{1}{(N+1)^D} \sum_{|\xi \cdot v_k| \geq \delta|\xi|} \hat{f}(\xi, v_k)\psi(v_k) \right| \\ &\qquad = \left| \frac{1}{(N+1)^D} \sum_{|\xi \cdot v_k| \geq \delta|\xi|} \frac{(-i)\hat{g}(\xi, v_k)}{\xi \cdot v_k} \, \psi(v_k) \right| \\ &\qquad \leq \|\psi\|_\infty \left(\frac{1}{(N+1)^D} \sum_k |\hat{g}(\xi, v_k)|^2 \right)^{1/2} \left(\frac{1}{(N+1)^D} \sum_{|\xi \cdot v_k| \geq \delta|\xi|} \frac{1}{|\xi \cdot v_k|^2} \right)^{1/2}. \quad (2\text{-}3) \end{aligned}$$

From now on we assume $\xi \neq 0$. Let $(e_1, \ldots, e_D)$ stand for the canonical basis of $\mathbb{R}^D$ so that $\xi = \sum_{j=1}^D \alpha_j e_j$ with $\alpha_j \in \mathbb{R}$. We distinguish the following two cases:

(i) ξ is aligned with an axis, that is, all but one the α_j vanish, or

(ii) ξ is generated by at least two vectors of the basis.

We start with the case (i), assuming for instance $\xi = \alpha e_1$. Then $\xi \cdot v_k = \alpha v_k^1$, where v_k^1 is the first component of the vector v_k.

We refer the reader to Figure 1 to complete the discussion. On each horizontal line we find $2\lfloor \delta N \rfloor + 1$ velocities such that $|\xi \cdot v_k| < \delta|\xi|$, where $\lfloor s \rfloor$ stands for the integer part of s. Thus, since there are $(N+1)^{D-1}$ such lines on the domain A_N, we obtain

$$\sum_{|\xi \cdot v_k| < \delta|\xi|} 1 = (2\lfloor \delta N \rfloor + 1)(N+1)^{D-1} = 2\Big(\delta + \frac{1}{N}\Big)(N+1)^D.$$

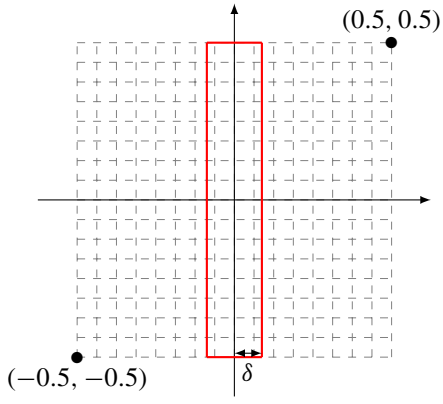

Figure 1. The delimited area corresponds to $|\xi \cdot v_k| < \delta|\xi|$ for ξ collinear to e_1.

Coming back to (2-2), we arrive at

$$\left|\frac{1}{(N+1)^D}\sum_{|\xi\cdot v_k|<\delta|\xi|}\hat{f}(\xi, v_k)\psi(v_k)\right| \le CF(\xi)\sqrt{\delta+\frac{1}{N}},$$

where $C > 0$ is a generic constant which does not depend on N and ξ.

Next, we cover the set of velocities such that $|v_k \cdot \xi| \ge \delta|\xi|$ by strips of width δ; see Figure 2 in dimension $D = 2$. We denote by S_p the p-th strip delimited by the straight lines $x = p\delta$ and $x = (p+1)\delta$. Each velocity on the strip S_p satisfies $p\delta \le v_k^1 \le (p+1)\delta$. Moreover, given a strip S_p, we cannot find more than $\lfloor \delta N \rfloor + 1$ abscissae in the strip and there are $(N+1)^{D-1}$ lines in the domain. It follows that

$$\sum_{|\xi\cdot v_k|\ge\delta|\xi|}\frac{1}{|\xi\cdot v_k|^2} = \sum_{|\xi\cdot v_k|\ge\delta|\xi|}\frac{1}{|\xi|^2}\,\frac{1}{|\xi/|\xi|.v_k|^2}$$
$$\le \frac{1}{|\xi|^2}\,2\left(\sum_{p\ge1}\frac{1}{(p\delta)^2}\right)(\delta N+1)(N+1)^{D-1} \le \frac{1}{|\xi|^2}\,2\left(\sum_{p\ge1}\frac{1}{p^2}\right)\frac{1}{\delta}\left(1+\frac{1}{\delta N}\right)(N+1)^D.$$

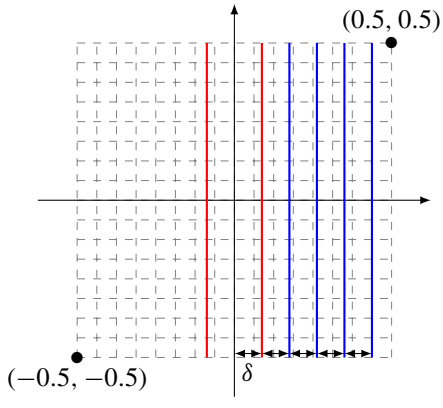

Figure 2. Splitting of the velocity space in strips of width δ. Since this space is symmetric, we only deal with the part corresponding to positive abscissae.

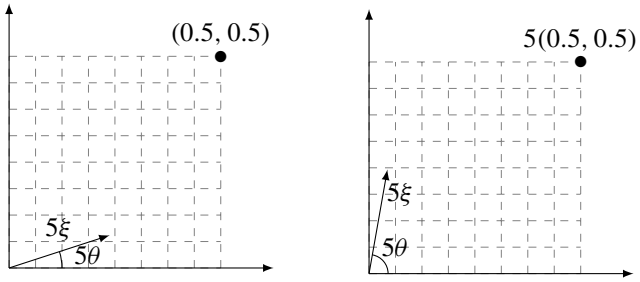

Figure 3. Representation of $\xi \in \mathbb{R}^2$ with $\theta \in \left]0, \frac{\pi}{4}\right]$ and $\theta \in \left]\frac{\pi}{4}, \frac{\pi}{2}\right]$ with $\cos\theta|\xi| = \xi \cdot e_1$.

Thus, we deduce from (2-3) that

$$\left|\frac{1}{(N+1)^D}\sum_{|\xi\cdot v_k|\geq\delta|\xi|}\hat{f}(\xi, v_k)\psi(v_k)\right| \leq CG(\xi)\frac{1}{|\xi|\sqrt{\delta}}\left(1+\frac{1}{\delta N}\right)^{1/2}.$$

We conclude that

$$\left|\hat{\rho}[\psi](\xi)\right| \leq C\left(F(\xi)\sqrt{\delta+\frac{1}{N}} + G(\xi)\frac{1}{|\xi|\sqrt{\delta}}\left(1+\frac{1}{\delta N}\right)^{1/2}\right) \tag{2-4}$$

holds when ξ is aligned with the axis.

We turn to the general case (ii). As illustrated in Figure 3, we can assume that the angle θ between ξ and one of the axes (say e_1) lies in $\left]0, \frac{\pi}{4}\right[$, the other cases follow by a symmetry argument.

The reasoning still consists in counting velocities in strips appropriately defined. As said above, without loss of generality we can assume that $\theta \in \left]0, \frac{\pi}{4}\right[$, where we have set $\cos\theta|\xi| = \xi \cdot e_1$. We set $\ell_1 := \delta/\cos\theta$. On a given strip, we can find at most $(\lfloor \ell_1 N \rfloor + 1) \times (N+1)^{D-1}$ velocities; see Figure 5.

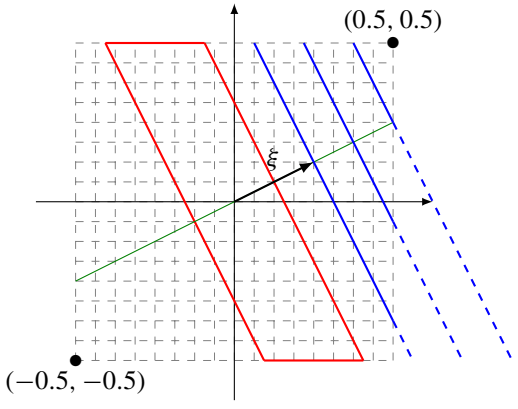

Figure 4. The area corresponding to $|\xi \cdot v_k| \leq \delta|\xi|$ is delimited as previously. The complementary set is split into strips of width δ.

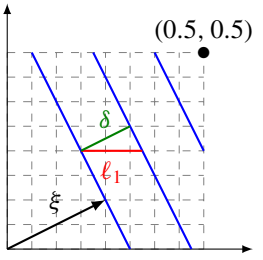

Figure 5. Representation of the parameter ℓ_1.

Therefore, bearing in mind that $0 < \theta < \frac{\pi}{4}$, we obtain

$$\sum_{|\xi\cdot v_k|\geq\delta|\xi|}\frac{1}{|\xi\cdot v_k|^2}=\sum_{|\xi\cdot v_k|\geq\delta|\xi|}\frac{1}{|\xi|^2}\,\frac{1}{|\xi/|\xi|\cdot v_k|^2}\leq\frac{1}{|\xi|^2}\,2\sum_{p\geq1}\frac{1}{(p\delta)^2}\left(\frac{\delta}{\cos\theta}N+1\right)(N+1)^{D-1}$$
$$\leq\frac{1}{|\xi|^2}\,2\sum_{p\geq1}\frac{1}{(p\delta)^2}\,\frac{1}{\delta\cos\theta}\left(1+\frac{1}{\delta N}\right)(N+1)^D\leq 2\sqrt{2}\,\frac{1}{|\xi|^2}\,\frac{1}{\delta}\left(1+\frac{1}{\delta N}\right)(N+1)^D$$

and

$$\sum_{|\xi\cdot v_k|<\delta|\xi|}1=(2\lfloor\ell_1 N\rfloor+1)(N+1)^{D-1}\leq 2\left(\frac{\delta}{\cos\theta}N+1\right)(N+1)^{D-1}\leq 2\sqrt{2}\left(\delta+\frac{1}{N}\right)(N+1)^D.$$

Thus, we deduce exactly like in case (i) that (2-4) holds for any $\xi\neq 0$.

Therefore, we have established that for all $\xi\neq 0$, we get (2-4) for all $\delta>0$. We take

$$\delta=\frac{1}{|\xi|}\mathbf{1}_{\{N\geq|\xi|\}}+\frac{1}{N}\mathbf{1}_{\{N<|\xi|\}}$$

and we define

$$\Theta_N(\xi):=\hat{\rho}[\psi](\xi)\mathbf{1}_{\{N\geq|\xi|\}},\quad \Delta_N(\xi):=\hat{\rho}[\psi](\xi)\mathbf{1}_{\{N<|\xi|\}}.$$

Then, we have

$$\Theta_N(\xi)\leq C\left(F(\xi)\sqrt{\frac{1}{|\xi|}+\frac{1}{N}}+G(\xi)\frac{1}{|\xi|\sqrt{1/|\xi|}}\left(1+\frac{1}{N/|\xi|}\right)^{1/2}\right)\mathbf{1}_{\{N\geq|\xi|\}}\leq C(F(\xi)+G(\xi))\frac{1}{\sqrt{|\xi|}}.$$

It implies that

$$|\xi|\Theta_N(\xi)^2\leq C(G^2(\xi)+F^2(\xi)),$$

which equally holds true for $\xi=0$. Then by the assumption on f and g, we deduce that $\Theta_N\in H^{1/2}(\mathbb{R}^D)$. Finally, we evaluate the remainder:

$$\Delta_N(\xi)\leq C\left(F(\xi)\sqrt{\frac{2}{N}}+G(\xi)\frac{1}{|\xi|\sqrt{1/N}}\left(1+\frac{1}{(1/N)N}\right)\right)\mathbf{1}_{\{N<|\xi|\}}\leq\frac{C}{\sqrt{N}}(F(\xi)+G(\xi)).$$

We conclude that

$$\Delta_N^2(\xi)\leq\frac{C}{N}(F^2(\xi)+G^2(\xi)),$$

which is also satisfied when $\xi=0$. Thus, by the assumption on f and g, we know $\|\Delta_N\|_{L^2}$ is dominated by $1/\sqrt{N}$, an observation which finishes the proof. □

A stochastic discrete velocity averaging lemma. Dealing with random discrete velocities we can expect to make the defect vanish when taking the expectation of the velocity averages. This is indeed the case as shown in the following statement.

Theorem 2.3. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let $V_1, \dots, V_{\mathscr{N}}$ be i.i.d. random variables, distributed according to the continuous uniform distribution on $[-0.5, 0.5]^D$. We set*

$$\mathrm{d}\mu = \frac{1}{\mathscr{N}} \sum_{k=1}^{\mathscr{N}} \delta(v = V_k).$$

Let $f, g \in L^2(\mathbb{R}^D \times \mathbb{R}^D \times \Omega, \mathrm{d}x\, \mathrm{d}\mu(v)\, \mathrm{d}\mathbb{P})$ satisfy, for all $x \in \mathbb{R}^D$, $\omega \in \Omega$, and $k \in \{1, \dots, \mathscr{N}\}$,

$$V_k \cdot \nabla_x f(x, V_k) = g(x, V_k). \tag{2-5}$$

Then, for all $\psi \in C_c^\infty(\mathbb{R}^D)$, the macroscopic quantity

$$\rho[\psi](x) := \frac{1}{\mathscr{N}} \sum_{k=1}^{\mathscr{N}} f(x, V_k)\psi(V_k) = \mathbb{R}^D f(x, v)\psi(v)\, \mathrm{d}\mu(v)$$

satisfies $\mathbb{E}\rho[\psi] \in H^{1/2}(\mathbb{R}^D)$ (and it is bounded in this space if the L^2 norm of f and g is bounded uniformly with respect to $\mathscr{N}$).

Remark 2.4. We point out that this statement has a different nature from the stochastic averaging lemma devised in [Debussche et al. 2015; 2016], where the velocity set still satisfies an assumption like (1-3) but the equation for $v \cdot \nabla_x f_n$ involves a stochastic term. Our analysis is closer in spirit to the results in [Lions et al. 2013], where the velocity variable is deterministic but is multiplied by a Brownian motion.

Proof. We apply the Fourier transform to (2-5). Then, for all k, we get

$$\xi \cdot V_k \hat{f}(\xi, V_k) = (-i)\hat{g}(\xi, V_k).$$

We set

$$F(\xi) := \left(\frac{1}{\mathscr{N}} \mathbb{E} \sum_k |\hat{f}(\xi, V_k)|^2\right)^{1/2}, \quad G(\xi) := \left(\frac{1}{\mathscr{N}} \mathbb{E} \sum_k |\hat{g}(\xi, V_k)|^2\right)^{1/2}.$$

Let us split

$$\begin{aligned}\mathbb{E}\hat{\rho}[\psi](\xi) &= \mathbb{E}\left[\frac{1}{\mathscr{N}} \sum_k \hat{f}(\xi, V_k)\psi(V_k)\right] \\ &= \mathbb{E}\left[\frac{1}{\mathscr{N}} \sum_{|\xi \cdot V_k| < \delta|\xi|} \hat{f}(\xi, V_k)\psi(V_k)\right] + \mathbb{E}\left[\frac{1}{\mathscr{N}} \sum_{|\xi \cdot V_k| \geq \delta|\xi|} \hat{f}(\xi, V_k)\psi(V_k)\right]\end{aligned}$$

for $\delta > 0$. The Cauchy–Schwarz inequality leads to the following estimates: on the one hand,

$$\left|\mathbb{E}\left[\frac{1}{\mathscr{N}} \sum_{|\xi \cdot V_k| < \delta|\xi|} \hat{f}(\xi, V_k)\psi(V_k)\right]\right| \leq \|\psi\|_\infty \left(\frac{1}{\mathscr{N}} \mathbb{E} \sum_k |\hat{f}(\xi, V_k)|^2\right)^{1/2} \left(\frac{1}{\mathscr{N}} \mathbb{E} \sum_{|\xi \cdot V_k| < \delta|\xi|} 1\right)^{1/2},$$

and, on the other hand,

$$\left|\mathbb{E}\left[\frac{1}{\mathscr{N}}\sum_{|\xi\cdot v_k|\geq\delta|\xi|}\hat{f}(\xi,V_k)\psi(V_k)\right]\right| = \left|\mathbb{E}\left[\frac{1}{\mathscr{N}}\sum_{|\xi\cdot V_k|\geq\delta|\xi|}\frac{(-i)\hat{g}(\xi,V_k)}{\xi\cdot V_k}\,\psi(V_k)\right]\right|$$
$$\leq \|\psi\|_\infty\left(\frac{1}{\mathscr{N}}\,\mathbb{E}\sum_k|\hat{g}(\xi,V_k)|^2\right)^{1/2}\left(\frac{1}{\mathscr{N}}\,\mathbb{E}\sum_{|\xi\cdot v_k|\geq\delta|\xi|}\frac{1}{|\xi\cdot V_k|^2}\right)^{1/2}.$$

We only detail the case where $\xi=\alpha e_1,\ \alpha\in\mathbb{R}$, the other cases being deduced by adapting the reasoning of the proof of Proposition 2.1. We have

$$\mathbb{E}\left[\sum_{|\xi\cdot V_k|\geq\delta|\xi|}\frac{1}{|\xi\cdot V_k|^2}\right]=\mathbb{E}\left[\sum_{|\xi\cdot V_k|\geq\delta|\xi|}\frac{1}{|\xi|^2}\frac{1}{|\xi/|\xi|\cdot V_k|^2}\right]\leq\mathbb{E}\left[\frac{1}{|\xi|^2}\,2\left(\sum_{p\geq1}\frac{1}{(p\delta)^2}\right)M_p\right],$$

where M_p is the number of velocities in the p-th strip (see Figure 2). We bear in mind that M_p is a random variable: since the V_i are distributed according to the uniform law, we have

$$\mathbb{P}(V_i\in S_p)=\delta$$

and, since the variables $V_1,\dots,V_{\mathscr{N}}$ are independent, M_p follows a binomial distribution of parameters $\mathscr{N}$ and δ. Therefore, we are led to

$$\mathbb{E}\left[\sum_{|\xi\cdot V_k|\geq\delta|\xi|}\frac{1}{|\xi\cdot V_k|^2}\right]\leq\frac{1}{|\xi|^2}\,2\left(\sum_{p\geq1}\frac{1}{(p\delta)^2}\right)\mathbb{E}[M_p]\leq C\frac{1}{|\xi|^2\delta}\mathscr{N},\tag{2-6}$$

which yields

$$\left|\mathbb{E}\left[\frac{1}{\mathscr{N}}\sum_{|\xi\cdot V_k|\geq\delta|\xi|}\hat{f}(\xi,V_k)\psi(V_k)\right]\right|\leq CG(\xi)\frac{1}{|\xi|\sqrt{\delta}}.$$

By the same token, we get

$$\mathbb{E}\left[\sum_{|\xi\cdot V_k|<\delta|\xi|}1\right]=2\delta\mathscr{N}\tag{2-7}$$

so that

$$\left|\mathbb{E}\left[\frac{1}{\mathscr{N}}\sum_{|\xi\cdot V_k|<\delta|\xi|}\hat{f}(\xi,V_k)\psi(V_k)\right]\right|\leq CF(\xi)\sqrt{\delta}.$$

Finally, we arrive at

$$\left|\mathbb{E}\hat{\rho}[\psi](\xi)\right|\leq C\left(F(\xi)\sqrt{\delta}+\frac{G(\xi)}{|\xi|\sqrt{\delta}}\right).$$

We apply this inequality with $\delta=G(\xi)/(|\xi|F(\xi))$, which leads to

$$\left|\mathbb{E}\hat{\rho}[\psi](\xi)\right|\leq C\sqrt{F(\xi)G(\xi)}\frac{1}{\sqrt{|\xi|}}.$$

This concludes the proof by using the assumptions on f and g. □

Remark 2.5. We can readily extend the result to nonuniform laws: we assume that the V_i are identically and independently distributed in $\mathbb{R}^D$ according to a continuous and bounded density of probability Φ. The number M_p of velocities in the strip S_p still follows a binomial law but now the expectation value depends on Φ and M_p can be shown to be dominated by $\mathscr{N}\|\Phi\|_\infty\delta$.

For certain applications, the variable v lies on the sphere. This is the case for the kinetic models arising in radiative transfer theory, where v represents the *direction* of flight of photons, which, of course, all travel with the speed of light. We can adapt the stochastic averaging lemma to this situation.

Theorem 2.6. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let $V_1, \dots, V_{\mathscr{N}}$ be i.i.d. random variables, distributed according to the continuous uniform distribution on $\mathbb{S}^{D-1}$. We set*

$$\mathrm{d}\mu = \frac{1}{\mathscr{N}} \sum_{k=1}^{\mathscr{N}} \delta(v = V_k).$$

Let $f, g \in L^2(\mathbb{R}^D \times \mathbb{R}^D \times \Omega, \mathrm{d}x\, \mathrm{d}\mu(v)\, \mathrm{d}\mathbb{P})$ satisfy, for all $x \in \mathbb{R}^D$, $\omega \in \Omega$, and $k \in \{1, \dots, \mathscr{N}\}$,

$$V_k \cdot \nabla_x f(x, V_k) = g(x, V_k).$$

Then, for all $\psi \in C_c^\infty(\mathbb{S}^{D-1})$, the macroscopic quantity

$$\rho[\psi](x) := \frac{1}{\mathscr{N}} \sum_{k=1}^{\mathscr{N}} f(x, V_k)\psi(V_k) = \mathbb{R}^D f(x, v)\psi(v)\, \mathrm{d}\mu(v)$$

satisfies $\mathbb{E}\rho[\psi] \in H^{1/2}(\mathbb{R}^D)$.

Proof. The proof follows the same arguments as those for Theorem 2.3; we only indicate the main changes. The proof still relies on counting the velocities produced by the random sampling in the domain

$$S_p = \left\{ v \in \mathbb{S}^{D-1} : \delta p|\xi| \le |v \cdot \xi| \le \delta(p+1)|\xi| \right\}$$

for given $\xi \in \mathbb{R}^D \backslash \{0\}$, $\delta > 0$ and $p \in \mathbb{Z}$. We define $\theta \in [0, 2\pi]$ such that

$$v \cdot \xi |\xi| = \cos\theta \in [-1, +1].$$

Considering the random vectors V_k, the associated variable θ_k is randomly distributed on $[0, 2\pi]$. For symmetry reasons, $\mathbb{P}(V_k \in S_p)$ is thus proportional to

$$\mathbb{P}\big(\delta|p| \le \cos\theta_k \le \delta(|p|+1)\big).$$

We start with the specific case of dimension $D = 2$, and we refer the reader to Figure 6. In this case, θ is uniformly distributed on $[0, 2\pi]$. Therefore, for any $p \in \mathbb{N}$, we know $\mathbb{P}(\delta p \le \cos\theta \le \delta(p+1))$ is proportional to

$$\Pi_{\delta,p} = \arccos(\delta(p+1))^{\arccos(\delta p)}\, \mathrm{d}\theta = \arccos(\delta p) - \arccos(\delta(p+1))$$

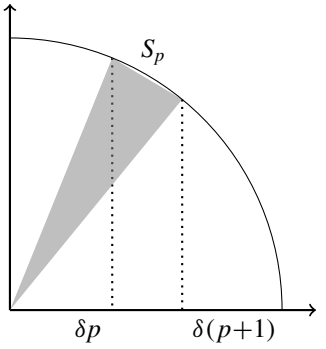

Figure 6. Velocities on the sphere $\mathbb{S}^1$, with domain S_p.

and $M_p = \#\{V_k \in S_p\}$ is driven by the binomial law with parameters $\mathscr{N}$ and $\alpha\Pi_{\delta,p}$ for a certain constant $\alpha > 0$. Hence, the analog of (2-7) is dominated, up to some constant, by

$$\mathscr{N}\Pi_{\delta,0} = \mathscr{N}\left(\tfrac{1}{2}\pi - \arccos\delta\right) = \mathscr{N}\delta\frac{\mathrm{d}x}{\sqrt{1-x^2}} \le C\mathscr{N}\delta$$

as far as $0 < \delta \le \delta_0 < 1$. Similarly, the analog of (2-6) involves the sum

$$\sum_{p\ge 1}\frac{\mathscr{N}}{\delta^2 p^2}\Pi_{\delta,p},$$

which we split into

$$\mathrm{I} = \sum_{1\le p\le 1/2\delta}\frac{\mathscr{N}}{\delta^2 p^2}\Pi_{\delta,p}, \quad \mathrm{II} = \sum_{1/2\delta< p\le 1/\delta}\frac{\mathscr{N}}{\delta^2 p^2}\Pi_{\delta,p}.$$

For I, we can still use the fact that $x \mapsto 1/\sqrt{1-x^2}$ is nonincreasing and bounded far away from $x = 1$ and we are led to the estimate

$$\mathrm{I} = \sum_{1\le p\le 1/2\delta}\frac{\mathscr{N}}{\delta^2 p^2}\int_{\delta(p+1)}^{\delta p}\frac{\mathrm{d}x}{\sqrt{1-x^2}} \le \sum_{1\le p\le 1/2\delta}\frac{\mathscr{N}}{\delta^2 p^2}\frac{\delta}{\sqrt{1-\delta^2(p+1)^2}} \le C\frac{\mathscr{N}}{\delta}.$$

For II, we use a summation by parts which yields

$$\begin{aligned}\mathrm{II} &= \sum_{1/2\delta< p\le 1/\delta}\frac{\mathscr{N}\arccos(\delta p)}{\delta^2}\left(\frac{1}{(p-1)^2} - \frac{1}{p^2}\right)\\ &\le \sum_{1/2\delta< p\le 1/\delta}\frac{\mathscr{N}\arccos(\delta p)}{\delta^2}\frac{2}{p(p-1)^2} \le \frac{4\delta}{\delta^2}\pi\mathscr{N}\sum_{p\ge 1}\frac{1}{p^2} \le C\frac{\mathscr{N}}{\delta}.\end{aligned}$$

Having these estimates at hand, we can repeat the same arguments as in the proof of Theorem 2.3.

For higher dimensions, the situation is actually simpler since θ is now distributed on $\left[0, \frac{\pi}{2}\right]$ according to the law with density $(\sin\theta)^{D-2}\,\mathrm{d}\theta$. Thus (with the simple estimate $0 \le (\sin\theta)^{D-2} \le \sin\theta$) we obtain directly the analog of estimates (2-6) and (2-7). □

The result can be extended to the L^p cases for $1 < p < \infty$ by using an interpolation argument as in [Golse et al. 1988, Theorem 2].

Corollary 2.7. *In Theorems 2.3 and 2.6, we assume that f and g belong to $L^p(\mathbb{R}^D \times \mathscr{V} \times \Omega, \mathrm{d}x\, \mathrm{d}\mu(v)\, \mathrm{d}\mathbb{P})$ for some $1 < p < \infty$, with $\mathscr{V}$ either $\mathbb{R}^D$ or $\mathbb{S}^{D-1}$. Then $\mathbb{E}\rho[\psi]$ lies in the Sobolev space $W^{s,p}(\mathbb{R}^D)$ with $0 < s < \min(1/p, 1 - 1/p) < 1$.*

Proof. We readily adapt the interpolation argument in [Golse et al. 1988]. Let $\mathscr{T}$ be the operator

$$\mathscr{T} : h \mapsto \mathbb{E} \int f(x, v)\psi(v)\, \mathrm{d}\mu(v),$$

where

$$f(x, V_k) + V_k \cdot \nabla_x f(x, V_k) = h(x, V_k).$$

Clearly $\mathscr{T}$ maps continuously $L^r(\mathbb{R}^D \times \mathscr{V} \times \Omega, \mathrm{d}x\, \mathrm{d}\mu(v)\, \mathrm{d}\mathbb{P})$ into $L^r(\mathbb{R}^D)$ for any $1 < r < \infty$. Moreover, Theorems 2.3 and 2.6 tell us that $\mathscr{T}$ is a continuous operator from $L^2(\mathbb{R}^D \times \mathscr{V} \times \Omega, \mathrm{d}x\, \mathrm{d}\mu(v)\, \mathrm{d}\mathbb{P})$ to $H^{1/2}(\mathbb{R}^D)$. We conclude by interpreting the Sobolev space $W^{s,p}$ by interpolation, as being an intermediate space between $L^r = W^{0,r}$ and $H^{1/2} = W^{1/2,2}$ [Bergh and Löfström 1976, Theorem 6.4.5, relation (7)], and L^p as being interpolated between L^r and L^2. □

We can equally extend the compactness statement to the L^1 framework by following [Golse and Saint-Raymond 2002].

Corollary 2.8. *We consider a random set of velocities defined as in Theorem 2.3 orTheorem 2.6. Let $(f_n)_{n\in\mathbb{N}}$ and $(g_n)_{n\in\mathbb{N}}$ be two sequences of functions defined on $\mathbb{R}^D \times \mathscr{V} \times \Omega$ such that*

(i) *$\{f_n : n \in \mathbb{N}\}$ is a relatively weakly compact set in $L^1(\mathbb{R}^D \times \mathscr{V} \times \Omega, \mathrm{d}x\, \mathrm{d}\mu(v)\, \mathrm{d}\mathbb{P})$,*

(ii) *$\{g_n : n \in \mathbb{N}\}$ is bounded in $L^1(\mathbb{R}^D \times \mathscr{V} \times \Omega, \mathrm{d}x\, \mathrm{d}\mu(v)\, \mathrm{d}\mathbb{P})$,*

(iii) *we have $V_k \cdot \nabla_x f_n(x, V_k) = g_n(x, V_k)$.*

Then $\mathbb{E}\rho_n[\psi](x) = \mathbb{E} \int f_n(x, v)\psi(v)\, \mathrm{d}\mu(v)$ lies in a relatively compact set of $L^1(B(0, R))$ for any $0 < R < \infty$ (for the strong topology).

Proof. The proof follows closely [Golse and Saint-Raymond 2002]; we sketch the arguments for the sake of completeness. For $\psi \in C_c^\infty(\mathscr{V})$, we denote by $\mathscr{A}$ the operator

$$\mathscr{A} : f \mapsto \mathbb{E} \int f(x, v)\psi(v)\, \mathrm{d}\mu(v).$$

For $\lambda > 0$, we also introduce the operator

$$R_\lambda : h \mapsto \int_0^\infty e^{-\lambda t} h(x - vt, v)\, \mathrm{d}t,$$

which returns the solution $f = R_\lambda h$ of $(\lambda + v \cdot \nabla_x) f = h$. It is a continuous operator on $L^p(\mathbb{R}^D \times \mathscr{V}, \mathrm{d}x\, \mathrm{d}\mu(v))$ spaces and we have

$$\|R_\lambda h\|_{L^p} \le \frac{\|h\|_{L^p}}{\lambda}. \tag{2-8}$$

Let us temporarily assume that the compactness statement holds for $\mathscr{A}R_\lambda g_n$, for any $\lambda > 0$, when (i)–(ii) is strengthened to

(ii′) $\{g_n : n \in \mathbb{N}\}$ is a relatively weakly compact set in $L^1(\mathbb{R}^D \times \mathscr{V} \times \Omega, \mathrm{d}x\,\mathrm{d}\mu(v)\,\mathrm{d}\mathbb{P})$.

Therefore, writing $(\lambda + v \cdot \nabla_x)R_\lambda f_n = f_n$, we deduce from (i) that $(\mathscr{A}R_\lambda f_n)_{n\in\mathbb{N}}$ is relatively compact in $L^1(B(0, R))$ for any $\lambda > 0$ and $0 < R < \infty$. Next, we write $f_n = \lambda R_\lambda f_n + R_\lambda(v \cdot \nabla_x f_n)$ so that, owing to (2-8), $\mathscr{A}f_n = \lambda\mathscr{A}R_\lambda f_n + \mathscr{A}R_\lambda(v \cdot \nabla_x f_n)$ appears as the sum of a sequence which is compact in $L^1(B(0, R))$ and a sequence whose norm is dominated by $1/\lambda$, uniformly with respect to n. Consequently, $(\mathscr{A}f_n)_{b\in\mathbb{N}}$ is relatively compact in $L^1(B(0, R))$.

We are thus left with the task of justifying the gain of compactness for $\mathscr{A}R_\lambda g_n$ when (i)–(ii) is replaced by (ii′); see [Golse et al. 1988, Proposition 3]. To this end, for $\lambda, M > 0$ we set $R_\lambda g_n = \gamma_n$ and we split

$$\gamma_n = \gamma_{n,M} + \gamma_n^M,$$

where

$$(\lambda + V_k \cdot \nabla_x)\gamma_{n,M}(x, V_k) = g_n(x, V_k)\mathbf{1}_{g_n(x,V_k)\le M},$$
$$(\lambda + V_k \cdot \nabla_x)\gamma_n^M(x, V_k) = g_n(x, V_k)\mathbf{1}_{g_n(x,V_k)>M}.$$

Since for any fixed $M > 0$, the set $\{g_n\mathbf{1}_{h_n\le M} : n \in \mathbb{N}\}$ is bounded in $L^1 \cap L^\infty \subset L^2$, we can apply Theorem 2.3 or Theorem 2.6, which imply that $(\mathscr{A}\gamma_{n,M})_{n\in\mathbb{N}}$ is compact in $L^1(B(0, R))$ for any finite R. We can conclude by showing that γ_n^M can be made arbitrarily small, in L^1 norm, uniformly with respect to $n \in \mathbb{N}$, for a suitable choice of $M > 0$. This is indeed the case because (ii′) implies

$$\lim_{M\to\infty}\left\{\sup_n \int |g_n|\mathbf{1}_{g_n>M}\,\mathrm{d}\mu(v)\,\mathrm{d}x\,\mathrm{d}\mathbb{P}(\omega)\right\} = 0$$

by virtue of the Dunford–Pettis theorem; see [Goudon 2011, §7.3.2]. Going back to (2-8) finishes the proof. □

3. Application to the Rosseland approximation

Let us go back to the asymptotic behavior of the solutions of (1-1). The problem (1-1) is completed with the initial condition

$$f_\varepsilon|_{t=0} = f_\varepsilon^0.$$

It satisfies $f_\varepsilon^0 \ge 0$ and $f_\varepsilon^0 \in L^1(\mathbb{R}^D \times \mathscr{V})$, as it is physically relevant, f_ε being a particle density. For the set $(\mathscr{V}, \mathrm{d}\mu)$, in what follows we suppose at least that $\mathscr{V}$ is a bounded subset in $\mathbb{R}^D$ and

$$\int_{\mathscr{V}} \mathrm{d}\mu(v) = 1, \quad \int_{\mathscr{V}} v\,\mathrm{d}\mu(v) = 0.$$

These assumptions are crucial for the analysis of the diffusion regime. Then, the connection to (1-2) can be established as follows.

Theorem 3.1. *We assume that* (1-3) *is fulfilled. Let σ be a function such that $\sigma(\rho) = \rho^\gamma \Sigma(\rho)$ with $|\gamma| < 1$ and $0 < \sigma_* \le \Sigma(\rho) \le \sigma^* < \infty$. Let $(f_\varepsilon^0)_{\varepsilon>0}$ satisfy*

$$\sup_{\varepsilon>0}\left(\int_{\mathbb{R}^d}\int_{\mathscr{V}} \left(1+\varphi(x)+|\ln f_\varepsilon^0| f_\varepsilon^0\right) \mathrm{d}\mu(v)\,\mathrm{d}x + \|f_\varepsilon^0\|_{L^\infty(\mathbb{R}^d\times\mathscr{V})}\right) = M_0 < +\infty$$

for a certain weight function such that $\lim_{|x|\to+\infty}\varphi(x) = +\infty$. Then (up to a subsequence) the solution f_ε of (1-1) *and ρ_ε converge to $\rho(t,x)$ in $L^p((0,T)\times\mathbb{R}^d\times\mathscr{V})$ and $L^p((0,T)\times\mathbb{R}^d)$ respectively, for any $1 \le p < \infty$, $0 < T < \infty$, where ρ is a solution to* (1-2) *with the initial data $\rho|_{t=0}$ given by the weak limit in $L^p(\mathbb{R}^d)$ of $\int_{\mathscr{V}} f_\varepsilon^0\,\mathrm{d}\mu(v)$ as $\varepsilon \to 0$.*

For instance this statement holds with $\mathscr{V} = \mathbb{S}^{D-1}$ endowed with the Lebesgue measure. We refer the reader to [Bardos et al. 1988] for a detailed proof, where the velocity averaging lemma is used to manage the passage to the limit in the nonlinearity. Assumption (1-3) can be replaced by

$$\text{for any } \xi \neq 0, \quad \mathrm{meas}\big(\{v \in \mathscr{V} \cap B(0,R) : v\cdot\xi \neq 0\}\big) > 0,$$

which allows us to deal with certain discrete velocity models. Then, the asymptotic regime can be analyzed with a *compensated compactness* argument, which relies on the structure of the system satisfied by the zeroth and first moments of f_ε, as pointed out in [Degond et al. 2000; Goudon and Poupaud 2001; Lions and Toscani 1997]; see also [Marcati and Milani 1990]. The question of the relation between the diffusion equation that corresponds to a discretization of the velocity set (discrete ordinate equation) and the diffusion equation that corresponds to the continuous model can be addressed. For the simple collision operator in (1-1), velocity grids, which differ from the simplest uniform mesh, can be constructed that lead to the *exact* diffusion coefficient, namely

$$\frac{1}{\mathscr{N}}\sum_{k=1}^{\mathscr{N}} v_k \otimes v_k = \int_{\mathbb{S}^{D-1}} v\otimes v\,\mathrm{d}v = \frac{1}{D}\mathbb{I};$$

we refer the reader to [Buet et al. 2002; Golse et al. 1999; Jin and Levermore 1991] for further discussion on this issue. However, for more general collision operators, it might happen that the equilibrium functions that make the collision operator vanish or the diffusion coefficient are not explicitly known; see [Bonnaillie-Noël et al. 2016; Degond et al. 2000].

We wish to revisit this question by means of a Monte Carlo approach: instead of the discrete ordinate viewpoint where a discrete velocity grid is adopted once and for all, we deal with a random set of velocities and we wonder whether it can provide, in expectation, a consistent approximation of the diffusion regime. The consistency analysis we propose uses Theorem 2.3 or Theorem 2.6 to justify the following claim.

Theorem 3.2. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let $V_1, \dots, V_{\mathscr{N}}$ be i.i.d. random variables distributed according to the continuous uniform law on $\mathscr{V}$. Then, we obtain a set $\mathscr{V}_{\mathscr{N}}$ of $2\mathscr{N}$ velocities in $\mathscr{V}$ by setting $V_{\mathscr{N}+j} = -V_j$ for all $j \in \{1,\dots,\mathscr{N}\}$. We denote the associated discrete measure on $\mathscr{V}$ by*

$$\mathrm{d}\mu_{\mathscr{N}}(v) = \frac{1}{2\mathscr{N}}\sum_{k=1}^{2\mathscr{N}} \delta(v = V_k).$$

Let $f_\varepsilon|_{t=0} = f_\varepsilon^0 \geq 0$ *satisfy*

$$\sup_{\varepsilon>0,\ \mathscr{N}\in\mathbb{N}} \left(\mathbb{E} \int_{\mathbb{R}^D} \int_{\mathscr{V}} \big(1+\varphi(x)+|\ln f_\varepsilon^0|\big) f_\varepsilon^0 \, \mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x + \|f_\varepsilon^0\|_{L^\infty(\Omega\times\mathbb{R}^d\times\mathscr{V})} \right) = M_0 < +\infty. \tag{3-1}$$

Let f_ε *be a solution of the equation*

$$\partial_t f_\varepsilon(t,x,V_j) + \frac{1}{\varepsilon} V_j \cdot \nabla_x f_\varepsilon(t,x,V_j) = \frac{1}{\varepsilon^2}\sigma(\rho_{\varepsilon,\mathscr{N}})[\rho_{\varepsilon,\mathscr{N}}(t,x) - f_\varepsilon(t,x,V_j)], \tag{3-2}$$

with

$$\rho_{\varepsilon,\mathscr{N}}(t,x) := \frac{1}{2\mathscr{N}} \sum_{i=1}^{2\mathscr{N}} f_\varepsilon(t,x,V_j).$$

We suppose that $\rho \in [0,\infty) \mapsto \sigma(\rho)$ *is a nonnegative function such that for any* $0 < R < \infty$, *there exists* $\sigma_\star(R) > 0$ *satisfying* $0 < 1/\sigma_\star(R) \leq \sigma(\rho) \leq \sigma_\star(R)$ *and* $|\sigma'(\rho)| \leq \sigma_\star(R)$ *for any* $0 \leq \rho \leq R$. *Then* $\mathbb{E}\rho_{\varepsilon,\mathscr{N}}$ *converges to* $\mathbb{E}\rho_{\mathscr{N}}$ *in* $L^2((0,T)\times\mathbb{R}^D)$ *as* ε *goes to* 0 *with* $0 < T < \infty$, *where* $\mathbb{E}\rho_{\mathscr{N}}$ *is solution of*

$$\partial_t \mathbb{E}\rho_{\mathscr{N}} + \operatorname{div}(\mathscr{J}_{\mathscr{N}}) = 0, \quad \sigma(\mathbb{E}\rho_{\mathscr{N}})\,\mathscr{J}_{\mathscr{N}} = -\mathbb{E}A_{\mathscr{N}} \nabla_x \mathbb{E}\rho_{\mathscr{N}} + O\left(\frac{1}{\sqrt{\mathscr{N}}}\right),$$

with $A_{\mathscr{N}}$ *the* $D \times D$ *matrix with random components defined by*

$$A_{\mathscr{N}} := \frac{1}{2\mathscr{N}} \sum_{j=1}^{2\mathscr{N}} V_j \otimes V_j,$$

and $\mathbb{E}\rho_{\mathscr{N}}|_{t=0}$ *is the weak limit of* $\int \mathbb{E} f_\varepsilon^0 \, \mathrm{d}\mu(v)$.

Note that the construction of the set $\mathscr{V}_{\mathscr{N}}$ ensures that the null flux condition $\int v \, \mathrm{d}\mu_{\mathscr{N}}(v) = 0$ is fulfilled, but the elements of $\mathscr{V}_{\mathscr{N}}$ are not independent. Nevertheless, the stochastic averaging lemma still applies to this situation, with a straightforward adaptation of the proof. It is likely that the assumptions on σ can be substantially weakened, but it not our aim here to seek refinements in this direction. We will make precise in the proof in which sense the consistency error $O(1/\sqrt{\mathscr{N}})$ should be understood.

Entropy estimates. In order to prove Theorem 3.2, the first step consists in establishing some a priori estimates, uniform with respect to the parameters ε and $\mathscr{N}$. We will then deduce the compactness needed to obtain the result. These estimates are quite classical; the proof that we sketch for the sake of completeness follows directly from [Bardos et al. 1988; Goudon and Poupaud 2001; Lions and Toscani 1997].

Proposition 3.3. *Let* f_ε^0 *satisfy* (3-1) *with* $\varphi(x) = (1+x^2)^\beta$, $0 < \beta < 1$. *Let* $0 < T < \infty$. *There exists a constant* $C(T)$ *which only depends on* T *such that*

$$\sup_{\varepsilon>0,\mathscr{N}\in\mathbb{N}} \left\{ \sup_{0\leq t\leq T} \mathbb{E} \int_{\mathbb{R}^D} \int_{\mathscr{V}} \big(1+\varphi(x)+|\ln f_\varepsilon|\big) f_\varepsilon \, \mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x + \|f_\varepsilon\|_{L^\infty(\Omega\times(0,T)\times\mathbb{R}^D\times\mathscr{V})} \right\} = C(T) < +\infty \tag{3-3}$$

and, furthermore,

$$\sup_{\varepsilon>0,\mathscr{N}\in\mathbb{N}} \mathbb{E} \int_0^T \int_{\mathbb{R}^D} \int_{\mathscr{V}} \frac{\sigma(\rho_{\varepsilon,\mathscr{N}})}{\varepsilon^2} (f_\varepsilon - \rho_{\varepsilon,\mathscr{N}}) \ln\left(\frac{f_\varepsilon}{\rho_{\varepsilon,\mathscr{N}}}\right) \mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x\,\mathrm{d}t \leq C(T). \tag{3-4}$$

Proof. As said above we crucially use the fact that

$$\int_{\mathscr{V}} \mathrm{d}\mu_{\mathscr{N}}(v) = 1, \quad \int_{\mathscr{V}} v \,\mathrm{d}\mu_{\mathscr{N}}(v) = 0.$$

As a matter of fact, the collision operator is mass-conserving in the sense that

$$\int_{\mathscr{V}} \sigma(\rho)(f-\rho)\,\mathrm{d}\mu_{\mathscr{N}}(v) = 0.$$

Accordingly, integrating immediately leads to

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}} f_\varepsilon \,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x = 0. \tag{3-5}$$

More generally, let $G : [0,\infty) \to \mathbb{R}$ be a convex function. We get

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}} G(f_\varepsilon)\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x = -\frac{1}{\varepsilon^2}\,\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}} \sigma(\rho_{\varepsilon,\mathscr{N}})(\rho_{\varepsilon,\mathscr{N}} - f_\varepsilon)(G'(\rho_{\varepsilon,\mathscr{N}}) - G'(f_\varepsilon))\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x \le 0.$$

With $G(z) = z^p$, $p \ge 1$, it gives an estimate on the L^p norm of the solution. Similarly, with $G(z) = [z - \|f_\varepsilon^0\|_\infty]_+^2$, we conclude that

$$\|f_\varepsilon\|_{L^\infty(\Omega\times(0,T)\times\mathbb{R}^D\times\mathscr{V})} \le \|f_\varepsilon^0\|_\infty.$$

Finally, with $G(z) = z\ln(z)$ we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}} f_\varepsilon \ln f_\varepsilon \,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x = -\frac{1}{\varepsilon^2}\,\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}} \sigma(\rho_{\varepsilon,\mathscr{N}})[\rho_{\varepsilon,\mathscr{N}} - f_\varepsilon]\ln\left(\frac{f_\varepsilon}{\rho_{\varepsilon,\mathscr{N}}}\right)\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x \le 0. \tag{3-6}$$

Let us focus on the following quantity obtained by multiplying (3-2) by φ and integrating

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}} \varphi(x) f_\varepsilon \,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x &= -\frac{1}{\varepsilon}\,\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}} \varphi(x) v\cdot\nabla_x f_\varepsilon \,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x\\ &= \frac{1}{\varepsilon}\,\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}} f_\varepsilon v\cdot\nabla_x\varphi(x)\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x\\ &= \mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}} v\cdot\nabla_x\varphi(x)\frac{f_\varepsilon - \rho_{\varepsilon,\mathscr{N}}}{\varepsilon}\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x.\end{aligned}$$

Note that we have used $\int v\,\mathrm{d}_{\mathscr{N}}(v) = 0$. By the Cauchy–Schwarz inequality, we know that

$$|\sqrt{b} - \sqrt{a}|^2 = \left|\int_a^b \frac{\mathrm{d}s}{2\sqrt{s}}\right|^2 \le \left|\int_a^b \frac{\mathrm{d}s}{4s}\right|\left|\int_a^b \mathrm{d}s\right| = \tfrac{1}{4}(b-a)\ln(b/a).$$

Thus, we get

$$\begin{aligned}\int_{\mathscr{V}} |f_\varepsilon - \rho_{\varepsilon,\mathscr{N}}|\,\mathrm{d}\mu_{\mathscr{N}}(v) &= \int_{\mathscr{V}} (\sqrt{f_\varepsilon} + \sqrt{\rho_{\varepsilon,\mathscr{N}}})\big|\sqrt{f_\varepsilon} - \sqrt{\rho_{\varepsilon,\mathscr{N}}}\big|\,\mathrm{d}\mu_{\mathscr{N}}(v)\\ &\le \left(\int_{\mathscr{V}} (\sqrt{f_\varepsilon} + \sqrt{\rho_{\varepsilon,\mathscr{N}}})^2\,\mathrm{d}\mu_{\mathscr{N}}(v)\right)^{1/2}\left(\int_{\mathscr{V}} (\sqrt{f_\varepsilon} - \sqrt{\rho_{\varepsilon,\mathscr{N}}})^2\,\mathrm{d}\mu_{\mathscr{N}}(v)\right)^{1/2}\\ &\le C\sqrt{\rho_{\varepsilon,\mathscr{N}}}\left(\int_{\mathscr{V}} (f_\varepsilon - \rho_{\varepsilon,\mathscr{N}})\ln(f_\varepsilon/\rho_{\varepsilon,\mathscr{N}})\,\mathrm{d}\mu_{\mathscr{N}}(v)\right)^{1/2},\end{aligned}$$

and we finally obtain the bound

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}}\varphi f_\varepsilon\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x\\
&\qquad\le \|v\|_{L^\infty(\Omega\times S)}\,\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}}|\nabla_x\varphi|\frac{|f_\varepsilon-\rho_{\varepsilon,\mathscr{N}}|}{\varepsilon}\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x\\
&\qquad\le C\,\mathbb{E}\int_{\mathbb{R}^D}|\nabla_x\varphi|\sqrt{\frac{\rho_{\varepsilon,\mathscr{N}}}{\sigma(\rho_{\varepsilon,\mathscr{N}})}}\left(\int_{\mathscr{V}}\frac{\sigma(\rho_{\varepsilon,\mathscr{N}})}{\varepsilon^2}(f_\varepsilon-\rho_{\varepsilon,\mathscr{N}})\ln(f_\varepsilon/\rho_{\varepsilon,\mathscr{N}})\,\mathrm{d}\mu_{\mathscr{N}}(v)\right)^{1/2}\mathrm{d}x\\
&\qquad\le C\,\mathbb{E}\left(\int_{\mathbb{R}^D}|\nabla_x\varphi|^2\frac{\rho_{\varepsilon,\mathscr{N}}}{\sigma(\rho_{\varepsilon,\mathscr{N}})}\,\mathrm{d}x\right)^{1/2}\left(\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}}\frac{\sigma(\rho_{\varepsilon,\mathscr{N}})}{\varepsilon^2}(f_\varepsilon-\rho_{\varepsilon,\mathscr{N}})\ln(f_\varepsilon/\rho_{\varepsilon,\mathscr{N}})\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x\right)^{1/2}.
\end{aligned}$$

By assumption, $1/\sigma(\rho_{\varepsilon,\mathscr{N}})$ is uniformly bounded. It follows that

$$\begin{aligned}
\mathbb{E}\int_{\mathbb{R}^D}|\nabla_x\varphi|^2\frac{\rho_{\varepsilon,\mathscr{N}}}{\sigma(\rho_{\varepsilon,\mathscr{N}})}\,\mathrm{d}x &\le C\left(\mathbb{E}\int_{\mathbb{R}^D}|\nabla_x\varphi|^{2q}\,\mathrm{d}x\right)^{1/q}\left(\mathbb{E}\int_{\mathbb{R}^D}\rho_{\varepsilon,\mathscr{N}}^p\,\mathrm{d}x\right)^{1/p}\\
&\le C\left(\mathbb{E}\int_{\mathbb{R}^D}|\nabla_x\varphi|^{2q}\,\mathrm{d}x\right)^{1/q}\left(\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}}|f_\varepsilon|^p\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x\right)^{1/p}\le C
\end{aligned}$$

holds provided the Hölder conjugate q of $p\ge 1$ satisfies $\beta\le 1/2-D/(4q)$.

The Young inequality

$$ab\le\frac{a^2}{4\theta}+\theta b^2$$

yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}}\varphi(x)f_\varepsilon(t,x,v)\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x\le C+\frac{1}{2}\,\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}}\frac{\sigma(\rho_{\varepsilon,\mathscr{N}})}{\varepsilon^2}(f_\varepsilon-\rho_{\varepsilon,\mathscr{N}})\ln(f_\varepsilon/\rho_{\varepsilon,\mathscr{N}})\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x.$$

Let us set

$$D_\varepsilon:=\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}}\frac{\sigma(\rho_{\varepsilon,\mathscr{N}})}{\varepsilon^2}(f_\varepsilon-\rho_{\varepsilon,\mathscr{N}})\ln(f_\varepsilon/\rho_{\varepsilon,\mathscr{N}})\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x\ge 0.$$

Coming back to (3-6), we get

$$\begin{aligned}
&\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}}f_\varepsilon(t,x,v)\ln f_\varepsilon(t,x,v)\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x+\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}}\varphi(x)f_\varepsilon(t,x,v)\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x+\frac{1}{2}\int_0^t D_\varepsilon(s)\,\mathrm{d}s\\
&\qquad\le Ct+\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}}f_\varepsilon^{\omega,0}(x,v)\ln f_\varepsilon^{\omega,0}(x,v)\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x+\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}}\varphi(x)f_\varepsilon^{\omega,0}(x,v)\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x.
\end{aligned}$$

Since $z|\ln z|=z\ln z-2z\ln z\,\mathbf{1}_{\{0\le z\le 1\}}$, we have

$$0\le-\int_{0\le f\le 1}f\ln f\,\mathrm{d}y=-\int_{0\le f\le e^{-\varphi}}f\ln f\,\mathrm{d}y-\int_{e^{-\varphi}\le f\le 1}f\ln f\,\mathrm{d}y\le\int\varphi f\,\mathrm{d}y+\int e^{-\varphi/2}\,\mathrm{d}y.$$

Then, we are led to

$$\begin{aligned}
&\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}} f_\varepsilon|\ln f_\varepsilon|\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x+\frac{1}{2}\int_0^t D_\varepsilon(s)\,\mathrm{d}s+\frac{1}{2}\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}}\varphi f_\varepsilon\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x\\
&\quad=\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}} f_\varepsilon\ln f_\varepsilon\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x-2\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}} f_\varepsilon\ln f_\varepsilon\mathbf{1}_{\{0\le f_\varepsilon\le 1\}}\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x\\
&\qquad\qquad+\frac{1}{2}\int_0^t D_\varepsilon(s)\,\mathrm{d}s+\frac{1}{2}\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}}\varphi f_\varepsilon\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x\\
&\quad\le\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}} f_\varepsilon\ln f_\varepsilon\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x+2\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}}\frac{\varphi}{4}f_\varepsilon\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x\\
&\qquad\qquad+2\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}} e^{-\varphi/8}\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x+\frac{1}{2}\int_0^t D_\varepsilon(s)\,\mathrm{d}s+\frac{1}{2}\mathbb{E}\int_{\mathbb{R}^D}\int_{\mathscr{V}}\varphi f_\varepsilon\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x\\
&\quad\le C(T).
\end{aligned}$$

□

Moreover, we can deduce from above that f_ε behaves like its macroscopic part $\rho_{\varepsilon,\mathscr{N}}$ for small ε.

Corollary 3.4. *We set $g_{\varepsilon,\mathscr{N}}:=(f_\varepsilon-\rho_{\varepsilon,\mathscr{N}})/\varepsilon$. Then, we have*

$$\sup_{\varepsilon>0,\ \mathscr{N}}\ \mathbb{E}\int_0^T\int_{\mathbb{R}^D}\left|\int_{\mathscr{V}} g_{\varepsilon,\mathscr{N}}\,\mathrm{d}\mu_{\mathscr{N}}(v)\right|^2\mathrm{d}x\,\mathrm{d}t\le C(T).$$

Proof. We write

$$\begin{aligned}
\mathbb{E}\int_0^T\int_{\mathbb{R}^D}\left|\int_{\mathscr{V}} g_{\varepsilon,\mathscr{N}}\,\mathrm{d}\mu_{\mathscr{N}}(v)\right|^2\mathrm{d}x\,\mathrm{d}t&=\mathbb{E}\int_0^T\int_{\mathbb{R}^D}\left(\int_{\mathscr{V}}\frac{|f_\varepsilon-\rho_{\varepsilon,\mathscr{N}}|}{\varepsilon}\,\mathrm{d}\mu_{\mathscr{N}}(v)\right)^2\mathrm{d}x\,\mathrm{d}t\\
&\le C\,\mathbb{E}\int_0^T\int_{\mathbb{R}^D}\rho_{\varepsilon,\mathscr{N}}\int_{\mathscr{V}}(f_\varepsilon-\rho_{\varepsilon,\mathscr{N}})\ln(f_\varepsilon/\rho_{\varepsilon,\mathscr{N}})\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x\,\mathrm{d}t\\
&\le C\,\mathbb{E}\int_0^T\int_{\mathbb{R}^D}\frac{\rho_{\varepsilon,\mathscr{N}}}{\sigma(\rho_{\varepsilon,\mathscr{N}})}\int_{\mathscr{V}}\sigma(\rho_{\varepsilon,\mathscr{N}})(f_\varepsilon-\rho_{\varepsilon,\mathscr{N}})\ln(f_\varepsilon/\rho_{\varepsilon,\mathscr{N}})\,\mathrm{d}\mu_{\mathscr{N}}(v)\,\mathrm{d}x\,\mathrm{d}t.
\end{aligned}$$

Since by the assumption on σ we know that $z\mapsto z/\sigma(z)$ is bounded on bounded sets and since $\rho_{\varepsilon,\mathscr{N}}$ is bounded in $L^\infty(\Omega\times(0,T)\times\mathbb{R}^D)$, we can conclude by using (3-4). □

Diffusive limit. We can now discuss how to pass to the limit $\varepsilon\to 0$.

Proof of Theorem 3.2. Applying the Dunford–Pettis theorem (see [Goudon 2011, §7.3.2]) we deduce from Proposition 3.3 that, possibly at the price of extracting a subsequence,

$$f_\varepsilon\rightharpoonup f_{\mathscr{N}}\quad\text{weakly in } L^1(\Omega\times(0,T)\times\mathbb{R}^D\times\mathscr{V}_{\mathscr{N}}).$$

Consequently, we also have

$$\rho_{\varepsilon,\mathscr{N}}=\int_{\mathscr{V}} f_\varepsilon\,\mathrm{d}\mu_{\mathscr{N}}(v)\rightharpoonup\rho_{\mathscr{N}}=\int_{\mathscr{V}} f_{\mathscr{N}}\,\mathrm{d}\mu_{\mathscr{N}}(v)\quad\text{weakly in } L^1(\Omega\times(0,T)\times\mathbb{R}^D)$$

and

$$\mathbb{E}\rho_{\varepsilon,\mathscr{N}}\rightharpoonup\mathbb{E}\rho_{\mathscr{N}}\quad\text{weakly in } L^1((0,T)\times\mathbb{R}^D).$$

Next, we consider the equations satisfied by the moments of f_ε. To this end, let us set

$$J_{\varepsilon,\mathscr{N}}(t,x) := \frac{1}{2\mathscr{N}} \sum_{i=1}^{2\mathscr{N}} \frac{V_i}{\varepsilon} f_\varepsilon(t,x,V_i), \quad \mathbb{P}_{\varepsilon,\mathscr{N}}(t,x) := \frac{1}{2\mathscr{N}} \sum_{i=1}^{2\mathscr{N}} V_i \otimes V_i f_\varepsilon(t,x,V_i).$$

Integrating (3-2) with respect to the velocity variable v yields

$$\partial_t \rho_{\varepsilon,\mathscr{N}} + \operatorname{div}(J_{\varepsilon,\mathscr{N}}) = 0. \tag{3-7}$$

Similarly, multiplying (3-2) by v and integrating leads to

$$\varepsilon^2 \partial_t J_{\varepsilon,\mathscr{N}} + \operatorname{div}(\mathbb{P}_{\varepsilon,\mathscr{N}}) = -\sigma(\rho_{\varepsilon,\mathscr{N}}) J_{\varepsilon,\mathscr{N}}. \tag{3-8}$$

Lemma 3.5. *The sequence $(J_{\varepsilon,\mathscr{N}})_{\varepsilon>0}$ is bounded in $L^2(\Omega\times(0,T)\times\mathbb{R}^D)$ and we can write $\mathbb{P}_{\varepsilon,\mathscr{N}} = A_{\mathscr{N}}\rho_{\varepsilon,\mathscr{N}} + \varepsilon\mathbb{K}_{\varepsilon,\mathscr{N}}$ with $A_{\mathscr{N}} = \frac{1}{2\mathscr{N}}\sum_{j=1}^{2\mathscr{N}} V_j\otimes V_j$ and the components of $(\mathbb{K}_{\varepsilon,\mathscr{N}})_{\varepsilon>0}$ are bounded in $L^2(\Omega\times(0,T)\times\mathbb{R}^D)$.*

Proof. The proof is based on the fact that $f_\varepsilon = \rho_{\varepsilon,\mathscr{N}} + \varepsilon g_{\varepsilon,\mathscr{N}}$. Since $\sum_{j=1}^{2\mathscr{N}} V_j = 0$, it allows us to write

$$J_{\varepsilon,\mathscr{N}} = \int v g_{\varepsilon,\mathscr{N}} \,\mathrm{d}\mu_{\mathscr{N}}(v),$$

and we deduce the bound on $J_{\varepsilon,\mathscr{N}}$ from Corollary 3.4 since $\|v\|_{L^\infty(\Omega\times S)} \le C$. In addition, we have

$$\mathbb{P}_{\varepsilon,\mathscr{N}} = \int v\otimes v \,\mathrm{d}\mu_{\mathscr{N}}(v)\rho_{\varepsilon,\mathscr{N}} + \varepsilon\int v\otimes v g_{\varepsilon,\mathscr{N}} \,\mathrm{d}\mu_{\mathscr{N}}(v).$$

We set

$$\mathbb{K}_{\varepsilon,\mathscr{N}}(t,x) := \int v\otimes v g_{\varepsilon,\mathscr{N}}(t,x,v)\,\mathrm{d}\mu_{\mathscr{N}}(v).$$

We conclude by using the estimates in Corollary 3.4 again. □

Owing to Lemma 3.5, (3-8) can be recast as

$$\varepsilon\big(\varepsilon\partial_t J_{\varepsilon,\mathscr{N}} + \operatorname{div}(\mathbb{K}_{\varepsilon,\mathscr{N}})\big) + A_{\mathscr{N}}\nabla_x\rho_{\varepsilon,\mathscr{N}} = -\nu_{\varepsilon,\mathscr{N}},$$

with $\nu_{\varepsilon,\mathscr{N}} := \sigma(\rho_{\varepsilon,\mathscr{N}})J_{\varepsilon,\mathscr{N}}$. Passing to the limit, up to subsequences, we are led to

$$\begin{cases} \partial_t\rho_{\mathscr{N}} + \operatorname{div}(J_{\mathscr{N}}) = 0, \\ A_{\mathscr{N}}\nabla\rho_{\mathscr{N}} = -\nu_{\mathscr{N}}, \end{cases} \tag{3-9}$$

where $\nu_{\mathscr{N}}$ is the weak limit as $\varepsilon\to 0$ of $\nu_{\varepsilon,\mathscr{N}}$, which is a bounded sequence in $L^2(\Omega\times(0,T)\times\mathbb{R}^D)$. It remains to establish a relation between $\nu_{\mathscr{N}}$, $\rho_{\mathscr{N}}$ and $J_{\mathscr{N}}$, or more precisely the expectation of these quantities. To this end, we are going to use the strong compactness of $\mathbb{E}\rho_{\varepsilon,\mathscr{N}}$ by using the averaging lemma. Indeed, we know that $\mathbb{E}\rho_{\varepsilon,\mathscr{N}}$ belongs to a bounded set in $L^2(0,T;H^{1/2}(\mathbb{R}^D))$; the proof follows exactly the same argument as for Theorem 2.3, taking the Fourier transform with respect to both the time and space variables t,x. However, because of the ε in front of the time derivative, we cannot expect a gain of regularity with respect to the time variable. Then, we need to combine this estimate with another argument as follows:

(i) By using the Weil–Kolmogorov–Fréchet theorem, see [Goudon 2011, Théorème 7.56], we deduce from the averaging lemma that

$$\lim_{|h|\to 0}\left(\sup_{\varepsilon}\int_0^T\int_{\mathbb{R}^D}\left|\mathbb{E}\rho_{\varepsilon,\mathscr{N}}(t,x+h)-\mathbb{E}\rho_{\varepsilon,\mathscr{N}}(t,x)\right|^2\mathrm{d}x\,\mathrm{d}t\right)=0.$$

(ii) Going back to (3-7), Lemma 3.5 tells us that $\partial_t\mathbb{E}\rho_{\varepsilon,\mathscr{N}}=-\mathrm{div}(\mathbb{E}J_{\varepsilon,\mathscr{N}})$ is bounded, uniformly with respect to ε, in $L^2(0,T;H^{-1}(\mathbb{R}^D))$.

Then, this is enough to deduce that $\mathbb{E}\rho_{\varepsilon,\mathscr{N}}$ strongly converges to $\mathbb{E}\rho_{\mathscr{N}}$ in $L^2((0,T)\times\mathbb{R}^D)$ (see, e.g., [Alonso et al. 2017, Appendix B] for a detailed proof).

Then, we rewrite

$$\mathbb{E}J_{\varepsilon,\mathscr{N}}=\mathbb{E}\left(\frac{\nu_{\varepsilon,\mathscr{N}}}{\sigma(\rho_{\varepsilon,\mathscr{N}})}\right)=\frac{\mathbb{E}\nu_{\varepsilon,\mathscr{N}}}{\sigma(\mathbb{E}\rho_{\varepsilon,\mathscr{N}})}+\mathbb{E}r_{\varepsilon,\mathscr{N}},\quad r_{\varepsilon,\mathscr{N}}=\left[\nu_{\varepsilon,\mathscr{N}}\left(\frac{1}{\sigma(\rho_{\varepsilon,\mathscr{N}})}-\frac{1}{\sigma(\mathbb{E}\rho_{\varepsilon,\mathscr{N}})}\right)\right]. \tag{3-10}$$

From the previous discussion, extracting further subsequences if necessary, we know that $\mathbb{E}\nu_{\varepsilon,\mathscr{N}}$ converges weakly to $\mathbb{E}\nu_N$ in $L^2((0,T)\times\mathbb{R}^D)$, while $\mathbb{E}\rho_{\varepsilon,\mathscr{N}}$ converges strongly in $L^2((0,T)\times\mathbb{R}^D)$ and a.e. to $\mathbb{E}\rho_{\mathscr{N}}$. Since σ is continuous and bounded from below, $1/\sigma(\mathbb{E}\rho_{\varepsilon,\mathscr{N}})$ converges to $1/\sigma(\mathbb{E}\rho_{\mathscr{N}})$ a.e. too, and it is bounded in $L^\infty((0,T)\times\mathbb{R}^D)$. We deduce that

$$\frac{\mathbb{E}\nu_{\varepsilon,\mathscr{N}}}{\sigma(\mathbb{E}\rho_{\varepsilon,\mathscr{N}})}\rightharpoonup\frac{\mathbb{E}\nu_{\mathscr{N}}}{\sigma(\mathbb{E}\rho_{\mathscr{N}})}\quad\text{weakly in } L^2((0,T)\times\mathbb{R}^D).$$

We are left with the task of proving that the last term in the right hand side of (3-10) tends to 0 as $\mathscr{N}\to\infty$, uniformly with respect to ε. The Cauchy–Schwarz inequality yields

$$\begin{aligned}
|\mathbb{E}\,r_{\varepsilon,\mathscr{N}}| &\le (\mathbb{E}[(\nu_{\varepsilon,\mathscr{N}})^2])^{1/2}\left(\mathbb{E}\left[\left(\frac{1}{\sigma(\rho_{\varepsilon,\mathscr{N}})}-\frac{1}{\sigma(\mathbb{E}\rho_{\varepsilon,\mathscr{N}})}\right)^2\right]\right)^{1/2}\\
&\le (\mathbb{E}\,[(\nu_{\varepsilon,\mathscr{N}})^2])^{1/2}\left(\mathbb{E}\left[\left(\int_{\mathbb{E}\rho_{\varepsilon,\mathscr{N}}}^{\rho_{\varepsilon,\mathscr{N}}}\frac{\mathrm{d}}{\mathrm{d}z}\left[\frac{1}{\sigma(z)}\right]\mathrm{d}z\right)\right]^2\right)^{1/2}\\
&\le (\mathbb{E}\,[(\nu_{\varepsilon,\mathscr{N}})^2])^{1/2}\big(\mathbb{E}\,[(\rho_{\varepsilon,\mathscr{N}}-\mathbb{E}\rho^2_{\varepsilon,\mathscr{N}}])\big)^{1/2}\\
&\le (\mathbb{E}\,[(\nu_{\varepsilon,\mathscr{N}})^2])^{1/2}\left(\mathbb{E}\left[\left(\frac{1}{2\mathscr{N}}\sum_{i=1}^{2\mathscr{N}}f_\varepsilon(V_i)-\mathbb{E}\rho_{\varepsilon,\mathscr{N}}\right)^2\right]\right)^{1/2}.
\end{aligned}\tag{3-11}$$

We remind the reader that the $2\mathscr{N}$ velocities are constructed by symmetry from $V_1,\dots,V_{\mathscr{N}}$, which are i.i.d. velocities in $[-0.5,0.5]^D$, and we write

$$\begin{aligned}
&\mathbb{E}\left[12\mathscr{N}\sum_{i=1}^{2\mathscr{N}}f_\varepsilon(V_i)-\mathbb{E}\rho_{\varepsilon,\mathscr{N}}\right]^2\\
&\qquad=\mathbb{E}\left[\frac{1}{4\mathscr{N}^2}\sum_{i,j=1}^{\mathscr{N}}\big\{\big(f_\varepsilon(V_i)+f_\varepsilon(-V_i)-2\mathbb{E}\rho_{\varepsilon,\mathscr{N}}\big)\big(f_\varepsilon(V_j)+f_\varepsilon(-V_j)-2\mathbb{E}\rho_{\varepsilon,\mathscr{N}}\big)\big\}\right].
\end{aligned}\tag{3-12}$$

When $i \neq j$, we know V_i and V_j are independent, which implies

$$\mathbb{E}\big[(f_\varepsilon(V_i)+f_\varepsilon(-V_i)-2\mathbb{E}\rho_{\varepsilon,\mathscr{N}})(f_\varepsilon(V_j)+f_\varepsilon(-V_j)-2\mathbb{E}\rho_{\varepsilon,\mathscr{N}})\big] \\ = \mathbb{E}\big[f_\varepsilon(V_i)+f_\varepsilon(-V_i)-2\mathbb{E}\rho_{\varepsilon,\mathscr{N}}\big]\,\mathbb{E}\big[f_\varepsilon(V_j)+f_\varepsilon(-V_j)-2\mathbb{E}\rho_{\varepsilon,\mathscr{N}}\big].$$

Now, we use the fact that the V_i are identically distributed so that

$$2\mathbb{E}\rho_{\varepsilon,\mathscr{N}} = 2\mathbb{E}\left(\frac{1}{2\mathscr{N}}\sum_{k=1}^{2\mathscr{N}} f_\varepsilon(V_k)\right) = \mathbb{E}\left(\frac{1}{\mathscr{N}}\sum_{k=1}^{\mathscr{N}}\big(f_\varepsilon(V_k)+f_\varepsilon(-V_k)\big)\right) \\ = \frac{1}{\mathscr{N}}\sum_{k=1}^{\mathscr{N}}\big(\mathbb{E}f_\varepsilon(V_k)+\mathbb{E}f_\varepsilon(-V_k)\big) = \mathbb{E}f_\varepsilon(V_j)+\mathbb{E}f_\varepsilon(-V_j)$$

for any $j \in \{1,\dots,\mathscr{N}\}$. It follows that

$$\mathbb{E}\big[f_\varepsilon(V_i)+f_\varepsilon(-V_i)-2\mathbb{E}\rho_{\varepsilon,\mathscr{N}})(f_\varepsilon(V_j)+f_\varepsilon(-V_j)-2\mathbb{E}\rho_{\varepsilon,\mathscr{N}})\big] = 0 \quad \text{when } i \neq j.$$

Going back to (3-12), we obtain

$$\mathbb{E}\left[\frac{1}{2\mathscr{N}}\sum_{i=1}^{2\mathscr{N}} f_\varepsilon(V_i)-\mathbb{E}\rho_{\varepsilon,\mathscr{N}}\right]^2 = \mathbb{E}\left[\frac{1}{4\mathscr{N}^2}\sum_{i=1}^{\mathscr{N}}\big(f_\varepsilon(V_i)+f_\varepsilon(-V_i)-2\mathbb{E}\rho_{\varepsilon,\mathscr{N}}\big)^2\right].$$

Since f_ε and $\rho_{\varepsilon,\mathscr{N}}$ are uniformly bounded, we conclude that the estimate

$$\mathbb{E}\left[\frac{1}{2\mathscr{N}}\sum_{i=1}^{2\mathscr{N}} f_\varepsilon(V_i)-\mathbb{E}\rho_{\varepsilon,\mathscr{N}}\right]^2 \leq \frac{C}{\mathscr{N}}$$

holds. Inserting this information in (3-11), we arrive at

$$\int_0^T\int_{\mathbb{R}^D} |\mathbb{E}r_{\varepsilon,\mathscr{N}}|^2\,\mathrm{d}x\,\mathrm{d}t \leq \frac{C}{\mathscr{N}}\,\mathbb{E}\int_0^T\int_{\mathbb{R}^D} v_{\varepsilon,\mathscr{N}}^2\,\mathrm{d}x\,\mathrm{d}t,$$

which is thus of order $O(1/\mathscr{N})$, uniformly with respect to ε.

Therefore, we can let ε run to 0 in (3-10) and, for a suitable subsequence, we are led to

$$\mathbb{E}J_{\varepsilon,\mathscr{N}} \rightharpoonup \mathbb{E}J_{\mathscr{N}} = \frac{\mathbb{E}\nu_{\mathscr{N}}}{\sigma(\mathbb{E}\rho_{\mathscr{N}})} + r_{\mathscr{N}} \quad \text{weakly in } L^2((0,T)\times\mathbb{R}^D) \text{ with } \|r_{\mathscr{N}}\|_{L^2((0,T)\times\mathbb{R}^D)} \leq \frac{C}{\sqrt{\mathscr{N}}}.$$

Finally, we take the expectation in (3-9) and we get

$$\mathbb{E}(A_{\mathscr{N}}\nabla_x\rho_{\mathscr{N}}) = -\mathbb{E}\nu_{\mathscr{N}} = -\sigma(\mathbb{E}\rho_{\mathscr{N}})\mathbb{E}J_{\mathscr{N}} + \sigma(\mathbb{E}\rho_{\mathscr{N}})r_{\mathscr{N}}.$$

Note that the last term is still of order $O(1/\sqrt{\mathscr{N}})$ in the $L^2((0,T)\times\mathbb{R}^D)$ norm. By reasoning similar to that above, we check that, for any $i, j \in \{1,\dots,D\}$,

$$\sqrt{\mathbb{E}\big[\big([A_{\mathscr{N}}]_{ij}-\mathbb{E}[A_{\mathscr{N}}]_{ij}\big)^2\big]} = O\left(\frac{1}{\sqrt{\mathscr{N}}}\right)$$

(this is the standard result about Monte Carlo integration). It implies that we can find a constant $C > 0$, which only depends on the dimension D, such that for any $\xi \in \mathbb{R}^D$,

$$\mathbb{E}\big[\big|A_{\mathscr{N}}\xi - \mathbb{E}[A_{\mathscr{N}}\xi]\big|^2\big] \le \frac{C|\xi|^2}{\mathscr{N}}.$$

Then we get

$$\mathbb{E}(A_{\mathscr{N}}\nabla_x\rho_{\mathscr{N}}) = \mathbb{E}A_{\mathscr{N}}\nabla_x\mathbb{E}\rho_{\mathscr{N}} + s_{\mathscr{N}}, \quad s_{\mathscr{N}} = \mathbb{E}\big[(A_{\mathscr{N}} - \mathbb{E}A_{\mathscr{N}})\nabla_x\rho_{\mathscr{N}}\big].$$

The remainder term should be analyzed in a weak sense, due to a lack of a priori regularity of $\nabla_x\rho_{\mathscr{N}}$ (we only know that the product $A_{\mathscr{N}}\nabla_x\rho_{\mathscr{N}}$ lies in L^2, but the invertibility of $A_{\mathscr{N}}$ is not guaranteed). We have, for any $\varphi \in C_c^\infty((0,T)\times\mathbb{R}^D)$,

$$\begin{aligned}\big|\langle \mathbb{E}s_{\mathscr{N}}|\varphi\rangle\big| &= \left|-\mathbb{E}\int_0^T\int_{\mathbb{R}^D} \rho_{\mathscr{N}}(A_{\mathscr{N}} - \mathbb{E}A_{\mathscr{N}})\nabla_x\varphi \,\mathrm{d}x\,\mathrm{d}t\right| \\ &\le \left(\mathbb{E}\int_0^T\int_{\mathbb{R}^D} \rho_{\mathscr{N}}^2 \,\mathrm{d}x\,\mathrm{d}t\right)^{1/2}\left(\int_0^T\int_{\mathbb{R}^D} |\nabla_x\varphi|^2 \,\mathrm{d}x\,\mathrm{d}t\right)^{1/2} \frac{C}{\sqrt{\mathscr{N}}}.\end{aligned}$$

Owing to the estimates (3-3) in Proposition 3.3, it means that $s_{\mathscr{N}}$ is therefore of order $O(1/\sqrt{\mathscr{N}})$ in the $L^2(0,T;H^{-1}(\mathbb{R}^D))$-norm. □

Remark 3.6. The random matrix $A_{\mathscr{N}}$ might be singular. However $\mathbb{E}A_{\mathscr{N}}$ is invertible. Indeed for any $\xi \neq 0$, we have

$$\mathbb{E}A_{\mathscr{N}}\xi\cdot\xi = \frac{1}{2\mathscr{N}}\sum_{j=1}^{2\mathscr{N}} \mathbb{E}[|V_j\cdot\xi|^2] \ge 0.$$

This quantity is actually positive since $\mathbb{P}(v\cdot\xi = 0) = 0$ for the continuous laws we are dealing with.

4. Comments and perspectives

The Monte Carlo procedure is widely used to numerically evaluate multidimensional integrals, precisely because, evaluating the numerical effort by the number $\mathscr{N}$ of quadrature points, it provides a result with an accuracy of order $O(1/\sqrt{\mathscr{N}})$, independently of the space dimension, in contrast to the deterministic quadrature methods where the error is $O(\mathscr{N}^{-k/D})$, k being the order of the method; see [Caflisch 1998; Lapeyre et al. 1998, Chapitre 1]. Application of such stochastic quadrature approaches to the numerical treatment of kinetic models for neutron transport dates back to the Manhattan project [Metropolis and Ulam 1949]. For applications to radiative transfer computations we refer the reader, e.g., to [Campbell 1967] and for a more recent overview to [Whitney 2011]. After the pioneering works by K. Nanbu [1980] and G. A. Bird [1970], Monte Carlo techniques are at the basis of the simulation of the Boltzmann equation for rarefied gases. (By the way, note that the construction of a suitable deterministic quadrature formula for approximating the Boltzmann operator can be a bit tricky, with unexpected connections to subtle number theory arguments [Michel and Schneider 2000].) Very comprehensive introductions can be found in [Graham and Méléard 1999; Pareschi 2005; Pareschi and Russo 1999] and in the textbook [Lapeyre et al. 1998]. The method can naturally be presented as a particulate method; roughly speaking,

it works according to a splitting approach [Lapeyre et al. 1998, Chapter 3]: first, particles (which, here, are "test" particles intended to actually represent a set of real particles) are displaced according to free transport over the time step Δt, and, second, the effects of the interaction between particles during the time step are evaluated by using a random sampling. Convergence of the method for the Boltzmann equation as the number of particles tends to ∞ is analyzed in [Graham and Méléard 1997; Pulvirenti et al. 1994; Wagner 1992; 2004]. However, the performance of Monte Carlo algorithms is known to degrade in near-continuum regimes, where the number of collision events per time unit increases; see [Caflisch 1998, §7; Lapeyre et al. 1998, §3.7.1 and §4.5]. This observation has motivated the development of hybrid methods [Dimarco and Pareschi 2008; Pareschi 2005].

As pointed out in the Introduction, the average lemma plays a central role in the analysis of nonlinear kinetic models and their hydrodynamic limits, with fundamental obstructions in extending to discrete velocity models. We expect that the stochastic average lemma established here might help in analyzing stochastic algorithms for kinetic models. Our first attempt remains at the level of space-time continuous models for the simplest radiative transfer equation: it is just a consistency result with the diffusion approximation. It is remarkable that the consistency error preserves the typical feature of the Monte Carlo error estimate in $O(1/\sqrt{\mathcal{N}})$, independently of the space dimension. A next step, likely inspired by the "time-discretized" version of the averaging lemma in [Bouchut and Desvillettes 1999; Horsin et al. 2003], would be to consider time-discretized models, where the random velocity grid is reconstructed at each time step.

References

[Agoshkov 1984] V. I. Agoshkov, "Spaces of functions with differential-difference characteristics and the smoothness of solutions of the transport equation", *Dokl. Akad. Nauk SSSR* **276**:6 (1984), 1289–1293. In Russian; translated in *Sov. Math. Dokl.* **29** (1984), 662–666. MR Zbl

[Alonso et al. 2017] A. Alonso, T. Goudon, and A. Vavasseur, "Damping of particles interacting with a vibrating medium", *Ann. Inst. H. Poincaré Anal. Non Linéaire* (online publication January 2017).

[Bardos et al. 1988] C. Bardos, F. Golse, B. Perthame, and R. Sentis, "The nonaccretive radiative transfer equations: existence of solutions and Rosseland approximation", *J. Funct. Anal.* **77**:2 (1988), 434–460. MR Zbl

[Bergh and Löfström 1976] J. Bergh and J. Löfström, *Interpolation spaces: an introduction*, Grundlehren der Mathematischen Wissenschaften **223**, Springer, 1976. MR Zbl

[Berthelin and Junca 2010] F. Berthelin and S. Junca, "Averaging lemmas with a force term in the transport equation", *J. Math. Pures Appl.* (9) **93**:2 (2010), 113–131. MR Zbl

[Bird 1970] G. A. Bird, "Direct simulation and the Boltzmann equation", *Phys. Fluids* **13**:11 (1970), 2676–2687. Zbl

[Bonnaillie-Noël et al. 2016] V. Bonnaillie-Noël, J. A. Carrillo, T. Goudon, and G. A. Pavliotis, "Efficient numerical calculation of drift and diffusion coefficients in the diffusion approximation of kinetic equations", *IMA J. Numer. Anal.* **36**:4 (2016), 1536–1569. MR

[Bouchut and Desvillettes 1999] F. Bouchut and L. Desvillettes, "Averaging lemmas without time Fourier transform and application to discretized kinetic equations", *Proc. Roy. Soc. Edinburgh Sect. A* **129**:1 (1999), 19–36. MR Zbl

[Buet et al. 2002] C. Buet, S. Cordier, B. Lucquin-Desreux, and S. Mancini, "Diffusion limit of the Lorentz model: asymptotic preserving schemes", *M2AN Math. Model. Numer. Anal.* **36**:4 (2002), 631–655. MR Zbl

[Caflisch 1998] R. E. Caflisch, "Monte Carlo and quasi-Monte Carlo methods", *Acta Numer.* **7** (1998), 1–49. MR Zbl

[Campbell 1967] P. M. Campbell, "Monte Carlo method for radiative transfer", *Int. J. Heat and Mass Transfer* **10**:4 (1967), 519–527.

[Debussche et al. 2015] A. Debussche, S. De Moor, and J. Vovelle, "Diffusion limit for the radiative transfer equation perturbed by a Wiener process", *Kinet. Relat. Models* **8**:3 (2015), 467–492. MR Zbl

[Debussche et al. 2016] A. Debussche, S. De Moor, and J. Vovelle, "Diffusion limit for the radiative transfer equation perturbed by a Markovian process", *Asymptot. Anal.* **98**:1-2 (2016), 31–58. MR Zbl

[Degond et al. 2000] P. Degond, T. Goudon, and F. Poupaud, "Diffusion limit for nonhomogeneous and non-micro-reversible processes", *Indiana Univ. Math. J.* **49**:3 (2000), 1175–1198. MR Zbl

[Dimarco and Pareschi 2008] G. Dimarco and L. Pareschi, "Hybrid multiscale methods, II: Kinetic equations", *Multiscale Model. Simul.* **6**:4 (2008), 1169–1197. MR Zbl

[DiPerna and Lions 1989a] R. J. DiPerna and P.-L. Lions, "Global weak solutions of Vlasov–Maxwell systems", *Comm. Pure Appl. Math.* **42**:6 (1989), 729–757. MR Zbl

[DiPerna and Lions 1989b] R. J. DiPerna and P.-L. Lions, "On the Cauchy problem for Boltzmann equations: global existence and weak stability", *Ann. of Math.* (2) **130**:2 (1989), 321–366. MR Zbl

[DiPerna et al. 1991] R. J. DiPerna, P.-L. Lions, and Y. Meyer, "L^p regularity of velocity averages", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **8**:3-4 (1991), 271–287. MR Zbl

[Golse 2000] F. Golse, "From kinetic to macroscopic models", pp. 41–121 in *Kinetic equations and asymptotic theory*, edited by B. Perthame and L. Desvillettes, Series in Applied Mathematics **4**, Gauthier-Villars, Paris, 2000. MR Zbl

[Golse and Saint-Raymond 2002] F. Golse and L. Saint-Raymond, "Velocity averaging in L^1 for the transport equation", *C. R. Math. Acad. Sci. Paris* **334**:7 (2002), 557–562. MR Zbl

[Golse and Saint-Raymond 2004] F. Golse and L. Saint-Raymond, "The Navier–Stokes limit of the Boltzmann equation for bounded collision kernels", *Invent. Math.* **155**:1 (2004), 81–161. MR Zbl

[Golse et al. 1988] F. Golse, P.-L. Lions, B. t. Perthame, and R. Sentis, "Regularity of the moments of the solution of a transport equation", *J. Funct. Anal.* **76**:1 (1988), 110–125. MR Zbl

[Golse et al. 1999] F. Golse, S. Jin, and C. D. Levermore, "The convergence of numerical transfer schemes in diffusive regimes, I: Discrete-ordinate method", *SIAM J. Numer. Anal.* **36**:5 (1999), 1333–1369. MR Zbl

[Goudon 2011] T. Goudon, *Intégration: intégrale de Lebesgue et introduction à l'analyse fonctionnelle*, Ellipses, Paris, 2011. Zbl

[Goudon and Poupaud 2001] T. Goudon and F. Poupaud, "Approximation by homogenization and diffusion of kinetic equations", *Comm. Partial Differential Equations* **26**:3-4 (2001), 537–569. MR Zbl

[Graham and Méléard 1997] C. Graham and S. Méléard, "Stochastic particle approximations for generalized Boltzmann models and convergence estimates", *Ann. Probab.* **25**:1 (1997), 115–132. MR Zbl

[Graham and Méléard 1999] C. Graham and S. Méléard, "Probabilistic tools and Monte-Carlo approximations for some Boltzmann equations", pp. 77–126 in *CEMRACS 1999* (Orsay, 1999), edited by F. Coquel and S. Cordier, ESAIM Proc. **10**, Soc. Math. Appl. Indust., Paris, 1999. MR Zbl

[Horsin et al. 2003] T. Horsin, S. Mischler, and A. Vasseur, "On the convergence of numerical schemes for the Boltzmann equation", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **20**:5 (2003), 731–758. MR Zbl

[Jin and Levermore 1991] S. Jin and D. Levermore, "The discrete-ordinate method in diffusive regimes", *Transport Theory Statist. Phys.* **20**:5-6 (1991), 413–439. MR Zbl

[Lapeyre et al. 1998] B. Lapeyre, E. Pardoux, and R. Sentis, *Méthodes de Monte-Carlo pour les équations de transport et de diffusion*, Mathématiques & Applications **29**, Springer, 1998. MR Zbl

[Lions and Toscani 1997] P. L. Lions and G. Toscani, "Diffusive limit for finite velocity Boltzmann kinetic models", *Rev. Mat. Iberoamericana* **13**:3 (1997), 473–513. MR Zbl

[Lions et al. 2013] P.-L. Lions, B. t. Perthame, and P. E. Souganidis, "Stochastic averaging lemmas for kinetic equations", exposé 26 , 17 p. in *Séminaire Laurent Schwartz: équations aux dérivées partielles et applications,* 2011–2012, École Polytech., Palaiseau, 2013. MR Zbl

[Marcati and Milani 1990] P. Marcati and A. Milani, "The one-dimensional Darcy's law as the limit of a compressible Euler flow", *J. Differential Equations* **84**:1 (1990), 129–147. MR Zbl

[Metropolis and Ulam 1949] N. Metropolis and S. Ulam, "The Monte Carlo method", *J. Amer. Statist. Assoc.* **44** (1949), 335–341. MR Zbl

[Michel and Schneider 2000] P. Michel and J. Schneider, "Approximation simultanée de réels par des nombres rationnels et noyau de collision de l'équation de Boltzmann", *C. R. Acad. Sci. Paris Sér. I Math.* **330**:9 (2000), 857–862. MR Zbl

[Mischler 1997] S. Mischler, "Convergence of discrete-velocity schemes for the Boltzmann equation", *Arch. Rational Mech. Anal.* **140**:1 (1997), 53–77. MR Zbl

[Nanbu 1980] K. Nanbu, "Direct simulation scheme derived from the Boltzmann equation, I: Monocomponent gases", *J. Phys. Soc. Jpn.* **49**:5 (1980), 2042–2049.

[Pareschi 2005] L. Pareschi, "Hybrid multiscale methods for hyperbolic and kinetic problems", pp. 87–120 in *GRIP—Research Group on Particle Interactions* (Sophia Antipolis, France, 2005), edited by T. Goudon et al., ESAIM Proc. **15**, EDP Sci., Les Ulis, 2005. MR Zbl

[Pareschi and Russo 1999] L. Pareschi and G. Russo, "An introduction to Monte Carlo methods for the Boltzmann equation", pp. 35–76 in *CEMRACS 1999* (Orsay, 1999), edited by F. Coquel and S. Cordier, ESAIM Proc. **10**, Soc. Math. Appl. Indust., Paris, 1999. MR Zbl

[Perthame and Souganidis 1998] B. Perthame and P. E. Souganidis, "A limiting case for velocity averaging", *Ann. Sci. École Norm. Sup.* (4) **31**:4 (1998), 591–598. MR Zbl

[Pulvirenti et al. 1994] M. Pulvirenti, W. Wagner, and M. B. Zavelani Rossi, "Convergence of particle schemes for the Boltzmann equation", *European J. Mech. B Fluids* **13**:3 (1994), 339–351. MR Zbl

[Saint-Raymond 2009] L. Saint-Raymond, *Hydrodynamic limits of the Boltzmann equation*, Lecture Notes in Mathematics **1971**, Springer, 2009. MR Zbl

[Tadmor and Tao 2007] E. Tadmor and T. Tao, "Velocity averaging, kinetic formulations, and regularizing effects in quasi-linear PDEs", *Comm. Pure Appl. Math.* **60**:10 (2007), 1488–1521. MR Zbl

[Villani 2002] C. Villani, "Limites hydrodynamiques de l'équation de Boltzmann", exposé 893 , pp. 365–405 in *Séminaire Bourbaki,* 2000/2001, Astérisque **282**, Soc. Mat. de France, Paris, 2002. MR Zbl

[Wagner 1992] W. Wagner, "A convergence proof for Bird's direct simulation Monte Carlo method for the Boltzmann equation", *J. Statist. Phys.* **66**:3-4 (1992), 1011–1044. MR Zbl

[Wagner 2004] W. Wagner, "Stochastic models and Monte Carlo algorithms for Boltzmann type equations", pp. 129–153 in *Monte Carlo and quasi-Monte Carlo methods 2002* (Singapore, 2002), edited by H. Niederreiter, Springer, 2004. MR Zbl

[Whitney 2011] B. A. Whitney, "Monte Carlo radiative transfer", *Bull. Astr. Soc. India* **39**:1 (2011), 101–127.

Received 7 Nov 2016. Revised 12 Feb 2017. Accepted 28 Mar 2017.

NATHALIE AYI: nathalie.ayi@inria.fr
Inria Rennes-Bretagne Atlantique, IPSO, Research team, IRMAR, UMR CNRS 6625, Campus de Beaulieu, Bâtiment 22/23, 263 Avenue du Général Leclerc, 35042 Rennes, France

THIERRY GOUDON: thierry.goudon@inria.fr
Université Côte d'Azur, Inria, CNRS, LJAD, Parc Valrose, 06108 Nice, France

ANALYSIS AND PDE
Vol. 10, No. 5, 2017
dx.doi.org/10.2140/apde.2017.10.1227

PERRON'S METHOD FOR NONLOCAL FULLY NONLINEAR EQUATIONS

CHENCHEN MOU

This paper is concerned with the existence of viscosity solutions of nonlocal fully nonlinear equations that are not translation-invariant. We construct a discontinuous viscosity solution of such a nonlocal equation by Perron's method. If the equation is uniformly elliptic, we prove the discontinuous viscosity solution is Hölder continuous and thus it is a viscosity solution.

1. Introduction

We investigate the existence of a viscosity solution of

$$\begin{cases} I(x, u(x), u(\cdot)) = 0 & \text{in } \Omega, \\ u = g & \text{in } \Omega^c, \end{cases} \tag{1-1}$$

where Ω is a bounded domain in $\mathbb{R}^n$, I is a nonlocal operator that is not translation-invariant and g is a bounded continuous function in $\mathbb{R}^n$.

An important example of (1-1) is the Dirichlet problem for nonlocal Bellman–Isaacs equations, i.e.,

$$\begin{cases} \sup_{a\in\mathcal{A}} \inf_{b\in\mathcal{B}} \{-I_{ab}[x,u] + b_{ab}(x)\cdot\nabla u(x) + c_{ab}(x)u(x) + f_{ab}(x)\} = 0 & \text{in } \Omega, \\ u = g & \text{in } \Omega^c, \end{cases} \tag{1-2}$$

where $\mathcal{A}, \mathcal{B}$ are two index sets, $b_{ab} : \mathbb{R}^n \to \mathbb{R}^n$, $c_{ab} : \mathbb{R}^n \to \mathbb{R}^+$, $f_{ab} : \mathbb{R}^n \to \mathbb{R}$ are uniformly continuous functions and I_{ab} is a Lévy operator. If the Lévy measures are symmetric and absolutely continuous with respect to the Lebesgue measure, then they can be represented as

$$I_{ab}[x,u] := \int_{\mathbb{R}^n} [u(x+z) - u(x)] K_{ab}(x,z)\, dz, \tag{1-3}$$

where $\{K_{ab}(x,\cdot) : x \in \Omega, a \in \mathcal{A}, b \in \mathcal{B}\}$ are kernels of Lévy measures satisfying

$$\int_{\mathbb{R}^n} \min\{|z|^2, 1\} K_{ab}(x,z)\, dz < +\infty \quad \text{for all } x \in \Omega. \tag{1-4}$$

In fact, we will not assume our Lévy measures to be symmetric in the following sections.

Existence of viscosity solutions has been well established for the Dirichlet problem for integro-differential equations by Perron's method when the equations satisfy the comparison principle. G. Barles and C. Imbert [Barles and Imbert 2008] studied the comparison principle for degenerate second-order

MSC2010: primary 35D40, 35J60, 35R09, 47G20, 49N70; secondary 45K05.
Keywords: viscosity solution, integro-PDE, Hamilton–Jacobi–Bellman–Isaacs equation, Perron's method, weak Harnack inequality.

integro-differential equations assuming the nonlocal operators are of Lévy–Itô type and the equations satisfy the coercive assumption. Then G. Barles, E. Chasseigne and C. Imbert [Barles et al. 2008] obtained the existence of viscosity solutions for such integro-differential equations by Perron's method. L. A. Caffarelli and L. Silvestre [2009, Section 5] proved the comparison principle for uniformly elliptic translation-invariant integro-differential equations where the nonlocal operators are of Lévy type. Then existence of viscosity solutions follows, if suitable barriers can be constructed, by Perron's method. Later H. Chang-Lara and G. Davila [2014a, Section 3; 2016b] extended the comparison and existence results of [Caffarelli and Silvestre 2009] to parabolic equations. The existence for (1-1) when I is a nonlocal operator that is not translation-invariant is much more difficult to tackle since we do not have a good comparison principle; see [Mou and Święch 2015], where the authors proved comparison assuming that either a viscosity subsolution or a supersolution is more regular. To our knowledge, the only available results for the existence of solutions for equations that are not translation-invariant are the following. D. Kriventsov [2013, Section 5] studied the existence of viscosity solutions of some uniformly elliptic nonlocal equations. J. Serra [2015b, Section 4] proved the existence of viscosity solutions of uniformly elliptic nonlocal Bellman equations. H. Chang-Lara and D. Kriventsov [2017, Section 5] extended existence results in [Kriventsov 2013] to a class of uniformly parabolic nonlocal equations. In all these proofs, the authors used fixed-point arguments. O. Alvarez and A. Tourin [1996] obtained the existence of viscosity solutions of degenerate parabolic nonlocal equations by Perron's method with a restrictive assumption that the Lévy measures are bounded. The boundedness of Lévy measures allowed them to obtain the comparison principle. The reader can consult [Crandall et al. 1992; Ishii 1987; 1989; Koike 2005] for Perron's method for viscosity solutions of fully nonlinear partial differential equations.

The probability literature on the existence of viscosity solutions of nonlocal Bellman–Isaacs equations is enormous. It is well known that Bellman–Isaacs equations arise when people study differential games, where the equations carry information about the value and strategies of the games. Probabilists represent viscosity solutions of nonlocal Bellman–Isaacs equations as value functions of certain stochastic differential games with jump diffusion via the dynamic programming principle. However, mostly in the probability literature, the nonlocal terms of nonlocal Bellman–Isaacs equations are of Lévy–Itô type and Ω is the whole space $\mathbb{R}^n$. We refer the reader to [Barles et al. 1997; Biswas 2012; Biswas et al. 2010; Buckdahn et al. 2011; Ishikawa 2004; Kharroubi and Pham 2015; Koike and Święch 2013; Øksendal and Sulem 2007; Pham 1998; Soner 1986; 1988; Święch and Zabczyk 2016] for stochastic representation formulas for viscosity solutions of nonlocal Bellman–Isaacs equations.

In Section 3, we adapt to the nonlocal case the approach from [Ishii 1987; 1989; Koike 2005] for obtaining existence of a discontinuous viscosity solution u of (1-1) without using the comparison principle. For applying Perron's method, we need to assume that there exist a continuous viscosity subsolution and a continuous supersolution of (1-1) and both satisfy the boundary condition. Since (1-1) involves the nonlocal term, the proof of the existence is more delicate than the PDE case.

In Section 4, we obtain a Hölder estimate for the discontinuous viscosity solution of (1-1) constructed by Perron's method assuming the equation is uniformly elliptic. In most of the literature, the nonlocal operator I is assumed to be uniformly elliptic with respect to a class of linear nonlocal operators of form

(1-3) with kernels K satisfying

$$(2-\sigma)\frac{\lambda}{|z|^{n+\sigma}} \le K(x,z) \le (2-\sigma)\frac{\Lambda}{|z|^{n+\sigma}}, \tag{1-5}$$

where $0 < \lambda \le \Lambda$. Various regularity results were obtained in recent years under the above uniform ellipticity, such as [Caffarelli and Silvestre 2009; 2011a; 2011b; Chang-Lara and Dávila 2014a; 2014b; 2016a; 2016b; Chang-Lara and Kriventsov 2017; Dong and Kim 2013; Jin and Xiong 2015; 2016; Kriventsov 2013; Serra 2015a; 2015b; Silvestre 2006; 2011; Dong and Zhang 2016] for both elliptic and parabolic integro-differential equations. In this paper, we follow [Schwab and Silvestre 2016] to assume a much weaker uniform ellipticity. Roughly speaking, we let I be uniformly elliptic with respect to a larger class of linear nonlocal operators where the kernels K satisfy the right-hand side of (1-5) in an integral sense and the left-hand side of that in a symmetric subset of each annulus domain with positive measure. The main tool we use is the weak Harnack inequality obtained in [Schwab and Silvestre 2016]. With the weak Harnack inequality, we are able to prove the oscillation between the upper- and lower-semicontinuous envelopes of the discontinuous viscosity solution u in the ball B_r is of order r^α for some $\alpha > 0$ and any small $r > 0$. This proves that u is Hölder continuous and thus it is a viscosity solution of (1-1). Recently, L. Silvestre [2016] applied the regularity for nonlocal equations under this weak ellipticity to obtain the regularity for the homogeneous Boltzmann equation without cut-off. We also want to mention that M. Kassmann, M. Rang and R. Schwab [Kassmann et al. 2014] studied Hölder regularity for a class of integro-differential operators with kernels which are positive along some given rays or cone-like sets.

To complete the existence results, we construct continuous sub/supersolutions in both uniformly elliptic and degenerate cases in Section 5. In the uniformly elliptic case, we follow the idea of [Ros-Oton and Serra 2016] to construct appropriate barrier functions. We then use them to construct a subsolution and a supersolution which satisfy the boundary condition. The weak uniform ellipticity and the lower-order terms of I make the proofs more involved. With all these ingredients in hand, we can conclude one of the main results in this manuscript, that (1-1) admits a viscosity solution if I is uniformly elliptic; see Theorem 5.6 in Section 5A. This main result generalizes nearly all the previous existence results for uniformly elliptic integro-differential equations. In the degenerate case, it is natural to construct a sub/supersolution only for (1-2) since we have little information about the nonlocal operator I. Moreover, we need to assume the nonlocal Bellman–Isaacs equation in (1-2) satisfies the coercive assumption, i.e., $c_{ab} \ge \gamma$ for some $\gamma > 0$. The coercive assumption is often made to study uniqueness, existence and regularity of viscosity solutions of degenerate elliptic PDEs and integro-PDEs; see [Barles et al. 2008; Barles and Imbert 2008; Crandall et al. 1992; Ishii 1987; 1989; Ishii and Lions 1990; Jakobsen and Karlsen 2006; Mou 2016; Mou and Święch 2015]. In Section 5B, we obtain a subsolution and a supersolution which satisfy the boundary condition in the degenerate case. The difficulty here lies in giving a degenerate assumption on the kernels which allows us to construct barrier functions. Roughly speaking, we only need to assume that the kernels $K_{ab}(x,\cdot)$ are nondegenerate in the outer-pointing normal direction of the boundary for the points x which are sufficiently close to the boundary. That means we allow our kernels K_{ab} to be degenerate in the whole domain. Then we can conclude the second main result, the existence of a discontinuous

viscosity solution of (1-2), given in Theorem 5.13. If the comparison principle holds for (1-2), we obtain that the discontinuous viscosity solution is a viscosity solution. Finally, we notice that our method could be adapted to the nonlocal parabolic equations for obtaining the corresponding existence results.

2. Notation and definitions

We write B_δ for the open ball centered at the origin with radius $\delta > 0$ and $B_\delta(x) := B_\delta + x$. We set $\Omega_\delta := \{x \in \Omega : \operatorname{dist}(x, \partial\Omega) > \delta\}$ for $\delta > 0$. For each nonnegative integer r and $0 < \alpha \le 1$, we denote by $C^{r,\alpha}(\Omega)$ ($C^{r,\alpha}(\bar\Omega)$) the subspace of $C^{r,0}(\Omega)$ ($C^{r,0}(\bar\Omega)$) consisting of functions whose r-th partial derivatives are locally (uniformly) α-Hölder continuous in Ω. For any $u \in C^{r,\alpha}(\bar\Omega)$, where r is a nonnegative integer and $0 \le \alpha \le 1$, define

$$[u]_{r,\alpha;\Omega} := \begin{cases} \sup_{x\in\Omega, |j|=r} |\partial^j u(x)| & \text{if } \alpha = 0, \\ \sup_{x,y\in\Omega, x\neq y, |j|=r} |\partial^j u(x) - \partial^j u(y)|/|x-y|^\alpha & \text{if } \alpha > 0, \end{cases}$$

and

$$\|u\|_{C^{r,\alpha}(\bar\Omega)} := \begin{cases} \sum_{j=0}^r [u]_{j,0,\Omega} & \text{if } \alpha = 0, \\ \|u\|_{C^{r,0}(\bar\Omega)} + [u]_{r,\alpha;\Omega} & \text{if } \alpha > 0. \end{cases}$$

For simplicity, we use the notation $C^\beta(\Omega)$ ($C^\beta(\bar\Omega)$), where $\beta > 0$, to denote the space $C^{r,\alpha}(\Omega)$ ($C^{r,\alpha}(\bar\Omega)$), where r is the largest integer smaller than β and $\alpha = \beta - r$. The set $C_b^\beta(\Omega)$ consist of functions from $C^\beta(\Omega)$ which are bounded. We write $\mathrm{USC}(\mathbb{R}^n)$ for the space of upper-semicontinuous functions in $\mathbb{R}^n$ and $\mathrm{LSC}(\mathbb{R}^n)$ for the space of lower-semicontinuous functions in $\mathbb{R}^n$.

We will give a definition of viscosity solutions of (1-1). We first state the general assumptions on the nonlocal operator I in (1-1). For any $\delta > 0$, $r, s \in \mathbb{R}$, $x, x_k \in \Omega$, $\varphi, \varphi_k, \psi \in C^2(B_\delta(x)) \cap L^\infty(\mathbb{R}^n)$, we assume:

(A0) The function $(x, r) \to I(x, r, \varphi(\cdot))$ is continuous in $B_\delta(x) \times \mathbb{R}$.

(A1) If $x_k \to x$ in Ω, $\varphi_k \to \varphi$ a.e. in $\mathbb{R}^n$, $\varphi_k \to \varphi$ in $C^2(B_\delta(x))$ and $\{\varphi_k\}_k$ is uniformly bounded in $\mathbb{R}^n$, then
$$I(x_k, r, \varphi_k(\cdot)) \to I(x, r, \varphi(\cdot)).$$

(A2) If $r \le s$, then $I(x, r, \varphi(\cdot)) \le I(x, s, \varphi(\cdot))$.

(A3) For any constant C, we have $I(x, r, \varphi(\cdot) + C) = I(x, r, \varphi(\cdot))$.

(A4) If φ touches ψ from above at x, then $I(x, r, \varphi(\cdot)) \le I(x, r, \psi(\cdot))$.

Remark 2.1. If I is uniformly elliptic and satisfies (A0), (A2), then (A0)–(A4) hold for I. See Lemma 4.2.

Remark 2.2. The nonlocal operator I in [Schwab and Silvestre 2016] has only two components, i.e., $(x, \varphi) \to I(x, \varphi(\cdot))$. Here we let our nonlocal operator I have three components and assume (A2)–(A3) hold. This is because we want to let I include the left-hand side of the nonlocal Bellman–Isaacs equation in (1-2) and, moreover, we want to describe the two properties

$$-I_{ab}[x, \varphi + C] + b_{ab}(x) \cdot \nabla(\varphi + C)(x) = -I_{ab}[x, \varphi] + b_{ab}(x) \cdot \nabla\varphi(x),$$
$$c_{ab}(x) r \le c_{ab}(x) s \quad \text{if } r \le s$$

in abstract forms.

Remark 2.3. The left-hand side of the nonlocal Bellman–Isaacs equation in (1-2) satisfies (A0)–(A4) if (1-4) holds and its coefficients K_{ab}, b_{ab}, c_{ab} and f_{ab} are uniformly continuous with respect to x in Ω, uniformly in $a \in \mathcal{A},\ b \in \mathcal{B}$. See [Guillen and Schwab 2016] for when the nonlocal operator I has a min-max structure.

Throughout the paper, we always assume the nonlocal operator I satisfies (A0)–(A4).

Definition 2.4. A bounded function $u \in \mathrm{USC}(\mathbb{R}^n)$ is a viscosity subsolution of $I = 0$ in Ω if whenever $u - \varphi$ has a maximum over $\mathbb{R}^n$ at $x \in \Omega$ for $\varphi \in C_b^2(\mathbb{R}^n)$, then

$$I(x, u(x), \varphi(\cdot)) \le 0.$$

A bounded function $u \in \mathrm{LSC}(\mathbb{R}^n)$ is a viscosity supersolution of $I = 0$ in Ω if whenever $u - \varphi$ has a minimum over $\mathbb{R}^n$ at $x \in \Omega$ for $\varphi \in C_b^2(\mathbb{R}^n)$, then

$$I(x, u(x), \varphi(\cdot)) \ge 0.$$

A bounded function u is a viscosity solution of $I = 0$ in Ω if it is both a viscosity subsolution and viscosity supersolution of $I = 0$ in Ω.

Remark 2.5. In Definition 2.4, all the maximums and minimums can be replaced by strict ones.

Definition 2.6. A bounded function u is a viscosity subsolution of (1-1) if u is a viscosity subsolution of $I = 0$ in Ω and $u \le g$ in Ω^c. A bounded function u is a viscosity supersolution of (1-1) if u is a viscosity supersolution of $I = 0$ in Ω and $u \ge g$ in Ω^c. A bounded function u is a viscosity solution of (1-1) if u is a viscosity subsolution and supersolution of (1-1).

We will use the following notations: if u is a function on Ω, then, for any $x \in \Omega$,

$$u^*(x) = \lim_{r \to 0} \sup\{u(y) : y \in \Omega \text{ and } |y - x| \le r\},$$
$$u_*(x) = \lim_{r \to 0} \inf\{u(y) : y \in \Omega \text{ and } |y - x| \le r\}.$$

One calls u^* the upper-semicontinuous envelope of u and u_* the lower semicontinuous envelope of u.

We then give a definition of discontinuous viscosity solutions of (1-1).

Definition 2.7. A bounded function u is a discontinuous viscosity subsolution of (1-1) if u^* is a viscosity subsolution of (1-1). A bounded function u is a discontinuous viscosity supersolution of (1-1) if u_* is a viscosity supersolution of (1-1). A function u is a discontinuous viscosity solution of (1-1) if it is both a discontinuous viscosity subsolution and a discontinuous viscosity supersolution of (1-1).

Remark 2.8. If u is a discontinuous viscosity solution of (1-1) and u is continuous in $\mathbb{R}^n$, then u is a viscosity solution of (1-1).

3. Perron's method

In this section, we obtain the existence of a discontinuous viscosity solution of (1-1) by Perron's method. We remind you that I satisfies (A0)–(A4).

Lemma 3.1. *Let $\mathcal{F}$ be a family of viscosity subsolutions of $I = 0$ in Ω. Let $w(x) = \sup\{u(x) : u \in \mathcal{F}\}$ in $\mathbb{R}^n$ and assume that $w^*(x) < \infty$ for all $x \in \mathbb{R}^n$. Then w is a discontinuous viscosity subsolution of $I = 0$ in Ω.*

Proof. Suppose that φ is a $C_b^2(\mathbb{R}^n)$ function such that $w^* - \varphi$ has a strict maximum (equal to 0) at $x_0 \in \Omega$ over $\mathbb{R}^n$. We can construct a uniformly bounded sequence of $C^2(\mathbb{R}^n)$ functions $\{\varphi_m\}_m$ such that $\varphi_m = \varphi$ in $B_1(x_0)$, $\varphi \le \varphi_m$ in $\mathbb{R}^n$, $\sup_{x \in B_2^c(x_0)}\{w^*(x) - \varphi_m(x)\} \le -\frac{1}{m}$ and $\varphi_m \to \varphi$ pointwise. Thus, for any positive integer m, we know $w^* - \varphi_m$ has a strict maximum (equal to 0) at x_0 over $\mathbb{R}^n$. Therefore, $\sup_{x \in B_1^c(x_0)}\{w^*(x) - \varphi_m(x)\} = \epsilon_m < 0$. By the definition of w^*, we have, for any $u \in \mathcal{F}$, $\sup_{x \in B_1^c(x_0)}\{u(x) - \varphi_m(x)\} \le \epsilon_m < 0$. Again, by the definition of w^*, we have, for any $\epsilon_m < \epsilon < 0$, there exist $u_\epsilon \in \mathcal{F}$ and $\bar{x}_\epsilon \in B_1(x_0)$ such that $u_\epsilon(\bar{x}_\epsilon) - \varphi(\bar{x}_\epsilon) > \epsilon$. Since $u_\epsilon \in \mathrm{USC}(\mathbb{R}^n)$ and $\varphi_m \in C_b^2(\mathbb{R}^n)$, there exists $x_\epsilon \in B_1(x_0)$ such that $u_\epsilon(x_\epsilon) - \varphi_m(x_\epsilon) = \sup_{x \in \mathbb{R}^n}\{u_\epsilon(x) - \varphi(x)\} \ge u_\epsilon(\bar{x}_\epsilon) - \varphi_m(\bar{x}_\epsilon) > \epsilon$. Since $w^* - \varphi_m$ attains a strict maximum (equal to 0) at x_0 over $\mathbb{R}^n$ and $u \le w^*$ for any $u \in \mathcal{F}$, we have $u_\epsilon(x_\epsilon) \to w^*(x_0)$ and $x_\epsilon \to x_0$ as $\epsilon \to 0^-$. Since u_ϵ is a viscosity subsolution of $I = 0$ in Ω, we have

$$I(x_\epsilon, u_\epsilon(x_\epsilon), \varphi_m(\cdot)) \le 0. \tag{3-1}$$

Since $x_\epsilon \to x_0$, $u_\epsilon(x_\epsilon) \to w^*(x_0)$ as $\epsilon \to 0^-$, $\varphi_m = \varphi$ in $B_1(x_0)$, $\varphi_m \to \varphi$ pointwise, $\{\varphi_m\}_m$ is uniformly bounded, $\varphi \in C_b^2(\mathbb{R}^n)$, (A0) and (A1) hold, we have, letting $\epsilon \to 0^-$ and $m \to +\infty$ in (3-1),

$$I(x_0, w^*(x_0), \varphi(\cdot)) \le 0.$$

Therefore, w is a discontinuous viscosity subsolution of $I = 0$. □

Theorem 3.2. *Let $\underline{u}, \bar{u}$ be bounded continuous functions and be respectively a viscosity subsolution and a viscosity supersolution of $I = 0$ in Ω. Assume moreover that $\bar{u} = \underline{u} = g$ in Ω^c for some bounded continuous function g and $\underline{u} \le \bar{u}$ in $\mathbb{R}^n$. Then*

$$w(x) = \sup_{u \in \mathcal{F}} u(x),$$

where

$$\mathcal{F} = \big\{u \in C^0(\mathbb{R}^n) : \underline{u} \le u \le \bar{u} \text{ in } \mathbb{R}^n \text{ and } u \text{ is a viscosity subsolution of } I = 0 \text{ in } \Omega\big\},$$

is a discontinuous viscosity solution of (1-1).

Proof. Since $\underline{u} \in \mathcal{F}$, we know $\mathcal{F} \neq \varnothing$. Thus, w is well defined, $\underline{u} \le w \le \bar{u}$ in $\mathbb{R}^n$ and $w = \bar{u} = \underline{u}$ in Ω^c. By Lemma 3.1, w is a discontinuous viscosity subsolution of $G = 0$ in Ω. We claim that w is a discontinuous viscosity supersolution of $G = 0$ in Ω. If not, there exist a point $x_0 \in \Omega$ and a function $\varphi \in C_b^2(\mathbb{R}^n)$ such that $w_* - \varphi$ has a strict minimum (equal to 0) at the point x_0 over $\mathbb{R}^n$ and

$$I(x_0, w_*(x_0), \varphi(\cdot)) < -\epsilon_0,$$

where ϵ_0 is a positive constant. Thus, we can find sufficiently small constants $\epsilon_1 > 0$ and $\delta_0 > 0$ such that $B_{\delta_0}(x_0) \subset \Omega$ and there exists a $C_b^2(\mathbb{R}^n)$ function φ_{ϵ_1} satisfying that $\varphi_{\epsilon_1} = \varphi$ in $B_{\delta_0}(x_0)$, $\varphi_{\epsilon_1} \le \varphi$ in $\mathbb{R}^n$, $\inf_{x \in B_{2\delta_0}^c(x_0)}\{w_*(x) - \varphi_{\epsilon_1}(x)\} \ge \epsilon_1 > 0$ and

$$I(x_0, \varphi_{\epsilon_1}(x_0), \varphi_{\epsilon_1}(\cdot)) < -\tfrac{1}{2}\epsilon_0. \tag{3-2}$$

Thus, by (A0), there exists $\delta_1 < \delta_0$ such that, for any $x \in B_{\delta_1}(x_0)$,

$$I(x, \varphi_{\epsilon_1}(x), \varphi_{\epsilon_1}(\cdot)) < -\tfrac{1}{4}\epsilon_0. \tag{3-3}$$

By the definition of w, we have $\varphi_{\epsilon_1} \le w_* \le \bar{u}$ in $\mathbb{R}^n$. If $\varphi_{\epsilon_1}(x_0) = w_*(x_0) = \bar{u}(x_0)$, then $\bar{u} - \varphi_{\epsilon_1}$ has a strict minimum at the point x_0 over $\mathbb{R}^n$. Since $\bar{u}$ is a viscosity supersolution of $I = 0$ in Ω, we have

$$I(x_0, \varphi_{\epsilon_1}(x_0), \varphi_{\epsilon_1}(\cdot)) \ge 0,$$

which contradicts (3-2). Thus, we have $\varphi_{\epsilon_1}(x_0) < \bar{u}(x_0)$. Since $\bar{u}$ and φ_{ϵ_1} are continuous functions in $\mathbb{R}^n$, we have $\varphi_{\epsilon_1}(x) < \bar{u}(x) - \epsilon_2$ in $B_{\delta_2}(x_0)$ for some $0 < \delta_2 < \delta_1$ and $\epsilon_2 > 0$. We define

$$\Delta_r = \sup_{x \in B_r^c(x_0)} \{\varphi_{\epsilon_1}(x) - w_*(x)\}.$$

Since $\inf_{x \in B^c_{2\delta_0}(x_0)}\{w_*(x) - \varphi_{\epsilon_1}(x)\} \ge \epsilon_1 > 0$, $w_* - \varphi_{\epsilon_1}$ has a strict minimum (equal to 0) at the point x_0 and $-w_* \in \mathrm{USC}(\mathbb{R}^n)$, we have $\Delta_r < 0$ for each $r > 0$. For any $y \in \overline{\Omega} \setminus B_r(x_0)$, there exists a function $v_y \in \mathcal{F}$ such that $v_y(y) - \varphi_{\epsilon_1}(y) \ge -\frac{3}{4}\Delta_r$. Since v_y and φ_{ϵ_1} are continuous in $\mathbb{R}^n$, there exists a positive constant δ_y such that $\inf_{x \in B_{\delta_y}(y)}\{v_y(x) - \varphi_{\epsilon_1}(x)\} \ge -\frac{1}{2}\Delta_r$. Since $\overline{\Omega} \setminus B_r(x_0)$ is a compact set in $\mathbb{R}^n$, there exists a finite set $\{y_i\}_{i=1}^{n_r} \subset \overline{\Omega} \setminus B_r(x_0)$ such that $\overline{\Omega} \setminus B_r(x_0) \subset \bigcup_{i=1}^{n_r} B_{\delta_{y_i}}(y_i)$. Thus, we define

$$v_r(x) = \sup_{1 \le i \le n_r} \{v_{y_i}(x)\}, \quad x \in \mathbb{R}^n.$$

By Lemma 3.1 and the definition of v_r, we have $v_r \in \mathcal{F}$ and $\inf_{x \in \overline{\Omega} \setminus B_r(x_0)}\{v_r(x) - \varphi_{\epsilon_1}(x)\} \ge -\frac{1}{2}\Delta_r$. Let α_r be a constant such that $0 < \alpha_r < \frac{1}{2}$ and $-\alpha_r \Delta_r < \epsilon_2$. Thus, we define

$$U(x) = \begin{cases} \max\{\varphi_{\epsilon_1}(x) - \alpha\Delta_r, v_r(x)\}, & x \in B_r(x_0), \\ v_r(x), & x \in B_r^c(x_0), \end{cases}$$

where $0 < r < \delta_2$ and $0 < \alpha < \alpha_r$. By the definition of U, we obtain $U \in C^0(\mathbb{R}^n)$, $\underline{u} \le U \le \bar{u}$ in $\mathbb{R}^n$, and there exists a sequence $\{x_n\}_n \subset B_r(x_0)$ such that $x_n \to x_0$ as $n \to +\infty$ and $U(x_n) > w(x_n)$.

We claim that U is a viscosity subsolution of $I = 0$ in Ω. For any $y \in \Omega$, suppose that there is a function $\psi \in C_b^2(\mathbb{R}^n)$ such that $U - \psi$ has a maximum (equal to 0) at y over $\mathbb{R}^n$. We then divide the proof into two cases.

Case 1: $U(y) = v_r(y)$. Since $v_r \le U \le \psi$ in $\mathbb{R}^n$, we know $v_r - \psi$ has a maximum (equal to 0) at y over $\mathbb{R}^n$. We recall that v_r is a viscosity subsolution of $I = 0$ in Ω. Therefore, we have

$$I(y, U(y), \psi(\cdot)) \le 0.$$

Case 2: $U(y) = \varphi_{\epsilon_1}(y) - \alpha\Delta_r$. We first notice that $y \in B_r(x_0)$. Since $\varphi_{\epsilon_1} - \alpha\Delta_r \le U \le \psi$ in $B_r(x_0)$, then $\varphi_{\epsilon_1} - \alpha\Delta_r - \psi \le 0$ in $B_r(x_0)$. By the definition of U, we have $\psi \ge U = v_r$ in $B_r^c(x_0)$. Thus, $\varphi_{\epsilon_1} - \alpha\Delta_r - \psi \le \varphi_{\epsilon_1} - \alpha\Delta_r - v_r \le \frac{1}{2}\Delta_r - \alpha\Delta_r \le 0$ in $B_r^c(x_0)$. Therefore, we have $\varphi_{\epsilon_1} - \alpha\Delta_r - \psi$ has a maximum (equal to 0) at $y \in B_r(x_0) \subset B_{\delta_1}(x_0)$ over $\mathbb{R}^n$. Since (3-3), (A0), (A3)–(A4) hold, we can choose α independent of ψ and sufficiently small that

$$I(y, \psi(y), \psi(\cdot)) \le I(y, \varphi_{\epsilon_1}(y) - \alpha\Delta_r, \varphi_{\epsilon_1}(\cdot)) \le 0.$$

Based on the two cases, we have that U is a viscosity subsolution of $I = 0$ in Ω. Therefore, $U \in \mathcal{F}$, which contradicts with the definition of w. Thus, w is a discontinuous viscosity supersolution of $I = 0$ in Ω. Therefore, w is a discontinuous viscosity solution of $I = 0$ in Ω. Since $w = g$ in Ω^c, we know w is a discontinuous viscosity solution of (1-1). □

Remark 3.3. Under the assumptions of Theorem 3.2, if the comparison principle holds for (1-1), the discontinuous viscosity solution w is the unique viscosity solution of (1-1). For example, if I is a translation-invariant nonlocal operator, (1-1) admits a unique viscosity solution.

Before applying Theorem 3.2 to (1-2), we now give the precise assumptions on its equation. For any $0 < \lambda \le \Lambda$ and $0 < \sigma < 2$, we consider the family of kernels $K : \mathbb{R}^n \to \mathbb{R}$ satisfying the following assumptions:

(H0) $K(z) \ge 0$ for any $z \in \mathbb{R}^n$.

(H1) For any $\delta > 0$,
$$\int_{B_{2\delta} \setminus B_\delta} K(z)\, dz \le (2-\sigma)\Lambda \delta^{-\sigma}.$$

(H2) For any $\delta > 0$,
$$\left| \int_{B_{2\delta} \setminus B_\delta} zK(z)\, dz \right| \le \Lambda |1-\sigma| \delta^{1-\sigma}.$$

We define our nonlocal operator
$$I_{ab}[x,u] := \int_{\mathbb{R}^n} \delta_z u(x) K_{ab}(x,z)\, dz, \tag{3-4}$$
where
$$\delta_z u(x) := \begin{cases} u(x+z) - u(x) & \text{if } \sigma < 1, \\ u(x+z) - u(x) - \mathbb{1}_{B_1}(z) \nabla u(x) \cdot z & \text{if } \sigma = 1, \\ u(x+z) - u(x) - \nabla u(x) \cdot z & \text{if } \sigma > 1. \end{cases}$$

We consider the following nonlocal Bellman–Isaacs equation
$$\sup_{a \in \mathcal{A}} \inf_{b \in \mathcal{B}} \{ -I_{ab}[x,u] + b_{ab}(x) \cdot \nabla u(x) + c_{ab}(x) u(x) + f_{ab}(x) \} = 0 \quad \text{in } \Omega. \tag{3-5}$$

Corollary 3.4. *Assume that $0 < \sigma < 2$, $b_{ab} \equiv 0$ in Ω if $\sigma < 1$ and $c_{ab} \ge 0$ in Ω. Let $\underline{u}$, $\bar{u}$ be bounded continuous functions and be respectively a viscosity subsolution and a viscosity supersolution of* (3-5)*, where $\{K_{ab}(\cdot, z)\}_{a,b,z}$, $\{b_{ab}\}_{a,b}$, $\{c_{ab}\}_{a,b}$ and $\{f_{ab}\}_{a,b}$ are sets of uniformly continuous functions in Ω, uniformly in $a \in \mathcal{A}$, $b \in \mathcal{B}$, and $\{K_{ab}(x,\cdot) : x \in \Omega,\ a \in \mathcal{A},\ b \in \mathcal{B}\}$ are kernels satisfying* (H0)–(H2)*. Assume moreover that $\bar{u} = \underline{u} = g$ in Ω^c for some bounded continuous function g and $\underline{u} \le \bar{u}$ in $\mathbb{R}^n$. Then*
$$w(x) = \sup_{u \in \mathcal{F}} u(x),$$
where
$$\mathcal{F} = \{ u \in C^0(\mathbb{R}^n) : \underline{u} \le u \le \bar{u} \text{ in } \mathbb{R}^n \text{ and } u \text{ is a viscosity subsolution of } (3\text{-}5) \},$$
is a discontinuous viscosity solution of (1-2)*.*

Proof. We define

$$I(x, r, u(\cdot)) := \sup_{a\in\mathcal{A}} \inf_{b\in\mathcal{B}} \{-I_{ab}[x, u] + b_{ab}(x)\cdot\nabla u(x) + c_{ab}(x)r + f_{ab}(x)\}.$$

It follows from (H1) and (H2) that I_{ab} satisfies (1-4); see Lemma 2.3 in [Schwab and Silvestre 2016]. Then, by (1-4) and uniform continuity of the coefficients, (A0) and (A1) hold. Since $c_{ab} \geq 0$ in Ω, (A2) holds. By (H0) and the structure of I_{ab}, (A3) and (A4) hold. □

4. Hölder estimates

In this section we give Hölder estimates of the discontinuous viscosity solution constructed by Perron's method in the previous section. To obtain Hölder estimates, we will assume that the nonlocal operator I is uniformly elliptic.

We define $\mathcal{L} := \mathcal{L}(\sigma, \lambda, \Lambda)$ to be the class of all the nonlocal operators of form

$$Lu(x) := \int_{\mathbb{R}^n} \delta_z u(x) K(z)\, dz,$$

where K is a kernel satisfying the assumptions (H0)–(H2) given above and the following assumption:

(H3) There exist positive constants λ and μ such that, for any $\delta > 0$, there is a set A_δ satisfying

(i) $A_\delta \subset B_{2\delta} \setminus B_\delta$;
(ii) $A_\delta = -A_\delta$;
(iii) $|A_\delta| \geq \mu |B_{2\delta} \setminus B_\delta|$;
(iv) $K(z) \geq (2-\sigma)\lambda\delta^{-n-\sigma}$ for any $z \in A_\delta$.

We note that we will also write $K \in \mathcal{L}$ if the corresponding nonlocal operator $L \in \mathcal{L}$. We then define the extremal operators

$$M^+_{\mathcal{L}} u(x) := \sup_{L\in\mathcal{L}} Lu(x), \quad M^-_{\mathcal{L}} u(x) := \inf_{L\in\mathcal{L}} Lu(x).$$

We denote by $m : [0, +\infty) \to [0, +\infty)$ a modulus of continuity. We say that the nonlocal operator I is uniformly elliptic if for every $r, s \in \mathbb{R}$, $x \in \Omega$, $\delta > 0$, $\varphi, \psi \in C^2(B_\delta(x)) \cap L^\infty(\mathbb{R}^n)$,

$$\begin{aligned} M^-_{\mathcal{L}}(\varphi-\psi)(x) - C_0|\nabla(\psi-\varphi)(x)| - m(|r-s|) &\leq I(x, r, \psi(\cdot)) - I(x, s, \varphi(\cdot)) \\ &\leq M^+_{\mathcal{L}}(\varphi-\psi)(x) + C_0|\nabla(\psi-\varphi)(x)| + m(|r-s|), \end{aligned}$$

where C_0 is a nonnegative constant such that $C_0 = 0$ if $\sigma < 1$.

Remark 4.1. The definition of uniform ellipticity is different from that in [Schwab and Silvestre 2016] since the nonlocal operator I contains the second component r.

Lemma 4.2. *If the nonlocal operator I is uniformly elliptic and satisfies* (A0), (A2), *then I satisfies* (A0)–(A4).

Proof. Suppose that $\delta > 0$, $x_k \to x$ in Ω, $\varphi_k \to \varphi$ a.e. in $\mathbb{R}^n$, $\varphi_k \to \varphi$ in $C^2(B_\delta(x))$ and $\{\varphi_k\}_k$ is uniformly bounded in $\mathbb{R}^n$. Since I is uniformly elliptic, we have, for any $r \in \mathbb{R}$,

$$\begin{aligned} M_{\mathcal{L}}^-(\varphi-\varphi_k)(x_k) - C_0|\nabla(\varphi_k-\varphi)(x_k)| &\le I(x_k, r, \varphi_k(\cdot)) - I(x_k, r, \varphi(\cdot)) \\ &\le M_{\mathcal{L}}^+(\varphi-\varphi_k)(x_k) + C_0|\nabla(\varphi_k-\varphi)(x_k)|. \end{aligned} \tag{4-1}$$

Since $K \in \mathcal{L}$, we know, by Lemma 2.3 in [Schwab and Silvestre 2016], that K satisfies (1-4). Letting $k \to +\infty$ in (4-1), we have, by (A0),

$$\lim_{k\to+\infty} I(x_k, r, \varphi_k(\cdot)) = I(x, r, \varphi(\cdot)).$$

Therefore, (A1) holds. For any constant C, we have

$$0 = M_{\mathcal{L}}^-(-C) - C_0|\nabla C| \le I(x, r, \varphi(\cdot)+C) - I(x, r, \varphi(\cdot)) \le M_{\mathcal{L}}^+(-C) + C_0|\nabla C| = 0.$$

Thus, (A3) holds. If φ touches a $C^2(B_\delta(x)) \cap L^\infty(\mathbb{R}^n)$ function ψ from above at x, then

$$I(x, r, \varphi) - I(x, r, \psi) \le M_{\mathcal{L}}^+(\psi-\varphi)(x) \le 0.$$

Therefore, (A4) holds. □

The following lemma is an elliptic version of Theorem 6.1 in [Schwab and Silvestre 2016].

Lemma 4.3. *Assume $0 < \sigma_0 \le \sigma < 2$, $C_0, C_1 \ge 0$, and further assume $C_0 = 0$ if $\sigma < 1$. Let u be a viscosity supersolution of*

$$M_{\mathcal{L}}^- u - C_0|\nabla u| = C_1 \quad \text{in } B_2$$

and $u \ge 0$ in $\mathbb{R}^n$. Then there exist constants C and ϵ_3 such that

$$\left(\int_{B_1} u^{\epsilon_3}\, dx\right)^{\frac{1}{\epsilon_3}} \le C(\inf_{B_1} u + C_1),$$

where ϵ_3 and C depend on σ_0, λ, Λ, C_0, n and μ.

The following lemma is a direct corollary of Lemma 4.3.

Corollary 4.4. *Assume $0 < \sigma_0 \le \sigma < 2$, $0 < r < 1$, $C_0, C_1 \ge 0$, and further assume $C_0 = 0$ if $\sigma < 1$. Let u be a viscosity supersolution of*

$$M_{\mathcal{L}}^- u - C_0|\nabla u| = C_1 \quad \text{in } B_{2r}$$

and $u \ge 0$ in $\mathbb{R}^n$. Then there exist constants C and ϵ_3 such that

$$\left(\left|\{u > t\} \cap B_r\right|\right) \le C r^n (u(0) + C_1 r^\sigma)^{\epsilon_3} t^{-\epsilon_3} \quad \text{for any } t \ge 0, \tag{4-2}$$

where ϵ_3 and C depend on σ_0, λ, Λ, C_0, n and μ.

Proof. Now let $v(x) = u(rx)$. By Lemma 2.2 in [Schwab and Silvestre 2016], we have

$$M_{\mathcal{L}}^- v - C_0 r^{\sigma-1}|\nabla v| \le C_1 r^\sigma \quad \text{in } B_2. \tag{4-3}$$

Now we apply Lemma 4.3 to (4-3). Thus, for any $t \geq 0$, we have

$$t\big|\{v > t\} \cap B_1\big|^{\frac{1}{\epsilon_3}} \leq \left(\int_{B_1} v^{\epsilon_3}\, dx\right)^{\frac{1}{\epsilon_3}} \leq C(\inf_{B_1} v + C_1 r^\sigma) \leq C(v(0) + C_1 r^\sigma).$$

Then

$$r^{-n}\big|\{u > t\} \cap B_r\big| \leq \big|\{v > t\} \cap B_1\big| \leq C(v(0) + C_1 r^\sigma)^{\epsilon_3} t^{-\epsilon_3} = C(u(0) + C_1 r^\sigma)^{\epsilon_3} t^{-\epsilon_3}.$$

Therefore, (4-2) holds. □

Then we follow the idea in [Caffarelli and Silvestre 2009] to obtain a Hölder estimate.

Theorem 4.5. *Assume $0 < \sigma_0 \leq \sigma < 2$, $C_0 \geq 0$, and further assume $C_0 = 0$ if $\sigma < 1$. For any $\epsilon > 0$, let $\mathcal{F}$ be a class of bounded continuous functions u in $\mathbb{R}^n$ such that $-\frac{1}{2} \leq u \leq \frac{1}{2}$ in $\mathbb{R}^n$, u is a viscosity subsolution of $M^+_{\mathcal{L}} u + C_0|\nabla u| = -\frac{1}{2}\epsilon$ in B_1 and $w = \sup_{u\in\mathcal{F}} u$ is a discontinuous viscosity supersolution of $M^-_{\mathcal{L}} w - C_0|\nabla w| = \frac{1}{2}\epsilon$ in B_1. Then there exist constants ϵ_4, α and C such that, if $\epsilon < \epsilon_4$,*

$$-C|x|^\alpha \leq w_*(x) - w^*(0) \leq w^*(x) - w_*(0) \leq C|x|^\alpha,$$

where ϵ_4, α and C depend on σ_0, λ, Λ, C_0, n and μ.

Proof. We claim that there exist an increasing sequence $\{m_k\}_k$ and a decreasing sequence $\{M_k\}_k$ such that $M_k - m_k = 8^{-\alpha k}$ and $m_k \leq \inf_{B_{8^{-k}}} w_* \leq \sup_{B_{8^{-k}}} w^* \leq M_k$. We will prove this claim by induction.

For $k = 0$, we choose $m_0 = -\frac{1}{2}$ and $M_0 = \frac{1}{2}$ since $-\frac{1}{2} \leq u \leq \frac{1}{2}$ for any $u \in \mathcal{F}$. Assume that we have the sequences up to m_k and M_k. In $B_{8^{-k-1}}$, we have either

$$\big|\{w_* \geq \tfrac{1}{2}M_k + m_k\} \cap B_{8^{-k-1}}\big| \geq \tfrac{1}{2}|B_{8^{-k-1}}| \tag{4-4}$$

or

$$\big|\{w_* \leq \tfrac{1}{2}M_k + m_k\} \cap B_{8^{-k-1}}\big| \geq \tfrac{1}{2}|B_{8^{-k-1}}|. \tag{4-5}$$

Case 1: (4-4) holds. We define

$$v(x) := \frac{w_*(8^{-k}x) - m_k}{\frac{1}{2}(M_k - m_k)}.$$

Thus, $v \geq 0$ in B_1 and

$$\big|\{v \geq 1\} \cap B_{\frac{1}{8}}\big| \geq \tfrac{1}{2}\big|B_{\frac{1}{8}}\big|.$$

Since w is a discontinuous viscosity supersolution of $M^-_{\mathcal{L}} w - C_0|\nabla w| = \frac{1}{2}\epsilon$ in B_1, we know v is a viscosity supersolution of

$$M^-_{\mathcal{L}} v - C_0 8^{k(1-\sigma)}|\nabla v| = 8^{k(\alpha-\sigma)}\epsilon \quad \text{in } B_{8^k}.$$

We notice that $C_0 = 0$ if $\sigma < 1$ and choose $\alpha < \sigma_0$. Thus, for any $0 < \sigma < 2$, v is a viscosity supersolution of

$$M^-_{\mathcal{L}} v - C_0|\nabla v| = \epsilon \quad \text{in } B_{8^k}.$$

By the inductive assumption, we have, for any $k \geq j \geq 0$,

$$v \geq \frac{m_{k-j} - m_k}{\frac{1}{2}(M_k - m_k)} \geq \frac{m_{k-j} - M_{k-j} + M_k - m_k}{\frac{1}{2}(M_k - m_k)} = 2(1 - 8^{\alpha j}) \quad \text{in } B_{8^j}. \tag{4-6}$$

Moreover, we have

$$v \geq 2 \cdot 8^{\alpha k}\left[-\tfrac{1}{2} - \left(\tfrac{1}{2} - 8^{-\alpha k}\right)\right] = 2(1 - 8^{\alpha k}) \quad \text{in } B^c_{8^k}. \tag{4-7}$$

By (4-6) and (4-7), we have

$$v(x) \geq -2(|8x|^\alpha - 1) \quad \text{for any } x \in B_1^c.$$

We define

$$v^+(x) := \max\{v(x), 0\} \quad \text{and} \quad v^-(x) := -\min\{v(x), 0\}.$$

Since $v \geq 0$ in B_1, we have $v^-(x) = 0$ and $\nabla v^-(x) = 0$ for any $x \in B_1$. By (H1), we can choose α independent of σ and sufficiently small that, for any $x \in B_{\frac{3}{4}}$ and $\sigma_0 \leq \sigma < 2$,

$$\begin{aligned}
M^-_{\mathcal{L}} v^+(x) &\leq M^-_{\mathcal{L}} v(x) + M^+_{\mathcal{L}} v^-(x) \\
&\leq M^-_{\mathcal{L}} v(x) + \sup_{K \in \mathcal{L}} \int_{\mathbb{R}^n} \delta_z v^-(x) K(z)\, dz \\
&\leq M^-_{\mathcal{L}} v(x) + \sup_{K \in \mathcal{L}} \int_{B^c_{\frac{1}{4}} \cap \{v(x+z) < 0\}} v^-(x+z) K(z)\, dz \\
&\leq M^-_{\mathcal{L}} v(x) + \sup_{K \in \mathcal{L}} \int_{B^c_{\frac{1}{4}}} \max\left\{2(|8(x+z)|^\alpha - 1), 0\right\} K(z)\, dz \\
&\leq M^-_{\mathcal{L}} v(x) + 2(2-\sigma)\Lambda \sum_{l=0}^{+\infty} \left(\frac{2^l}{4}\right)^{-\sigma} (2^{(l+4)\alpha} - 1) \\
&\leq M^-_{\mathcal{L}} v(x) + 2^{13}(2-\sigma_0)\Lambda \left(\frac{2^{4(\alpha-\sigma_0)}}{1 - 2^{\alpha-\sigma_0}} - \frac{2^{-4\sigma_0}}{1 - 2^{-\sigma_0}}\right) \leq M^-_{\mathcal{L}} v(x) + \epsilon.
\end{aligned}$$

Therefore, we have

$$M^-_{\mathcal{L}} v^+ - C_0 |\nabla v^+| \leq 2\epsilon \quad \text{in } B_{\frac{3}{4}}.$$

Given any point $x \in B_{1/8}$, we can apply Corollary 4.4 in $B_{1/4}(x)$ to obtain

$$C(v^+(x) + 2\epsilon)^{\epsilon_3} \geq \left|\{v^+ > 1\} \cap B_{\frac{1}{4}}(x)\right| \geq \left|\{v^+ > 1\} \cap B_{\frac{1}{8}}\right| \geq \tfrac{1}{2}\left|B_{\frac{1}{8}}\right|.$$

Thus, we can choose ϵ_4 sufficiently small that $v^+ \geq \epsilon_4$ in $B_{1/8}$ if $\epsilon < \epsilon_4$. Therefore,

$$v(x) = \frac{w_*(8^{-k}x) - m_k}{\frac{1}{2}(M_k - m_k)} \geq \epsilon_4 \quad \text{in } B_{\frac{1}{8}}.$$

If we set $m_{k+1} = m_k + \frac{1}{2}\epsilon_4(M_k - m_k)$ and $M_{k+1} = M_k$, we must have

$$m_{k+1} \le \inf_{B_{8^{-k-1}}} w_* \le \sup_{B_{8^{-k-1}}} w^* \le M_{k+1}.$$

Case 2: (4-5) holds. For any $u \in \mathcal{F}$, we obtain that $u \in C^0(\mathbb{R}^n)$ is a viscosity subsolution of $M_{\mathcal{L}}^+ u + C_0|\nabla u| = -\frac{1}{2}\epsilon$ in B_1 and $u \le w_*$ in $\mathbb{R}^n$. Thus, we have

$$\left|\left\{u \le \tfrac{1}{2}(M_k + m_k)\right\} \cap B_{8^{-k-1}}\right| \ge \tfrac{1}{2}|B_{8^{-k-1}}|.$$

We define

$$v_u(x) := \frac{M_k - u(8^{-k}x)}{\frac{1}{2}(M_k - m_k)}.$$

Thus, $v_u \ge 0$ in B_1 and

$$\left|\{v_u \ge 1\} \cap B_{\frac{1}{8}}\right| \ge \tfrac{1}{2}\left|B_{\frac{1}{8}}\right|.$$

Since u is a viscosity subsolution of $M_{\mathcal{L}}^+ u + C_0|\nabla u| = -\frac{1}{2}\epsilon$ in B_1, then v_u is a viscosity supersolution of

$$M_{\mathcal{L}}^- v_u - C_0|\nabla v_u| = \epsilon \quad \text{in } B_{8^k}.$$

Similar to Case 1, we have, if $\epsilon < \epsilon_4$,

$$v_u(x) = \frac{M_k - u(8^{-k}x)}{\frac{1}{2}(M_k - m_k)} \ge \epsilon_4 \quad \text{in } B_{\frac{1}{8}},$$

which implies

$$u(8^{-k}x) \le M_k - \tfrac{1}{2}\epsilon_4(M_k - m_k) \quad \text{in } B_{\frac{1}{8}}.$$

By the definition of w, we have

$$w^*(8^{-k}x) \le M_k - \tfrac{1}{2}\epsilon_4(M_k - m_k) \quad \text{in } B_{\frac{1}{8}}.$$

If we set $m_{k+1} = m_k$ and $M_{k+1} = M_k - \frac{1}{2}\epsilon_4(M_k - m_k)$, we must have

$$m_{k+1} \le \inf_{B_{8^{-k-1}}} w_* \le \sup_{B_{8^{-k-1}}} w^* \le M_{k+1}.$$

Therefore, in both of the cases, we have $M_{k+1} - m_{k+1} = \left(1 - \frac{1}{2}\epsilon_4\right)8^{-\alpha k}$. We then choose α and ϵ_4 sufficiently small that $\left(1 - \frac{1}{2}\epsilon_4\right) = 8^{-\alpha}$. Thus we have $M_{k+1} - m_{k+1} = 8^{-\alpha(k+1)}$. □

Theorem 4.6. *Assume that $0 < \sigma_0 \le \sigma < 2$ and $I(x, 0, 0)$ is bounded in Ω. Assume that I is uniformly elliptic and satisfies* (A0), (A2). *Let w be the bounded discontinuous viscosity solution of* (1-1) *constructed in Theorem 3.2. Then, for any sufficiently small $\tilde{\delta} > 0$, there exists a constant C such that $w \in C^\alpha(\Omega)$ and*

$$\|w\|_{C^\alpha(\bar{\Omega}_{\tilde{\delta}})} \le C\left(C_2 + m(C_2) + \|I(\cdot, 0, 0)\|_{L^\infty(\Omega)}\right),$$

where α is given in Theorem 4.5, $C_2 := \max\{\|\underline{u}\|_{L^\infty(\mathbb{R}^n)}, \|\bar{u}\|_{L^\infty(\mathbb{R}^n)}\}$ and C depends on σ_0, $\tilde{\delta}$, λ, Λ, C_0, n, μ.

Proof. It is obvious that $\|u\|_{L^\infty(\mathbb{R}^n)} \le C_2$ if $u \in \mathcal{F}$. Since I is uniformly elliptic, we have

$$I(x,0,0) - I(x,u(x),u(\cdot)) \le M_{\mathcal{L}}^+ u(x) + C_0|\nabla u(x)| + m(C_2) \quad \text{in } \Omega.$$

Since u is a viscosity subsolution of $I = 0$ in Ω, we have

$$-m(C_2) - \|I(\cdot,0,0)\|_{L^\infty(\Omega)} \le M_{\mathcal{L}}^+ u + C_0|\nabla u| \quad \text{in } \Omega.$$

Similarly, we have

$$M_{\mathcal{L}}^- w_* - C_0|\nabla w_*| \le m(C_2) + \|I(\cdot,0,0)\|_{L^\infty(\Omega)} \quad \text{in } \Omega.$$

By normalization, the result follows from Theorem 4.5. □

By applying Theorem 4.6 to Bellman–Isaacs equation, we have the following corollary.

Corollary 4.7. *Assume that $0 < \sigma_0 \le \sigma < 2$, $b_{ab} \equiv 0$ in Ω if $\sigma < 1$ and $c_{ab} \ge 0$ in Ω. Assume that $\{K_{ab}(\cdot,z)\}_{a,b,z}$, $\{b_{ab}\}_{a,b}$, $\{c_{ab}\}_{a,b}$, $\{f_{ab}\}_{a,b}$ are sets of uniformly bounded and continuous functions in Ω, uniformly in $a \in \mathcal{A}$, $b \in \mathcal{B}$, and $\{K_{ab}(x,\cdot) : x \in \Omega,\ a \in \mathcal{A},\ b \in \mathcal{B}\}$ are kernels satisfying* (H0)–(H3). *Let w be the bounded discontinuous viscosity solution of* (1-2) *constructed in Corollary 3.4. Then, for any sufficiently small $\tilde{\delta} > 0$, there exists a constant C such that $w \in C^\alpha(\Omega)$ and*

$$\|w\|_{C^\alpha(\bar{\Omega}_{\tilde{\delta}})} \le C\big(C_2 + \sup_{a\in\mathcal{A},b\in\mathcal{B}} \|f_{ab}\|_{L^\infty(\Omega)}\big),$$

where α and C_2 are given in Theorem 4.6 and C depends on σ_0, $\tilde{\delta}$, λ, Λ, $\sup_{a\in\mathcal{A},b\in\mathcal{B}} \|b_{ab}\|_{L^\infty(\Omega)}$, $\sup_{a\in\mathcal{A},b\in\mathcal{B}} \|c_{ab}\|_{L^\infty(\Omega)}$, n, μ.

Remark 4.8. In this section we assume our nonlocal equations satisfy the weak uniform ellipticity introduced in [Schwab and Silvestre 2016] mainly because, to our knowledge, this is the weakest assumption to get the weak Harnack inequality. In fact, our approach to get Hölder continuity of the discontinuous viscosity solution constructed by Perron's method could be applied to more general nonlocal equations as long as the weak Harnack inequality holds for such an equation.

5. Continuous sub/supersolutions

In this section we construct continuous sub/supersolutions in both uniformly elliptic and degenerate cases.

5A. ***Uniformly elliptic case.*** In the uniformly elliptic case, we follow the idea in [Ros-Oton and Serra 2016] to establish barrier functions. We define $v_\alpha(x) = ((x_1 - 1)^+)^\alpha$, where $0 < \alpha < 1$ and $x = (x_1, x_2, \dots, x_n)$.

Lemma 5.1. *Assume that $0 < \sigma < 2$. Then there exists a sufficiently small $\alpha > 0$ such that*

$$M_{\mathcal{L}}^+ v_\alpha((1+r)e_1) \le -\epsilon_5 r^{\alpha-\sigma}$$

for any $r > 0$, where $e_1 = (1,0,\dots,0)$ and ϵ_5 is some positive constant.

Proof. Case 1: $0<\sigma<1$. By Lemma 2.2 in [Schwab and Silvestre 2016], we have, for any $r>0$ and $\alpha>0$,

$$\begin{aligned}
M_{\mathcal{L}}^{+}v_\alpha((1+r)e_1) &= \sup_{K\in\mathcal{L}}\int_{\mathbb{R}^n}\big(v_\alpha((1+r)e_1+z)-v_\alpha((1+r)e_1)\big)K(z)\,dz\\
&= \sup_{K\in\mathcal{L}}\int_{\mathbb{R}^n}\big(((r+z_1)^+)^\alpha-r^\alpha\big)K(z)\,dz\\
&= r^{\alpha-\sigma}\sup_{K\in\mathcal{L}}\int_{\mathbb{R}^n}\big(((1+z_1)^+)^\alpha-1\big)r^{n+\sigma}K(rz)\,dz\\
&= r^{\alpha-\sigma}\sup_{K\in\mathcal{L}}\int_{\mathbb{R}^n}\big(((1+z_1)^+)^\alpha-1\big)K(z)\,dz\\
&\le r^{\alpha-\sigma}\bigg(\sup_{K\in\mathcal{L}}\int_{z_1>-1}\big((1+z_1)^\alpha-1\big)K(z)\,dz-\inf_{K\in\mathcal{L}}\int_{z_1\le-1}K(z)\,dz\bigg).
\end{aligned}$$

By (H3), we have, for any $K\in\mathcal{L}$ and any $\delta>0$, there is a set A_δ satisfying $A_\delta\subset B_{2\delta}\setminus B_\delta$, $A_\delta=-A_\delta$, $|A_\delta|\ge\mu|B_{2\delta}\setminus B_\delta|$ and $K(z)\ge(2-\sigma)\lambda\delta^{-n-\sigma}$ in A_δ. It is obvious that

$$\mu_\delta:=\frac{\big|(B_{2\delta}\setminus B_\delta)\cap\{z:|z_1|<1\}\big|}{|B_{2\delta}\setminus B_\delta|}\to 0\quad\text{as }\delta\to+\infty.$$

Thus, there exists $\delta_3>0$ such that $\mu_\delta<\frac{1}{2}\mu$ if $\delta\ge\delta_3$. Then

$$\frac{\big|\{z:|z_1|\ge1\}\cap A_{\delta_3}\big|}{|B_{2\delta_3}\setminus B_{\delta_3}|}\ge\frac{|A_{\delta_3}|-\big|(B_{2\delta_3}\setminus B_{\delta_3})\cap\{z:|z_1|<1\}\big|}{|B_{2\delta_3}\setminus B_{\delta_3}|}\ge\frac{\mu}{2}.$$

By the symmetry of A_{δ_3}, we have

$$\frac{\big|\{z:z_1\le-1\}\cap A_{\delta_3}\big|}{|B_{2\delta_3}\setminus B_{\delta_3}|}\ge\frac{\mu}{4}.$$

Therefore, we have, for any $K\in\mathcal{L}$,

$$\int_{z_1\le-1}K(z)\,dz\ge\int_{\{z:z_1\le-1\}\cap A_{\delta_3}}K(z)\,dz\ge\frac{(2-\sigma)\lambda\mu}{4}\delta_3^{-n-\sigma}|B_{2\delta_3}\setminus B_{\delta_3}|=:2\epsilon_5. \tag{5-1}$$

By (H1) and (H2), we have, for any $K\in\mathcal{L}$,

$$\begin{aligned}
\int_{z_1>-1}((1+z_1)^\alpha-1)K(z)\,dz &= \int_{\{z:z_1>-1\}\cap B_{\frac12}}+\int_{\{z:z_1>-1\}\cap B_{\frac12}^c}\\
&\le\alpha2^{1-\alpha}\bigg|\int_{B_{\frac12}}zK(z)\,dz\bigg|+\int_{\{z:z_1>-1\}\cap B_{\frac12}^c}((1+z_1)^\alpha-1)K(z)\,dz\\
&\le\alpha2^{1-\alpha}(1-\sigma)\Lambda\sum_{l=0}^{+\infty}\bigg(\frac{1}{2^{l+2}}\bigg)^{1-\sigma}+(2-\sigma)\Lambda\sum_{l=0}^{+\infty}(2^{l-1})^{-\sigma}((1+2^l)^\alpha-1)\\
&\le2\alpha\Lambda\frac{1-\sigma}{1-2^{\sigma-1}}+8\Lambda\bigg(\frac{2^{\alpha-\sigma}}{1-2^{\alpha-\sigma}}-\frac{2^{-\sigma}}{1-2^{-\sigma}}\bigg).
\end{aligned}\tag{5-2}$$

Thus, we have

$$\lim_{\alpha\to 0^+}\sup_{K\in\mathcal{L}}\int_{z_1>-1}((1+z_1)^\alpha-1)K(z)\,dz-\inf_{K\in\mathcal{L}}\int_{z_1\le -1}K(z)\,dz\le -2\epsilon_5.$$

Then there exists a sufficiently small α such that

$$M_{\mathcal{L}}^+ v_\alpha((1+r)e_1)\le -\epsilon_5 r^{\alpha-\sigma}.$$

Case 2: $\sigma=1$. Using (H2), we have, for any $r>0$ and $\alpha>0$,

$$\begin{aligned}
M_{\mathcal{L}}^+ v_\alpha((1+r)e_1) &= \sup_{K\in\mathcal{L}}\int_{\mathbb{R}^n}\big(v_\alpha((1+r)e_1+z)-v_\alpha((1+r)e_1)-\mathbb{1}_{B_1}(z)\nabla v_\alpha((1+r)e_1)\cdot z\big)K(z)\,dz\\
&= \sup_{K\in\mathcal{L}}\int_{\mathbb{R}^n}\big(((r+z_1)^+)^\alpha-r^\alpha-\mathbb{1}_{B_1}(z)\alpha r^{\alpha-1}z_1\big)K(z)\,dz\\
&= r^{\alpha-1}\sup_{K\in\mathcal{L}}\int_{\mathbb{R}^n}\big(((1+z_1)^+)^\alpha-1-\mathbb{1}_{B_{\frac{1}{r}}}(z)\alpha z_1\big)r^{n+1}K(rz)\,dz\\
&= r^{\alpha-1}\sup_{K\in\mathcal{L}}\int_{\mathbb{R}^n}\big(((1+z_1)^+)^\alpha-1-\mathbb{1}_{B_{\frac{1}{2}}}(z)\alpha z_1\big)K(z)\,dz\\
&\le r^{\alpha-1}\bigg(\sup_{K\in\mathcal{L}}\int_{z_1>-1}\big((1+z_1)^\alpha-1-\mathbb{1}_{B_{\frac{1}{2}}}(z)\alpha z_1\big)K(z)\,dz-\inf_{K\in\mathcal{L}}\int_{z_1\le -1}K(z)\,dz\bigg).
\end{aligned}$$

By (H1), we have, for any $K\in\mathcal{L}$,

$$\begin{aligned}
&\int_{z_1>-1}\big((1+z_1)^\alpha-1-\mathbb{1}_{B_{\frac{1}{2}}}(z)\alpha z_1\big)K(z)\,dz\\
&\quad=\int_{\{z:z_1>-1\}\cap B_{\frac{1}{2}}}((1+z_1)^\alpha-1-\alpha z_1)K(z)\,dz+\int_{\{z:z_1>-1\}\cap B_{\frac{1}{2}}^c}((1+z_1)^\alpha-1)K(z)\,dz\\
&\quad\le\alpha(1-\alpha)2^{2-\alpha}\int_{B_{\frac{1}{2}}}|z|^2K(z)\,dz+\int_{\{z:z_1>-1\}\cap B_{\frac{1}{2}}^c}((1+z_1)^\alpha-1)K(z)\,dz\\
&\quad\le\alpha(1-\alpha)2^{2-\alpha}\Lambda\sum_{l=0}^{+\infty}\bigg(\frac{1}{2^{l+2}}\bigg)^{-1}\bigg(\frac{1}{2^{l+1}}\bigg)^2+\Lambda\sum_{l=0}^{+\infty}(2^{l-1})^{-1}((1+2^l)^\alpha-1)\\
&\quad\le 8\alpha\Lambda+4\Lambda\bigg(\frac{2^{\alpha-1}}{1-2^{\alpha-1}}-\frac{2^{-1}}{1-2^{-1}}\bigg).
\end{aligned}$$

Then the rest of proof is similar to Case 1.

Case 3: $1<\sigma<2$. For any $r>0$ and $\alpha>0$, we have

$$\begin{aligned}
M_{\mathcal{L}}^+ v_\alpha((1+r)e_1) &= \sup_{K\in\mathcal{L}}\int_{\mathbb{R}^n}\big(v_\alpha((1+r)e_1+z)-v_\alpha((1+r)e_1)-\nabla v_\alpha((1+r)e_1)\cdot z\big)K(z)\,dz\\
&= \sup_{K\in\mathcal{L}}\int_{\mathbb{R}^n}\big(((r+z_1)^+)^\alpha-r^\alpha-\alpha r^{\alpha-1}z_1\big)K(z)\,dz
\end{aligned}$$

$$= r^{\alpha-\sigma} \sup_{K\in\mathcal{L}} \int_{\mathbb{R}^n} \big(((1+z_1)^+)^\alpha - 1 - \alpha z_1\big) K(z)\, dz$$
$$\le r^{\alpha-\sigma} \left(\sup_{K\in\mathcal{L}} \int_{z_1>-1} \big(((1+z_1)^+)^\alpha - 1 - \alpha z_1\big) K(z)\, dz - \inf_{K\in\mathcal{L}} \int_{z_1\le -1} (1+\alpha z_1) K(z)\, dz \right).$$

Using (5-1) and (H2), we have

$$\inf_{K\in\mathcal{L}} \int_{z_1\le -1} (1+\alpha z_1) K(z)\, dz \ge \inf_{K\in\mathcal{L}} \int_{z_1\le -1} K(z)\, dz - \alpha \sup_{K\in\mathcal{L}} \left| \int_{B_1^c} z K(z)\, dz \right| \ge 2\epsilon_5 - \frac{\alpha\Lambda(\sigma-1)}{1-2^{1-\sigma}}.$$

By (H1) and (H2), we have, for any $K \in \mathcal{L}$,

$$\begin{aligned}
\int_{z_1>-1} \big((1+z_1)^\alpha - 1 - \alpha z_1\big) K(z)\, dz &= \int_{\{z:z_1>-1\}\cap B_{\frac12}} + \int_{\{z:z_1>-1\}\cap B_{\frac12}^c} \\
&\le \alpha(1-\alpha) 2^{2-\alpha} \int_{B_{\frac12}} |z|^2 K(z)\, dz + \alpha \left| \int_{\{z:z_1>-1\}\cap B_{\frac12}^c} z K(z)\, dz \right| \\
&\quad + \int_{\{z:z_1>-1\}\cap B_{\frac12}^c} ((1+z_1)^\alpha - 1) K(z)\, dz \\
&\le \frac{16\alpha(2-\sigma)\Lambda}{1-2^{\sigma-2}} + \frac{2\alpha\Lambda(\sigma-1)}{1-2^{1-\sigma}} + 16(2-\sigma)\Lambda \left(\frac{2^{\alpha-\sigma}}{1-2^{\alpha-\sigma}} - \frac{2^{-\sigma}}{1-2^{-\sigma}} \right).
\end{aligned}$$

Then we have

$$\begin{aligned}
&\lim_{\alpha\to 0^+} \sup_{K\in\mathcal{L}} \int_{z_1>-1} \big(((1+z_1)^+)^\alpha - 1 - \alpha z_1\big) K(z)\, dz - \inf_{K\in\mathcal{L}} \int_{z_1\le -1} (1+\alpha z_1) K(z)\, dz \\
&\quad \le \lim_{\alpha\to 0^+} \frac{16\alpha(2-\sigma)\Lambda}{1-2^{\sigma-2}} + \frac{2\alpha\Lambda(\sigma-1)}{1-2^{1-\sigma}} + 16(2-\sigma)\Lambda \left(\frac{2^{\alpha-\sigma}}{1-2^{\alpha-\sigma}} - \frac{2^{-\sigma}}{1-2^{-\sigma}} \right) - 2\epsilon_5 + \frac{\alpha\Lambda(\sigma-1)}{1-2^{1-\sigma}} \\
&\quad = -2\epsilon_5.
\end{aligned}$$

Similar to Case 1, there exists a sufficiently small α such that

$$M_{\mathcal{L}}^+ v_\alpha((1+r)e_1) \le -\epsilon_5 r^{\alpha-\sigma}. \qquad \square$$

Lemma 5.2. *Assume that $0<\sigma<2$, $C_0 \ge 0$ and further assume $C_0 = 0$ if $\sigma < 1$. Then there are $\alpha > 0$ and $0 < r_0 < 1$ sufficiently small so that the function $u_\alpha(x) := ((|x|-1)^+)^\alpha$ satisfies $M_{\mathcal{L}}^+ u_\alpha + C_0|\nabla u_\alpha| \le -1$ in $\overline{B}_{1+r_0} \setminus \overline{B}_1$.*

Proof. We notice that u_α and $|\nabla|$ are rotation invariant. By Lemma 2.2 in [Schwab and Silvestre 2016], $M_{\mathcal{L}}^+$ is also rotation invariant. Then we only need to prove that $M_{\mathcal{L}}^+ u_\alpha((1+r)e_1) + C_0|\nabla u_\alpha((1+r)e_1)| \le -1$ for any $r \in (0, r_0]$, where r_0 and α are sufficiently small positive constants. Note that, for all $r > 0$, $u_\alpha((1+r)e_1) = v_\alpha((1+r)e_1)$, $\nabla u_\alpha((1+r)e_1) = \nabla v_\alpha((1+r)e_1)$ and

$$\big|\big(|(1+r)e_1 + z| - 1\big)^+ - (r+z_1)^+\big| \le C|z'|^2 \quad \text{for any } z \in B_1,$$

where $z = (z_1, z')$. Therefore, we have

$$0 \le (u_\alpha - v_\alpha)((1+r)e_1 + z) \le \begin{cases} C r^{\alpha-1}|z'|^2, & z \in B_{\frac{r}{2}}, \\ C|z'|^{2\alpha}, & z \in B_1 \setminus B_{\frac{r}{2}}, \\ C|z|^\alpha, & z \in \mathbb{R}^n \setminus B_1. \end{cases}$$

Using (H1), we have, for any $0 < \sigma < 2$ and $L \in \mathcal{L}$,

$$\begin{aligned} 0 &\le L(u_\alpha - v_\alpha)((1+r)e_1) \\ &= \int_{\mathbb{R}^n} (u_\alpha - v_\alpha)((1+r)e_1 + z) K(z)\, dz \\ &\le C\left(\int_{B_{\frac{r}{2}}} r^{\alpha-1}|z'|^2 K(z)\, dz + \int_{B_1 \setminus B_{\frac{r}{2}}} |z'|^{2\alpha} K(z)\, dz + \int_{\mathbb{R}^n \setminus B_1} |z|^\alpha K(z)\, dz \right) \\ &\le C\left(\int_{B_{\frac{r}{2}}} r^{\alpha-1}|z|^2 K(z)\, dz + \int_{B^c_{\frac{r}{2}}} |z|^{2\alpha} K(z)\, dz \right) \le C\Lambda(r^{\alpha-\sigma+1} + r^{2\alpha-\sigma}). \end{aligned}$$

Thus, we have $M^+_{\mathcal{L}}(u_\alpha - v_\alpha)((1+r)e_1) \le C\Lambda(r^{\alpha-\sigma+1} + r^{2\alpha-\sigma})$. Therefore, by Lemma 5.1, there exists a sufficiently small $\alpha > 0$ such that

$$\begin{aligned} M^+_{\mathcal{L}} u_\alpha((1+r)e_1) + C_0|\nabla u_\alpha((1+r)e_1)| & \\ \le M^+_{\mathcal{L}}(u_\alpha - v_\alpha)((1+r)e_1) + M^+_{\mathcal{L}} v_\alpha((1+r)e_1) + C_0|\nabla u_\alpha((1+r)e_1)| & \\ \le C\Lambda(r^{\alpha-\sigma+1} + r^{2\alpha-\sigma}) - \epsilon_5 r^{\alpha-\sigma} + \alpha C_0 r^{\alpha-1}. & \end{aligned}$$

We notice that $\alpha - \sigma + 1 > \alpha - \sigma$, $2\alpha - \sigma > \alpha - \sigma$ and

(i) if $0 < \sigma < 1$, then $C_0 = 0$;

(ii) if $\sigma = 1$, then $\alpha C_0 \to 0$ as $\alpha \to 0$;

(iii) if $1 < \sigma < 2$, then $\alpha - 1 > \alpha - \sigma$.

Thus, there exist sufficiently small $0 < r_0 < 1$ such that we have, for any $r \in (0, r_0]$,

$$M^+_{\mathcal{L}} u_\alpha((1+r)e_1) + C_0|\nabla u_\alpha((1+r)e_1)| \le -1. \tag{5-3}$$

This completes the proof. □

In the rest of this section, we assume that Ω satisfies the uniform exterior ball condition, i.e., there is a constant $r_\Omega > 0$ such that, for any $x \in \partial\Omega$ and $0 < r \le r_\Omega$, there exists $y^r_x \in \Omega^c$ satisfying $\overline{B}_r(y^r_x) \cap \overline{\Omega} = \{x\}$. Without loss of generality, we can assume that $r_\Omega < 1$. Since Ω is a bounded domain, there exists a sufficiently large constant $R_0 > 0$ such that $\Omega \subset \{y : |y_1| < R_0\}$.

Remark 5.3. At this stage, we are not sure about whether the exterior ball condition is necessary for the construction of sub/supersolutions. In future work, we plan to construct sub/supersolutions under a weaker assumption on Ω, such as the cone condition.

Lemma 5.4. *Assume that* $0<\sigma<2$, $C_0\geq 0$ *and further assume* $C_0=0$ *if* $\sigma<1$. *There exists an* $\epsilon_7>0$ *such that, for any* $x\in\partial\Omega$ *and* $0<r<r_\Omega$, *there is a continuous function* $\varphi_{x,r}$ *satisfying*

$$\begin{cases}\varphi_{x,r}\equiv 0 & \text{in } \bar{B}_r(y_x^r),\\ \varphi_{x,r}>0 & \text{in } \bar{B}_r^c(y_x^r),\\ \varphi_{x,r}\geq 1 & \text{in } B_{2r}^c(y_x^r),\\ M_{\mathcal{L}}^+\varphi_{x,r}+C_0|\nabla\varphi_{x,r}|\leq -\epsilon_7 & \text{in } \Omega.\end{cases}$$

Proof. We define a uniformly continuous function φ in $\mathbb{R}^n$ such that $1\leq\varphi\leq 2$ and

$$\varphi(y)=1 \quad \text{in } y_1>R_0+1, \qquad \varphi(y)=2 \quad \text{in } y_1\leq R_0.$$

We pick some sufficiently large $C_3>2/r_0^\alpha$ and we define

$$\varphi_{x,r}(y)=\min\left\{\varphi(y), C_3u_\alpha\left(\frac{y-y_x^r}{r}\right)\right\},$$

where α and r_0 are defined in Lemma 5.2. It is easy to verify that $\varphi_{x,r}\equiv 0$ in $\bar{B}_r(y_x^r)$, $\varphi_{x,r}>0$ in $\bar{B}_r^c(y_x^r)$, and $\varphi_{x,r}\geq 1$ in $B_{2r}^c(y_x^r)$. By Lemma 5.2, we have $M_{\mathcal{L}}^+u_\alpha+C_0|\nabla u_\alpha|\leq -1$ in $\bar{B}_{1+r_0}\setminus\bar{B}_1$. It is obvious that, for any $y\in\bar{B}_{(1+r_0)r}(y_x^r)\setminus\bar{B}_r(y_x^r)$, we have

$$\left(M_{\mathcal{L}}^+u_\alpha\left(\frac{\cdot-y_x^r}{r}\right)\right)(y)+C_0r^{1-\sigma}\left|\left(\nabla u_\alpha\left(\frac{\cdot-y_x^r}{r}\right)\right)(y)\right|\leq -r^{-\sigma} \quad \text{for any } 0<r<r_\Omega.$$

Since $C_0=0$ if $0<\sigma<1$, and $0<r<1$, we have

$$\left(M_{\mathcal{L}}^+u_\alpha\left(\frac{\cdot-y_x^r}{r}\right)\right)(y)+C_0\left|\left(\nabla u_\alpha\left(\frac{\cdot-y_x^r}{r}\right)\right)(y)\right|\leq -1 \quad \text{for any } 0<r<r_\Omega.$$

For any $y\in\bar{B}_{(1+(2/C_3)^{1/\alpha})r}(y_x^r)\setminus\bar{B}_r(y_x^r)$, we have $\varphi_{x,r}(y)=C_3u_\alpha((y-y_x^r)/r)$. Suppose that there exists a test function $\psi\in C_b^2(\mathbb{R}^n)$ that touches $\varphi_{x,r}$ from below at y. Thus, ψ/C_3 touches $u_\alpha((\cdot-y_x^r)/r)$ from below at y. Hence, $M_{\mathcal{L}}^+\psi(y)+C_0|\nabla\psi(y)|\leq -C_3$. For any $y\in\Omega\cap\bar{B}^c_{(1+(2/C_3)^{1/\alpha})r}(y_x^r)$, we have $\varphi_{x,r}(y)=\varphi(y)=\max_{\mathbb{R}^n}\varphi_{x,r}=2$. Therefore, for any $0<\sigma<2$, we have

$$\begin{aligned}(M_{\mathcal{L}}^+\varphi_{x,r})(y)+C_0|\nabla\varphi_{x,r}(y)| &= \sup_{K\in\mathcal{L}}\int_{\mathbb{R}^n}(\varphi_{x,r}(y+z)-\varphi_{x,r}(y))K(z)\,dz\\ &= \sup_{K\in\mathcal{L}}\int_{\mathbb{R}^n}(\varphi_{x,r}(y+z)-2)K(z)\,dz\\ &\leq -\inf_{K\in\mathcal{L}}\int_{\{z|z_1>-y_1+R_0+1\}}K(z)\,dz\\ &\leq -\inf_{K\in\mathcal{L}}\int_{\{z|z_1>2R_0+1\}}K(z)\,dz.\end{aligned}$$

By a similar estimate to (5-1), there exists a positive constant ϵ_6 such that, for any $K\in\mathcal{L}$, we have

$$\int_{\{z|z_1>2R_0+1\}}K(z)\,dz\geq\epsilon_6.$$

Then, for any $y \in \Omega \cap \bar{B}^c_{(1+(2/C_3)^{1/\alpha})r}(y^r_x)$, we have

$$M^+_{\mathcal{L}}\varphi_{x,r}(y) + C_0|\nabla\varphi_{x,r}(y)| \le -\epsilon_6. \tag{5-4}$$

Based on the above estimates, if we set $\epsilon_7 = \min\{C_3, \epsilon_6\}$, we have

$$M^+_{\mathcal{L}}\varphi_{x,r} + C_0|\nabla\varphi_{x,r}| \le -\epsilon_7 \quad \text{in } \Omega. \qquad \square$$

Theorem 5.5. *Assume that* $0 < \sigma < 2$, $I(x,0,0)$ *is bounded in* Ω *and* g *is a bounded continuous function in* $\mathbb{R}^n$. *Assume that* I *is uniformly elliptic and satisfies* (A0), (A2). *Then* (1-1) *admits a continuous viscosity supersolution* $\bar{u}$ *and a continuous viscosity subsolution* $\underline{u}$ *and* $\bar{u} = \underline{u} = g$ *in* Ω^c.

Proof. We only prove (1-1) admits a viscosity supersolution $\bar{u}$ and $\bar{u} = g$ in Ω^c. For a viscosity subsolution, the construction is similar. Since I is uniformly elliptic, we have, for any $x \in \Omega$,

$$-m(\|g\|_{L^\infty(\mathbb{R}^n)}) \le I(x, -\|g\|_{L^\infty(\mathbb{R}^n)}, 0) - I(x,0,0) \le m(\|g\|_{L^\infty(\mathbb{R}^n)}).$$

Thus, we have $\|I(\cdot, -\|g\|_{L^\infty(\mathbb{R}^n)}, 0)\|_{L^\infty(\Omega)} < +\infty$. Since g is a continuous function, let ρ_R be a modulus of continuity of g in B_R. Let R_1 be a sufficiently large constant such that $\Omega \subset B_{R_1-1}$. For any $x \in \partial\Omega$, we let

$$u_{x,r} = \rho_{R_1}(3r) + g(x) + \max\left\{2\|g\|_{L^\infty(\mathbb{R}^n)}, \frac{\left\|I(\cdot, -\|g\|_{L^\infty(\mathbb{R}^n)}, 0)\right\|_{L^\infty(\Omega)}}{\epsilon_7}\right\}\varphi_{x,r},$$

where $\varphi_{x,r}$ and ϵ_7 are given in Lemma 5.4. It is obvious that $u_{x,r}(x) = \rho_{R_1}(3r) + g(x)$, $u_{x,r} \ge g$ in $\mathbb{R}^n$ and

$$M^+_{\mathcal{L}}u_{x,r} + C_0|\nabla u_{x,r}| \le -\left\|I(\cdot, -\|g\|_{L^\infty(\mathbb{R}^n)}, 0)\right\|_{L^\infty(\Omega)} \quad \text{in } \Omega.$$

Now we define $\tilde{u} = \inf_{x\in\partial\Omega, 0<r<r_\Omega}\{u_{x,r}\}$. Therefore, $\tilde{u} = g$ in $\partial\Omega$ and $\tilde{u} \ge g$ in $\mathbb{R}^n$. For any $x \in \partial\Omega$ and $y \in \mathbb{R}^n$, we have

$$\begin{aligned} g(y) - g(x) &\le \tilde{u}(y) - \tilde{u}(x) = \tilde{u}(y) - g(x) \\ &\le \rho_{R_1}(3r) + \max\left\{2\|g\|_{L^\infty(\mathbb{R}^n)}, \frac{\left\|I(\cdot, -\|g\|_{L^\infty(\mathbb{R}^n)}, 0)\right\|_{L^\infty(\Omega)}}{\epsilon_7}\right\}\varphi_{x,r}(y) \end{aligned}$$

for any $0 < r < r_\Omega$. Therefore, $\tilde{u}$ is continuous on $\partial\Omega$. For any $y \in \Omega$, we define $d_y = \operatorname{dist}(y, \partial\Omega) > 0$. If $r < \frac{1}{2}d_y$, then we have, for any $z \in B_{d_y/2}(y)$,

$$u_{x,r}(z) = \rho_{R_1}(3r) + g(x) + 2\max\left\{2\|g\|_{L^\infty(\mathbb{R}^n)}, \frac{\left\|I(\cdot, -\|g\|_{L^\infty(\mathbb{R}^n)}, 0)\right\|_{L^\infty(\Omega)}}{\epsilon_7}\right\}, \quad \text{for any } x \in \partial\Omega.$$

Thus, we have, for any $z \in B_{d_y/2}(y)$,

$$\inf_{x\in\partial\Omega, \frac{d_y}{2}<r<r_\Omega}\{u_{x,r}(z) - u_{x,r}(y), 0\} \le \tilde{u}(z) - \tilde{u}(y) \le \sup_{x\in\partial\Omega, \frac{d_y}{2}<r<r_\Omega}\{u_{x,r}(z) - u_{x,r}(y), 0\}.$$

Since $\{u_{x,r}\}_{x\in\partial\Omega, d_y/2<r<r_\Omega}$ has a uniform modulus of continuity, $\tilde{u}$ is continuous in Ω. Therefore, $\tilde{u}$ is a bounded continuous function in $\overline{\Omega}$. By Lemma 3.1, in Ω we have

$$M^+_{\mathcal{L}}\tilde{u} + C_0|\nabla\tilde{u}| \le -\|I(\cdot, -\|g\|_{L^\infty(\mathbb{R}^n)}, 0)\|_{L^\infty(\Omega)}.$$

Now we define

$$\bar{u} := \begin{cases} \tilde{u} & \text{in } \Omega, \\ g & \text{in } \Omega^c. \end{cases}$$

By the properties of $\tilde{u}$, we have $\bar{u}$ is a bounded continuous function in $\mathbb{R}^n$, $\bar{u} = g$ in Ω^c and

$$M_{\mathcal{L}}^+ \bar{u} + C_0|\nabla \bar{u}| \le -\big\| I(\cdot, -\|g\|_{L^\infty(\mathbb{R}^n)}, 0)\big\|_{L^\infty(\Omega)}$$

in Ω. Using (A2) and uniform ellipticity, we have, for any $x \in \Omega$,

$$\begin{aligned} I(x, -\|g\|_{L^\infty(\mathbb{R}^n)}, 0) - I(x, \bar{u}(x), \bar{u}(\cdot)) &\le I(x, \bar{u}(x), 0) - I(x, \bar{u}(x), \bar{u}(\cdot)) \\ &\le M_{\mathcal{L}}^+ \bar{u}(x) + C_0|\nabla \bar{u}(x)| \le -\big\| I(\cdot, -\|g\|_{L^\infty(\mathbb{R}^n)}, 0)\big\|_{L^\infty(\Omega)}. \end{aligned}$$

Thus, $I(x, \bar{u}(x), \bar{u}(\cdot)) \ge 0$ in Ω. □

Now we have enough ingredients to conclude:

Theorem 5.6. *Let Ω be a bounded domain satisfying the uniform exterior ball condition. Assume that $0 < \sigma < 2$, $I(x, 0, 0)$ is bounded in Ω and g is a bounded continuous function. Assume that I is uniformly elliptic and satisfies* (A0), (A2). *Then* (1-1) *admits a viscosity solution u.*

Proof. The result follows from Theorems 3.2, 4.6 and 5.5. □

Corollary 5.7. *Let Ω be a bounded domain satisfying the uniform exterior ball condition. Assume that $0 < \sigma < 2$, $b_{ab} \equiv 0$ in Ω if $\sigma < 1$ and $c_{ab} \ge 0$ in Ω. Assume that g is a bounded continuous function in $\mathbb{R}^n$, $\{K_{ab}(\cdot, z)\}_{a,b,z}$, $\{b_{ab}\}_{a,b}$, $\{c_{ab}\}_{a,b}$, $\{f_{ab}\}_{a,b}$ are sets of uniformly bounded and continuous functions in Ω, uniformly in $a \in \mathcal{A}$, $b \in \mathcal{B}$, and $\{K_{ab}(x, \cdot) : x \in \Omega,\ a \in \mathcal{A},\ b \in \mathcal{B}\}$ are kernels satisfying* (H0)–(H3). *Then* (1-2) *admits a viscosity solution u.*

5B. ***Degenerate case.*** In the degenerate case, it is natural to construct a sub/supersolution only for (1-2) when $c_{ab} \ge \gamma$ for some $\gamma > 0$. Recall that Ω is a bounded domain satisfying the uniform exterior ball condition with a uniform radius r_Ω and, for any $x \in \partial\Omega$ and $0 < r \le r_\Omega$, we have y_x^r is a point satisfying $\overline{B}_r(y_x^r) \cap \overline{\Omega} = \{x\}$. From now on, we will hide the dependence on x for all variables and functions to make the notation simpler. For example, we will let $y^r := y_x^r$. For any $x \in \partial\Omega$, $y \in \Omega$ and $0 < r \le r_\Omega$, we let

$$n := \frac{x - y^r}{|x - y^r|}, \quad n_y^r := \frac{y - y^r}{|y - y^r|}, \quad \text{and} \quad v_\alpha^r(y) := \left(\left(\frac{(y - y^r)\cdot n}{r} - 1\right)^+\right)^\alpha$$

(see Figure 1).

Instead of letting $\{K_{ab}(x, \cdot) : x \in \Omega,\ a \in \mathcal{A},\ b \in \mathcal{B}\}$ satisfy (H3), we let the set of kernels satisfy the following weaker assumption:

($\overline{\text{H3}}$) There exist $C_4 > 0$, $0 < r_1 < r_\Omega$, $\lambda > 0$ and $\mu > 0$ such that, for any $x \in \partial\Omega$, $0 < r < r_1$ and $y \in \Omega \cap B_{2r}(y^r)$, there is a set A_y^r satisfying

(i) $A_y^r \subset \{z : z_{n_y^r} < -rs_y^r\} \cap (B_{C_4 r s_y^r} \setminus B_{r s_y^r})$, where $z_{n_y^r} := z \cdot n_y^r$ and $s_y^r := |y - y^r|/r - 1$;

(ii) $|A_y^r| \ge \mu |B_{r s_y^r}|$;

(iii) $K(y, z) \ge (2 - \sigma)\lambda (r s_y^r)^{-n-\sigma}$ for any $z \in A_y^r$.

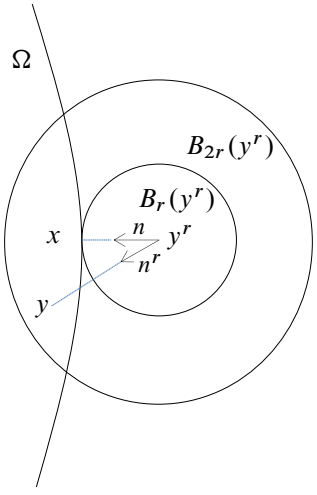

Figure 1. The exterior ball centered at y^r.

Lemma 5.8. *Suppose that* $\{K_{ab}(x,\cdot) : a \in \mathcal{A},\ b \in \mathcal{B},\ x \in \{y \in \Omega : \operatorname{dist}(y, \partial\Omega) < r_1\}\}$ *satisfies* (H3) *for some* $r_1 \in (0, r_\Omega)$. *Then* ($\overline{\text{H3}}$) *holds for the set of kernels.*

Proof. For any $x \in \partial\Omega,\ 0 < r < r_1$ and $y \in \Omega \cap B_{2r}(y^r)$, we define

$$\mu_{C_4} := \frac{\left|(B_{C_4 r s_y^r} \setminus B_{\frac{C_4 r s_y^r}{2}}) \cap \{z : |z_{n_y^r}| \le r s_y^r\}\right|}{|B_{C_4 r s_y^r} \setminus B_{\frac{C_4 r s_y^r}{2}}|}. \tag{5-5}$$

We notice that the right-hand side of (5-5) depends only on C_4. It is obvious that

$$\lim_{C_4 \to +\infty} \mu_{C_4} = 0.$$

By (H3), there exists a set A satisfying

$$A \subset B_{C_4 r s_y^r} \setminus B_{\frac{C_4 r s_y^r}{2}}, \quad A = -A, \quad |A| \ge \mu \left|B_{C_4 r s_y^r} \setminus B_{\frac{C_4 r s_y^r}{2}}\right|,$$

and, for any $z \in A$,

$$K(y,z) \ge (2-\sigma)\lambda\left(\tfrac{1}{2}C_4 r s_y^r\right)^{-n-\sigma} = (2-\sigma)\lambda\left(\tfrac{1}{2}C_4\right)^{-n-\sigma}(r s_y^r)^{-n-\sigma} := (2-\sigma)\bar{\lambda}(r s_y^r)^{-n-\sigma}.$$

There exists a sufficiently large constant $C_4 (\ge 2)$ such that $\mu_{C_4} < \frac{1}{2}\mu$. Then

$$\frac{\left|\{z : |z_{n_y^r}| > r s_y^r\} \cap A\right|}{\left|B_{C_4 r s_y^r} \setminus B_{\frac{C_4 r s_y^r}{2}}\right|} \ge \frac{|A| - \left|(B_{C_4 r s_y^r} \setminus B_{\frac{C_4 r s_y^r}{2}}) \cap \{z : |z_{n_y^r}| \le r s_y^r\}\right|}{\left|B_{C_4 r s_y^r} \setminus B_{\frac{C_4 r s_y^r}{2}}\right|} \ge \frac{\mu}{2}.$$

Let $A_y^r := A \cap \{z : z_{n_y^r} < -r s_y^r\}$. By the symmetry of A, we have

$$|A_y^r| \ge \tfrac{1}{4}\mu \left|B_{C_4 r s_y^r} \setminus B_{\frac{C_4 r s_y^r}{2}}\right| \ge \tfrac{1}{4}\mu |B_{r s_y^r}| := \bar{\mu}|B_{r s_y^r}|.$$

Therefore, ($\overline{\text{H3}}$) holds for the set of kernels with C_4, r_1, $\bar{\lambda}$ and $\bar{\mu}$. □

Lemma 5.9. *Assume that* $0 < \sigma < 2$ *and* $\{K_{ab}(x,\cdot) : x \in \Omega, a \in \mathcal{A}, b \in \mathcal{B}\}$ *are kernels satisfying* (H0)–(H2), ($\overline{\text{H3}}$). *Then there exists a sufficiently small* $\alpha > 0$ *such that, for any* $x \in \partial\Omega$, $0 < r < r_1$ *and* $s \in \{l \in (0,1) : y^r + (1+l)rn \in \Omega\}$, *we have* $I_{ab}[y^r + (1+s)rn, v_\alpha^r] \le -\epsilon_8 r^{-\sigma} s^{\alpha-\sigma}$, *where* ϵ_8 *is some positive constant.*

Proof. We only prove the result for the case $0 < \sigma < 1$. For the rest of cases, the proofs are similar to those in Lemma 5.1. For any $x \in \partial\Omega$, $0 < r < r_1$ and $s \in \{l \in (0,1) : y^r + (1+l)rn \in \Omega\}$, we have

$$\begin{aligned}
I_{ab}[y^r+(1+s)rn, v_\alpha^r] &= \int_{\mathbb{R}^n} \big(v_\alpha^r(y^r+(1+s)rn+z) - v_\alpha^r(y^r+(1+s)rn)\big) K_{ab}(y^r+(1+s)rn, z)\,dz \\
&= \int_{\mathbb{R}^n} \left[\left(\left(s+\frac{\tilde z_n}{r}\right)^+\right)^\alpha - s^\alpha\right] K_{ab}(y^r+(1+s)rn, z)\,dz \\
&= r^{-\sigma} s^{\alpha-\sigma} \int_{\mathbb{R}^n} \big[((1+\tilde z_n)^+)^\alpha - 1\big](rs)^{n+\sigma} K_{ab}(y^r+(1+s)rn, rsz)\,dz \\
&= r^{-\sigma} s^{\alpha-\sigma} \left\{ \int_{\tilde z_n > -1} [(1+\tilde z_n)^\alpha - 1](rs)^{n+\sigma} K_{ab}(y^r+(1+s)rn, rsz)\,dz \right. \\
&\qquad \left. - \int_{\tilde z_n \le -1} (rs)^{n+\sigma} K_{ab}(y^r+(1+s)rn, rsz)\,dz \right\},
\end{aligned}$$

where $\tilde z_n := z \cdot n$. Using ($\overline{\text{H3}}$), we have

$$\begin{aligned}
\int_{\tilde z_n \le -1} (rs)^{n+\sigma} K_{ab}(y^r + (1+s)rn, rsz)\,dz &= (rs)^\sigma \int_{\tilde z_n \le -rs} K_{ab}(y^r + (1+s)rn, z)\,dz \\
&\ge (rs)^\sigma \int_{A^r_{y^r+(1+s)rn}} K_{ab}(y^r + (1+s)rn, z)\,dz \\
&\ge (2-\sigma)\lambda\mu(rs)^{-n}|B_{rs}| := 2\epsilon_8.
\end{aligned}$$

We notice that the kernel $(rs)^{n+\sigma} K_{ab}(y^r + (1+s)rn, rs\,\cdot)$ still satisfies (H1) and (H2). By a similar calculation to (5-2), we have

$$\int_{\tilde z_n > -1} [(1+\tilde z_n)^\alpha - 1](rs)^{n+\sigma} K_{ab}(y^r + (1+s)rn, rsz)\,dz \le \epsilon(\alpha),$$

where $\epsilon(\alpha)$ is a positive constant satisfying that $\epsilon(\alpha) \to 0$ as $\alpha \to 0$. Then there exists a sufficiently small α such that

$$I_{ab}[y^r + (1+s)rn, v_\alpha^r] \le -\epsilon_8 r^{-\sigma} s^{\alpha-\sigma}. \qquad \square$$

Lemma 5.10. *Assume that* $0 < \sigma < 2$, *and* $b_{ab} \equiv 0$ *in* Ω *if* $\sigma < 1$. *Assume that* $\{b_{ab}\}_{a,b}$ *are sets of uniformly bounded functions in* Ω *and* $\{K_{ab}(x,\cdot) : x \in \Omega,\ a \in \mathcal{A},\ b \in \mathcal{B}\}$ *are kernels satisfying* (H0)–(H2), ($\overline{\text{H3}}$). *Then there are* $\alpha > 0$ *and* $0 < s_0 < 1$ *sufficiently small so that, for any* $x \in \partial\Omega$ *and* $0 < r < r_1$, *the function*

$$u_\alpha^r(y) := \left(\left(\frac{|y - y^r|}{r} - 1\right)^+\right)^\alpha$$

satisfies, for any $a \in \mathcal{A}$ and $b \in \mathcal{B}$,

$$-I_{ab}[y, u_\alpha^r] + b_{ab}(y) \cdot \nabla u_\alpha^r(y) \geq 1 \quad \text{in } \Omega \cap \big(\bar{B}_{(1+s_0)r}(y^r) \setminus \bar{B}_r(y^r)\big).$$

Proof. Note that, for all $s > 0$, we have $u_\alpha^r(y^r + (1+s)rn) = v_\alpha^r(y^r + (1+s)rn)$, $\nabla u_\alpha^r(y^r + (1+s)rn) = \nabla v_\alpha^r(y^r + (1+s)rn)$ and

$$\left| \left(\frac{|(1+s)rn + z|}{r} - 1 \right)^+ - \left(s + \frac{\tilde{z}_n}{r} \right)^+ \right| \leq C \frac{|z - \tilde{z}_n|^2}{r^2} \quad \text{for any } z \in B_r.$$

Thus, we have

$$0 \leq (u_\alpha^r - v_\alpha^r)(y^r + (1+s)rn + z) \leq \begin{cases} C s^{\alpha-1} |z - \tilde{z}_n|^2 / r^2, & z \in B_{\frac{rs}{2}}, \\ C |z - \tilde{z}_n|^{2\alpha} / r^{2\alpha}, & z \in B_r \setminus B_{\frac{rs}{2}}, \\ C |z|^\alpha / r^\alpha, & z \in \mathbb{R}^n \setminus B_r. \end{cases}$$

Using (H1), we have, for any $0 < \sigma < 2$, $a \in \mathcal{A}$, $b \in \mathcal{B}$ and $s \in \{l \in (0,1) : y^r + (1+l)rn \in \Omega\}$,

$$\begin{aligned}
0 &\leq I_{ab}[y^r + (1+s)rn, u_\alpha^r - v_\alpha^r] \\
&\leq \int_{\mathbb{R}^n} (u_\alpha^r - v_\alpha^r)(y^r + (1+s)rn + z) K_{ab}(y^r + (1+s)rn, z)\, dz \\
&\leq C \left(\int_{B_{\frac{rs}{2}}} s^{\alpha-1} \frac{|z - \tilde{z}_n|^2}{r^2} K_{ab}(y^r + (1+s)rn, z)\, dz + \int_{B_{\frac{r}{2}} \setminus B_{\frac{rs}{2}}} \frac{|z - \tilde{z}_n|^{2\alpha}}{r^{2\alpha}} K_{ab}(y^r + (1+s)rn, z)\, dz \right. \\
&\qquad \left. + \int_{\mathbb{R}^n \setminus B_r} \frac{|z|^\alpha}{r^\alpha} K_{ab}(y^r + (1+s)rn, z)\, dz \right) \\
&\leq C \left(\int_{B_{\frac{rs}{2}}} s^{\alpha-1} \frac{|z|^2}{r^2} K_{ab}(y^r + (1+s)rn, z)\, dz + \int_{\mathbb{R}^n \setminus B_{\frac{rs}{2}}} \frac{|z|^{2\alpha}}{r^{2\alpha}} K_{ab}(y^r + (1+s)rn, z)\, dz \right) \\
&\leq C \Lambda r^{-\sigma} (s^{\alpha-\sigma+1} + s^{2\alpha-\sigma}).
\end{aligned}$$

By Lemma 5.9, we have

$$\begin{aligned}
-I_{ab}[y^r + (1+s)rn, u_\alpha^r] &\geq -I_{ab}[y^r + (1+s)rn, v_\alpha^r] - I_{ab}[y^r + (1+s)rn, u_\alpha^r - v_\alpha^r] \\
&\geq r^{-\sigma} [\epsilon_8 s^{\alpha-\sigma} - C\Lambda (s^{\alpha-\sigma+1} + s^{2\alpha-\sigma})].
\end{aligned} \tag{5-6}$$

For any $y \in \Omega \cap (B_{2r}(y^r) \setminus \bar{B}_r(y^r))$, we have

$$\begin{aligned}
-I_{ab}[y, u_\alpha^r] &= -\int_{\mathbb{R}^n} \delta_z u_\alpha^r(y) K_{ab}(y, z)\, dz \\
&= -\int_{\mathbb{R}^n} \delta_z u_\alpha^r(y^r + (1 + s_y^r) r n_y^r) K_{ab}(y, z)\, dz \\
&= -\int_{\mathbb{R}^n} \delta_z u_\alpha^r(y^r + (1 + s_y^r) rn) K_{ab}\left(y, \left(\frac{z}{|z|} + n_y^r - n \right) |z| \right) dz.
\end{aligned}$$

Using ($\overline{\text{H3}}$) and a similar estimate to (5-6), we have

$$-I_{ab}[y, u_\alpha^r] \geq r^{-\sigma} \big[\epsilon_8 (s_y^r)^{\alpha-\sigma} - C\Lambda ((s_y^r)^{\alpha-\sigma+1} + (s_y^r)^{2\alpha-\sigma}) \big].$$

By a similar estimate to (5-3), there exists a sufficiently small constant $0 < s_0 < 1$ such that we have, for any $y \in \Omega \cap (\overline{B}_{(1+s_0)r}(y^r) \setminus \overline{B}_r(y^r))$,

$$-I_{ab}[y, u^r_\alpha] + b_{ab}(y) \cdot \nabla u^r_\alpha(y) \geq 1. \qquad \square$$

Lemma 5.11. *Assume that $0 < \sigma < 2$, $b_{ab} \equiv 0$ in Ω if $\sigma < 1$ and $c_{ab} \geq \gamma$ in Ω for some $\gamma > 0$. Assume that $\{K_{ab}(\cdot, z)\}_{a,b,z}$, $\{b_{ab}\}_{a,b}$, $\{c_{ab}\}_{a,b}$, $\{f_{ab}\}_{a,b}$ are sets of uniformly bounded and continuous functions in Ω, uniformly in $a \in \mathcal{A}$, $b \in \mathcal{B}$, and $\{K_{ab}(x, \cdot) : x \in \Omega,\ a \in \mathcal{A},\ b \in \mathcal{B}\}$ are kernels satisfying* (H0)–(H2), ($\overline{\text{H3}}$). *Then, for any $x \in \partial\Omega$ and $0 < r < r_1$, there is a continuous viscosity supersolution ψ_r of* (3-5) *such that $\psi_r \equiv 0$ in $\overline{B}_r(y^r)$, $\psi_r > 0$ in $\overline{B}^c_r(y^r)$ and*

$$\psi_r \equiv \frac{\sup_{a\in\mathcal{A},b\in\mathcal{B}} \|f_{ab}\|_{L^\infty(\Omega)} + 1}{\gamma} \quad \text{in } B^c_{(1+s_0)r}(y^r), \tag{5-7}$$

where s_0 is given by Lemma 5.10.

Proof. Without loss of generality, we assume that $0 < \gamma < 1$. We pick a sufficiently large $C_5 > 0$ that

$$C_5 > \frac{\sup_{a\in\mathcal{A},b\in\mathcal{B}} \|f_{ab}\|_{L^\infty(\Omega)} + 1}{s_0^\alpha \gamma}. \tag{5-8}$$

We then define, for any $x \in \partial\Omega$ and $0 < r < r_1$,

$$\psi_r(y) = \min\left\{ \frac{\sup_{a\in\mathcal{A},b\in\mathcal{B}} \|f_{ab}\|_{L^\infty(\Omega)} + 1}{\gamma}, C_5 u^r_\alpha(y) \right\}.$$

It is easy to verify that $\psi_r \equiv 0$ in $\overline{B}_r(y^r)$, $\psi_r > 0$ in $\overline{B}^c_r(y^r)$ and ψ_r is a continuous function in $\mathbb{R}^n$. Using (5-8), we know that

$$C_5 u^r_\alpha \geq C_5 s_0^\alpha \geq \frac{\sup_{a\in\mathcal{A},b\in\mathcal{B}} \|f_{ab}\|_{L^\infty(\Omega)} + 1}{\gamma} \quad \text{in } B^c_{(1+s_0)r}(y^r).$$

Therefore, (5-7) holds. Since $c_{ab} \geq \gamma > 0$ in Ω, $(\sup_{a\in\mathcal{A},b\in\mathcal{B}} \|f_{ab}\|_{L^\infty(\Omega)} + 1)/\gamma$ is a viscosity supersolution of (3-5) in Ω. By Lemma 5.10 and (5-7), we have, for any $y \in \Omega \cap (\overline{B}_{(1+s_0)r}(y^r) \setminus \overline{B}_r(y^r))$,

$$\begin{aligned} \sup_{a\in\mathcal{A}} \inf_{b\in\mathcal{B}} \{-I_{ab}[y, C_5 u^r_\alpha] + C_5 b_{ab}(x) \cdot \nabla u^r_\alpha(y) + C_5 c_{ab}(x) u^r_\alpha(y) + f_{ab}(y)\} \\ \geq \sup_{a\in\mathcal{A},b\in\mathcal{B}} \|f_{ab}\|_{L^\infty(\Omega)} + 1 + f_{ab}(y) \geq 0. \end{aligned} \tag{5-9}$$

Therefore, ψ_r is a continuous viscosity supersolution of (3-5) in Ω. $\square$

Theorem 5.12. *Assume that $0 < \sigma < 2$, $b_{ab} \equiv 0$ in Ω if $\sigma < 1$ and $c_{ab} \geq \gamma$ in Ω for some $\gamma > 0$. Assume that g is a bounded continuous function in $\mathbb{R}^n$, $\{K_{ab}(\cdot, z)\}_{a,b,z}$, $\{b_{ab}\}_{a,b}$, $\{c_{ab}\}_{a,b}$, $\{f_{ab}\}_{a,b}$ are sets of uniformly bounded and continuous functions in Ω, uniformly in $a \in \mathcal{A}$, $b \in \mathcal{B}$, and $\{K_{ab}(x, \cdot) : x \in \Omega,\ a \in \mathcal{A},\ b \in \mathcal{B}\}$ are kernels satisfying* (H0)–(H2), ($\overline{\text{H3}}$). *Then* (1-2) *admits a continuous viscosity supersolution $\bar{u}$ and a continuous viscosity subsolution $\underline{u}$ and $\bar{u} = \underline{u} = g$ in Ω^c.*

Proof. We only prove (1-2) admits a viscosity supersolution $\bar{u}$ such that $\bar{u} = g$ in Ω^c. Since g is a continuous function, let ρ_R be a modulus of continuity of g in B_R. Let R_1 be a sufficiently large constant such that $\Omega \subset B_{R_1-1}$. For any $x \in \partial\Omega$, we let

$$u_r = \rho_{R_1}(3r) + g(x) + \left(2\|g\|_{L^\infty(\mathbb{R}^n)} \frac{\gamma}{\sup_{a\in\mathcal{A},b\in\mathcal{B}} \|f_{ab}\|_{L^\infty(\Omega)} + 1} + 1\right)\psi_r,$$

where ψ_r is given in Lemma 5.11. Using Lemma 5.11, $u_r(x) = \rho_{R_1}(3r) + g(x)$, $u_r \geq g$ in $\mathbb{R}^n$ and u_r is a continuous viscosity supersolution of (3-5) in Ω. Then the rest of the proof is similar to Theorem 5.5. □

Theorem 5.13. *Let Ω be a bounded domain satisfying the uniform exterior ball condition. Assume that $0 < \sigma < 2$, $b_{ab} \equiv 0$ in Ω if $\sigma < 1$ and $c_{ab} \geq \gamma$ in Ω for some $\gamma > 0$. Assume that g is a bounded continuous function in $\mathbb{R}^n$, $\{K_{ab}(\cdot, z)\}_{a,b,z}$, $\{b_{ab}\}_{a,b}$, $\{c_{ab}\}_{a,b}$, $\{f_{ab}\}_{a,b}$ are sets of uniformly bounded and continuous functions in Ω, uniformly in $a \in \mathcal{A}$, $b \in \mathcal{B}$, and $\{K_{ab}(x, \cdot) : x \in \Omega, a \in \mathcal{A}, b \in \mathcal{B}\}$ are kernels satisfying* (H0)–(H2), ($\overline{\text{H3}}$). *Then* (1-2) *admits a discontinuous viscosity solution u.*

Proof. The result follows from Corollary 3.4 and Theorem 5.12. □

Acknowledgement

We would like to thank the referee for valuable comments which improved the paper.

References

[Alvarez and Tourin 1996] O. Alvarez and A. Tourin, "Viscosity solutions of nonlinear integro-differential equations", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **13**:3 (1996), 293–317. MR Zbl

[Barles and Imbert 2008] G. Barles and C. Imbert, "Second-order elliptic integro-differential equations: viscosity solutions' theory revisited", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **25**:3 (2008), 567–585. MR Zbl

[Barles et al. 1997] G. Barles, R. Buckdahn, and E. Pardoux, "Backward stochastic differential equations and integral-partial differential equations", *Stochastics Stochastics Rep.* **60**:1-2 (1997), 57–83. MR Zbl

[Barles et al. 2008] G. Barles, E. Chasseigne, and C. Imbert, "On the Dirichlet problem for second-order elliptic integro-differential equations", *Indiana Univ. Math. J.* **57**:1 (2008), 213–246. MR Zbl

[Biswas 2012] I. H. Biswas, "On zero-sum stochastic differential games with jump-diffusion driven state: a viscosity solution framework", *SIAM J. Control Optim.* **50**:4 (2012), 1823–1858. MR Zbl

[Biswas et al. 2010] I. H. Biswas, E. R. Jakobsen, and K. H. Karlsen, "Viscosity solutions for a system of integro-PDEs and connections to optimal switching and control of jump-diffusion processes", *Appl. Math. Optim.* **62**:1 (2010), 47–80. MR Zbl

[Buckdahn et al. 2011] R. Buckdahn, Y. Hu, and J. Li, "Stochastic representation for solutions of Isaacs' type integral-partial differential equations", *Stochastic Process. Appl.* **121**:12 (2011), 2715–2750. MR Zbl

[Caffarelli and Silvestre 2009] L. Caffarelli and L. Silvestre, "Regularity theory for fully nonlinear integro-differential equations", *Comm. Pure Appl. Math.* **62**:5 (2009), 597–638. MR Zbl

[Caffarelli and Silvestre 2011a] L. Caffarelli and L. Silvestre, "The Evans–Krylov theorem for nonlocal fully nonlinear equations", *Ann. of Math.* (2) **174**:2 (2011), 1163–1187. MR Zbl

[Caffarelli and Silvestre 2011b] L. Caffarelli and L. Silvestre, "Regularity results for nonlocal equations by approximation", *Arch. Ration. Mech. Anal.* **200**:1 (2011), 59–88. MR Zbl

[Chang-Lara and Dávila 2014a] H. A. Chang-Lara and G. Dávila, "Regularity for solutions of non local parabolic equations", *Calc. Var. Partial Differential Equations* **49**:1-2 (2014), 139–172. MR Zbl

[Chang-Lara and Dávila 2014b] H. A. Chang-Lara and G. Dávila, "Regularity for solutions of nonlocal parabolic equations, II", *J. Differential Equations* **256**:1 (2014), 130–156. MR Zbl

[Chang-Lara and Dávila 2016a] H. A. Chang-Lara and G. Dávila, "$C^{\sigma,\alpha}$ estimates for concave, non-local parabolic equations with critical drift", *J. Integral Equations Appl.* **28**:3 (2016), 373–394. MR Zbl

[Chang-Lara and Dávila 2016b] H. A. Chang-Lara and G. Dávila, "Hölder estimates for non-local parabolic equations with critical drift", *J. Differential Equations* **260**:5 (2016), 4237–4284. MR Zbl

[Chang-Lara and Kriventsov 2017] H. A. Chang-Lara and D. Kriventsov, "Further time regularity for fully non-linear parabolic equations", *Comm. Pure Appl. Math.* **70**:5 (2017), 950–977. MR

[Crandall et al. 1992] M. G. Crandall, H. Ishii, and P.-L. Lions, "User's guide to viscosity solutions of second order partial differential equations", *Bull. Amer. Math. Soc.* (*N.S.*) **27**:1 (1992), 1–67. MR Zbl

[Dong and Kim 2013] H. Dong and D. Kim, "Schauder estimates for a class of non-local elliptic equations", *Discrete Contin. Dyn. Syst.* **33**:6 (2013), 2319–2347. MR Zbl

[Dong and Zhang 2016] H. Dong and H. Zhang, "On Schauder estimates for a class of nonlocal fully nonlinear parabolic equations", preprint, 2016. arXiv

[Guillen and Schwab 2016] N. Guillen and R. W. Schwab, "Min-max formulas for nonlocal elliptic operators", preprint, 2016. arXiv

[Ishii 1987] H. Ishii, "Perron's method for Hamilton–Jacobi equations", *Duke Math. J.* **55**:2 (1987), 369–384. MR Zbl

[Ishii 1989] H. Ishii, "On uniqueness and existence of viscosity solutions of fully nonlinear second-order elliptic PDEs", *Comm. Pure Appl. Math.* **42**:1 (1989), 15–45. MR Zbl

[Ishii and Lions 1990] H. Ishii and P.-L. Lions, "Viscosity solutions of fully nonlinear second-order elliptic partial differential equations", *J. Differential Equations* **83**:1 (1990), 26–78. MR Zbl

[Ishikawa 2004] Y. Ishikawa, "Optimal control problem associated with jump processes", *Appl. Math. Optim.* **50**:1 (2004), 21–65. MR Zbl

[Jakobsen and Karlsen 2006] E. R. Jakobsen and K. H. Karlsen, "A 'maximum principle for semicontinuous functions' applicable to integro-partial differential equations", *NoDEA Nonlinear Differential Equations Appl.* **13**:2 (2006), 137–165. MR Zbl

[Jin and Xiong 2015] T. Jin and J. Xiong, "Schauder estimates for solutions of linear parabolic integro-differential equations", *Discrete Contin. Dyn. Syst.* **35**:12 (2015), 5977–5998. MR Zbl

[Jin and Xiong 2016] T. Jin and J. Xiong, "Schauder estimates for nonlocal fully nonlinear equations", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **33**:5 (2016), 1375–1407. MR Zbl

[Kassmann et al. 2014] M. Kassmann, M. Rang, and R. W. Schwab, "Integro-differential equations with nonlinear directional dependence", *Indiana Univ. Math. J.* **63**:5 (2014), 1467–1498. MR Zbl

[Kharroubi and Pham 2015] I. Kharroubi and H. Pham, "Feynman-Kac representation for Hamilton-Jacobi-Bellman IPDE", *Ann. Probab.* **43**:4 (2015), 1823–1865. MR Zbl

[Koike 2005] S. Koike, "Perron's method for L^p-viscosity solutions", *Saitama Math. J.* **23** (2005), 9–28. MR Zbl

[Koike and Święch 2013] S. Koike and A. Święch, "Representation formulas for solutions of Isaacs integro-PDE", *Indiana Univ. Math. J.* **62**:5 (2013), 1473–1502. MR Zbl

[Kriventsov 2013] D. Kriventsov, "$C^{1,\alpha}$ interior regularity for nonlinear nonlocal elliptic equations with rough kernels", *Comm. Partial Differential Equations* **38**:12 (2013), 2081–2106. MR Zbl

[Mou 2016] C. Mou, "Semiconcavity of viscosity solutions for a class of degenerate elliptic integro-differential equations in $\mathbb{R}^n$", *Indiana Univ. Math. J.* **65**:6 (2016), 1891–1920. MR Zbl

[Mou and Święch 2015] C. Mou and A. Święch, "Uniqueness of viscosity solutions for a class of integro-differential equations", *NoDEA Nonlinear Differential Equations Appl.* **22**:6 (2015), 1851–1882. MR Zbl

[Øksendal and Sulem 2007] B. Øksendal and A. Sulem, *Applied stochastic control of jump diffusions*, 2nd ed., Springer, 2007. MR Zbl

[Pham 1998] H. Pham, "Optimal stopping of controlled jump diffusion processes: a viscosity solution approach", *J. Math. Systems Estim. Control* **8**:1 (1998), art. id. 42281. MR Zbl

[Ros-Oton and Serra 2016] X. Ros-Oton and J. Serra, "Boundary regularity for fully nonlinear integro-differential equations", *Duke Math. J.* **165**:11 (2016), 2079–2154. MR Zbl

[Schwab and Silvestre 2016] R. W. Schwab and L. Silvestre, "Regularity for parabolic integro-differential equations with very irregular kernels", *Anal. PDE* **9**:3 (2016), 727–772. MR Zbl

[Serra 2015a] J. Serra, "$C^{\sigma+\alpha}$ regularity for concave nonlocal fully nonlinear elliptic equations with rough kernels", *Calc. Var. Partial Differential Equations* **54**:4 (2015), 3571–3601. MR Zbl

[Serra 2015b] J. Serra, "Regularity for fully nonlinear nonlocal parabolic equations with rough kernels", *Calc. Var. Partial Differential Equations* **54**:1 (2015), 615–629. MR Zbl

[Silvestre 2006] L. Silvestre, "Hölder estimates for solutions of integro-differential equations like the fractional Laplace", *Indiana Univ. Math. J.* **55**:3 (2006), 1155–1174. MR Zbl

[Silvestre 2011] L. Silvestre, "On the differentiability of the solution to the Hamilton–Jacobi equation with critical fractional diffusion", *Adv. Math.* **226**:2 (2011), 2020–2039. MR Zbl

[Silvestre 2016] L. Silvestre, "A new regularization mechanism for the Boltzmann equation without cut-off", *Comm. Math. Phys.* **348**:1 (2016), 69–100. MR Zbl

[Soner 1986] H. M. Soner, "Optimal control with state-space constraint, II", *SIAM J. Control Optim.* **24**:6 (1986), 1110–1122. MR Zbl

[Soner 1988] H. M. Soner, "Optimal control of jump-Markov processes and viscosity solutions", pp. 501–511 in *Stochastic differential systems, stochastic control theory and applications* (Minneapolis, MN, 1986), IMA Vol. Math. Appl. **10**, Springer, 1988. MR Zbl

[Święch and Zabczyk 2016] A. Święch and J. Zabczyk, "Integro-PDE in Hilbert spaces: existence of viscosity solutions", *Potential Anal.* **45**:4 (2016), 703–736. MR Zbl

Received 24 Nov 2016. Revised 1 Feb 2017. Accepted 24 Apr 2017.

CHENCHEN MOU: muchenchen@math.ucla.edu
Department of Mathematics, UCLA, Los Angeles, CA 90095, United States

ANALYSIS AND PDE
Vol. 10, No. 5, 2017
dx.doi.org/10.2140/apde.2017.10.1255

A SPARSE DOMINATION PRINCIPLE FOR ROUGH SINGULAR INTEGRALS

JOSÉ M. CONDE-ALONSO, AMALIA CULIUC, FRANCESCO DI PLINIO AND YUMENG OU

We prove that bilinear forms associated to the rough homogeneous singular integrals

$$T_\Omega f(x) = \text{p.v.} \int_{\mathbb{R}^d} f(x-y)\Omega\left(\frac{y}{|y|}\right)\frac{\mathrm{d}y}{|y|^d},$$

where $\Omega \in L^q(S^{d-1})$ has vanishing average and $1 < q \le \infty$, and to Bochner–Riesz means at the critical index in $\mathbb{R}^d$ are dominated by sparse forms involving $(1, p)$ averages. This domination is stronger than the weak-L^1 estimates for T_Ω and for Bochner–Riesz means, respectively due to Seeger and Christ. Furthermore, our domination theorems entail as a corollary new sharp quantitative A_p-weighted estimates for Bochner–Riesz means and for homogeneous singular integrals with unbounded angular part, extending previous results of Hytönen, Roncal and Tapiola for T_Ω. Our results follow from a new abstract sparse domination principle which does not rely on weak endpoint estimates for maximal truncations.

1. Introduction and main results

Singular integral operators of Calderón–Zygmund type, which are a priori *signed* and *nonlocal*, can be dominated in norm [Lerner 2013], pointwise [Conde-Alonso and Rey 2016; Lacey 2017; Lerner and Nazarov 2015], or dually [Bernicot et al. 2016; Culiuc et al. 2016a; 2016b] by sparse averaging operators (forms), which are in contrast *positive* and *localized*. For $1 \le p_1, p_2 < \infty$, we define the *sparse* (p_1, p_2)*-averaging form* to be the bisublinear form

$$\mathsf{PSF}_{\mathcal{S};p_1,p_2}(f_1, f_2) := \sum_{Q\in\mathcal{S}} |Q|\langle f_1\rangle_{p_1,Q}\langle f_2\rangle_{p_2,Q}, \quad \langle f\rangle_{p,Q} := |Q|^{-\frac{1}{p}}\|f\mathbf{1}_Q\|_p,$$

associated to a (countable) sparse collection $\mathcal{S}$ of cubes of $\mathbb{R}^d$. The collection $\mathcal{S}$ is η-sparse if there exist $0 < \eta \le 1$ (a number which will not play a relevant role) and measurable, pairwise disjoint sets $\{E_I : I \in \mathcal{S}\}$ such that

$$E_I \subset I, \quad |E_I| \ge \eta|I|.$$

In this article, we prove a sparse domination principle of type

$$|\langle Tf_1, f_2\rangle| \lesssim \sup_{\mathcal{S}} \mathsf{PSF}_{\mathcal{S};p_1,p_2}(f_1, f_2) \tag{1-1}$$

Conde-Alonso was supported in part by ERC Grant 32501 and by MTM-2013-44304-P project. Di Plinio was partially supported by the National Science Foundation under the grants NSF-DMS-1500449 and NSF-DMS-1650810.
MSC2010: primary 42B20; secondary 42B25.
Keywords: positive sparse operators, rough singular integrals, weighted norm inequalities.

for singular integral operators T whose (possible) lack of kernel smoothness forbids the avenue exploited in [Lacey 2017; Lerner 2016]. Our principle, summarized in Theorem C below, can be employed in a rather direct fashion to recover the best known, and sharp, sparse domination results for Dini- and Hörmander-type Calderón–Zygmund operators [Bui et al. 2017; Hytönen et al. 2017; Lacey 2017; Volberg and Zorin-Kranich 2016].

However, the main purpose of our work is to suitably extend (1-1) to the class of rough singular integrals introduced in the seminal paper of Calderón and Zygmund [1956], and further studied, notably, in [Duoandikoetxea and Rubio de Francia 1986; Christ 1988; Christ and Rubio de Francia 1988; Seeger 1996]. Prime examples from this class include the rough homogeneous singular integrals on $\mathbb{R}^d$

$$T_\Omega f(x) = \text{p.v.} \int_{\mathbb{R}^d} f(x-y)\Omega\left(\frac{y}{|y|}\right)\frac{\mathrm{d}y}{|y|^d}, \tag{1-2}$$

with $\Omega \in L^q(S^{d-1})$ having zero average, as well as the critical Bochner–Riesz means in dimension d, defined by the multiplier operator

$$B_\delta f = \mathcal{F}^{-1}\big[\hat{f}(\cdot)(1-|\cdot|^2)_+^\delta\big], \quad \delta = \frac{d-1}{2}. \tag{1-3}$$

For the singular integrals (1-2) no sparse domination results were known prior to this article, although some quantitative weighted estimates were established in the recent works [Hytönen et al. 2017; Pérez et al. 2016]; see below for details. For the Bochner–Riesz means (1-3), the recent results of [Benea et al. 2017] and [Carro and Domingo-Salazar 2016] are far from being optimal at the critical exponent.

The main difficulty encountered by previous approaches in this setting is the following: first, notice that an estimate of the type (1-1) is already stronger than the weak-L^{p_1} bound for T. In particular, if $p_1 = 1$ then (1-1) recovers the weak-L^1 endpoint bound. On the other hand, the preexisting techniques for sparse domination [Benea et al. 2017; Bernicot et al. 2016; Hytönen et al. 2017; Lacey 2017; Lerner 2016] essentially rely on weak-L^p estimates for a grand maximal truncation of the singular integral operator T, but those do not seem attainable in the context, for instance, of [Seeger 1996], as observed in [Lerner 2016]. In fact, the rough singular integrals we consider below are not known to satisfy such an estimate for $p = 1$, and therefore a different approach is required in order to obtain the sparse bounds that we want.

As a corollary of our domination results, we obtain quantitative A_p-weighted estimates for homogeneous singular integrals (1-2) whose angular part belongs to $L^q(S^{d-1})$ for some $1 < q \le \infty$. These are novel, and sharp, when $q < \infty$, while in the case $q = \infty$ we recover the best known result recently proved in [Hytönen et al. 2017] by other methods. Although our result for the Bochner–Riesz means (1-3) seemingly yields the best known quantitative A_p estimates, we do not know whether our results are sharp in this case.

Main results. Our main results consist of estimates for the bilinear forms associated to T_Ω and B_δ by sparse operators involving L^p-averages. The formulation of our first theorem requires the Orlicz–Lorentz norms

$$\|\Omega\|_{L^{q,1}\log L(S^{d-1})} := q\int_0^\infty t\log(\mathrm{e}+t)\big|\{\theta\in S^{d-1} : |\Omega(\theta)| > t\}\big|^{\frac{1}{q}}\frac{\mathrm{d}t}{t}, \quad 1 \le q < \infty.$$

Theorem A. *There exists an absolute dimensional constant $C > 0$ such that the following holds. Let $\Omega \in L^1(S^{d-1})$ have zero average. Then for all $1 < t < \infty$, $f_1 \in L^t(\mathbb{R}^d)$, $f_2 \in L^{t'}(\mathbb{R}^d)$, we have*

$$|\langle T_\Omega f_1, f_2\rangle| \le \frac{Cp}{p-1} \sup_{\mathcal{S}} \mathsf{PSF}_{\mathcal{S};1,p}(f_1, f_2) \begin{cases} \|\Omega\|_{L^{q,1}\log L(S^{d-1})}, & 1 < q < \infty,\ p \ge q', \\ \|\Omega\|_{L^\infty(S^{d-1})}, & 1 < p < \infty. \end{cases}$$

Remark 1.1. To avoid Lorentz norms in the statement, one may recall the continuous embeddings $L^{q+\varepsilon}(S^{d-1}) \hookrightarrow L^{q,1}\log L(S^{d-1}) \hookrightarrow L^q(S^{d-1})$ for all $1 \le q < \infty$ and $\varepsilon > 0$.

Theorem B. *There exists an absolute dimensional constant $C > 0$ such that the following holds. For all $1 < t < \infty$, $f_1 \in L^t(\mathbb{R}^d)$, $f_2 \in L^{t'}(\mathbb{R}^d)$, the critical Bochner–Riesz means* (1-3) *satisfy*

$$|\langle B_\delta f_1, f_2\rangle| \le \frac{Cp}{p-1} \sup_{\mathcal{S}} \mathsf{PSF}_{\mathcal{S};1,p}(f_1, f_2), \quad 1 < p < \infty.$$

The weak-L^1 estimate for T_Ω is the main result of [Seeger 1996], while the same endpoint estimate for (1-3) has been established in [Christ 1988]. Theorems A and B recover such results; see Appendix B for a proof of this implication, which we include for future reference. This is not surprising as the localized estimates for (1-2), (1-3) which are needed to apply our abstract result are a distillation and an improvement of the microlocal techniques of [Seeger 1996] and of the previous works [Christ 1988; Christ and Rubio de Francia 1988], and of the oscillatory integral estimates of [Christ 1988] respectively.

We reiterate that the commonly used techniques for sparse domination, which rely on the weak-L^1 estimate for the maximal truncation of the singular integral operator, fail to be applicable in the context of Theorem A as the maximal truncations of T_Ω in (1-2) are not known to satisfy such an estimate even when $\Omega \in L^\infty(S^{d-1})$ [Grafakos and Stefanov 1999]. Our abstract result, Theorem C, whose statement is more technical and is postponed until Section 2, only relies on the uniform L^2-boundedness (or L^r-boundedness for any r) of the truncated operators, and thus might be considered stronger than the approaches of the mentioned references. See Remark 2.5 for additional discussion on this point.

Theorems A and B give as corollaries a family of quantitative weighted estimates.

Corollary A.1. *If Ω lies in the unit ball of $L^{q,1}\log L(S^{d-1})$ for some $1 < q < \infty$ and has zero average, we have the weighted norm inequalities*

$$\|T_\Omega\|_{L^t(w)\to L^t(w)} \le C_{t,q}[w]_{A_{\frac{t}{q'}}}^{\max\{1,\frac{1}{t-q'}\}}, \quad q' < t < \infty. \tag{1-4}$$

If furthermore $\|\Omega\|_{L^\infty(S^{d-1})} \le 1$,

$$\|T_\Omega\|_{L^t(w)\to L^t(w)} \le C_t[w]_{A_t}^{\frac{1}{t-1}\max\{t,2\}}, \quad 1 < t < \infty. \tag{1-5}$$

Corollary B.1. *Referring to* (1-3), *we have the weighted norm inequalities*

$$\|B_\delta\|_{L^t(w)\to L^t(w)} \le C_t[w]_{A_t}^{\frac{1}{t-1}\max\{t,2\}}, \quad 1 < t < \infty. \tag{1-6}$$

Proof of Corollaries A.1, B.1. To prove (1-4), applying Theorem A for $p = q'$ (strictly speaking, to the adjoint of T_Ω) yields that the bilinear form associated to T_Ω is dominated by

$$\sup_{\mathcal{S}} \mathsf{PSF}_{\mathcal{S};q',1}.$$

The proof of the weighted estimate can then be found, for instance, in [Bernicot et al. 2016, Proposition 6.4]. We prove (1-5), and (1-6) follows via the same argument: below, C denotes a positive absolute constant which may vary between occurrences. Combining the inequality [Di Plinio and Lerner 2014, Proposition 4.1]

$$\langle f\rangle_{1+\varepsilon,Q} \le \langle f\rangle_{1,Q} + C\,\varepsilon \langle \mathrm{M}_{1+\varepsilon} f\rangle_{1,Q},$$

which is valid for all $\varepsilon > 0$, with the estimate of Theorem A for $p = 1+\varepsilon$ we obtain

$$|\langle T_\Omega f_1, f_2\rangle| \le \frac{C}{\varepsilon} \sup_{\mathcal{S}} \mathsf{PSF}_{\mathcal{S};1,1}(f_1, f_2) + C \sup_{\mathcal{S}} \mathsf{PSF}_{\mathcal{S};1,1}(\mathrm{M}_{1+\varepsilon} f_1, f_2), \quad \varepsilon > 0.$$

The above display leads via standard reasoning [Cruz-Uribe et al. 2011; Hytönen et al. 2012; Moen 2012] to the chain of inequalities

$$\begin{aligned}\|T\|_{L^t(w)\to L^t(w)} &\le C_t [w]_{A_t}^{\max\{1,\frac{1}{t-1}\}} \inf_{0<\varepsilon<t-1} \left(\frac{1}{\varepsilon} + \|\mathrm{M}_{1+\varepsilon}\|_{L^t(w)\to L^t(w)}\right)\\ &\le C_t [w]_{A_t}^{\max\{1,\frac{1}{t-1}\}} \inf_{0<\varepsilon<t-1} \left(\frac{1}{\varepsilon} + [w]_{A_{\frac{t}{1+\varepsilon}}}^{\frac{1+\varepsilon}{t-(1+\varepsilon)}}\right) \le C_t [w]_{A_t}^{\frac{1}{t-1}\max\{t,2\}}. \qquad \square\end{aligned}$$

Corollary A.1 is a quantification of the weighted inequalities due to Watson [1990] and Duoandikoetxea [1993]: if $1 < q \le \infty$ and $\Omega \in L^q(S^{d-1})$ then

$$\left.\begin{aligned} &w \in A_{\frac{t}{q'}}, && q' \le t < \infty,\ t \ne 1,\\ &w^{\frac{1}{1-t}} \in A_{\frac{t'}{q'}}, && 1 < t \le q,\ t \ne \infty,\\ &w^{q'} \in A_t, && 1 < t < \infty \end{aligned}\right\} \implies \|T_\Omega\|_{L^t(w)\to L^t(w)} < \infty.$$

Estimate (1-5) was first established by Hytönen, Roncal and Tapiola [Hytönen et al. 2017] via a different two-step technique involving sparse domination for Dini-type kernels, a Littlewood–Paley decomposition along the lines of [Christ and Rubio de Francia 1988] and interpolation with change of measure. In [Pérez et al. 2016], these ideas were extended to obtain A_1 estimates for T_Ω and commutators of T_Ω and BMO symbols. At this time, we do not know whether the power of the Muckenhoupt constant in (1-5) is sharp.

Qualitative A_p-bounds for critical Bochner–Riesz means are classical [Shi and Sun 1992]; see also [Vargas 1996]. On the other hand, Corollary B.1 seems to be the first quantitative A_p estimate for B_δ. We do not know whether the power of the A_p constant in (1-6) is sharp; the construction in [Luque et al. 2015, Corollary 3.1] shows that the optimal power α_p must obey $\alpha_p \ge \max\{1, 1/(p-1)\}$. The article [Benea et al. 2017] contains sparse domination estimates and weighted inequalities for the supercritical regime $0 < \delta' < \delta$ which are not informative in the critical case. An extension of our methods to the supercritical cases will appear in forthcoming work.

Finally, we mention that our argument for (1-5) and (1-6) shows that improvements of powers as those in Corollaries A.1 and B.1 are tied to the blowup rate as $p \to 1^+$ of the main estimate of Theorems A and B.

A remark on the proof and plan of the article. Theorems A and B fall under the scope of the same abstract result, Theorem C, which is stated and proved in Section 2. Theorem C is obtained by means of an iterative scheme reminiscent of the arguments used in [Culiuc et al. 2016a] by three of us to prove a sparse domination estimate for the bilinear Hilbert transform, and later adapted to dyadic and continuous Calderón–Zygmund singular integrals in [Culiuc et al. 2016b]. At each iteration, a decomposition of Calderón–Zygmund type is performed, and the operator itself is decomposed into small scales (scales falling within the exceptional set) which will be estimated at subsequent steps of the iteration, and large scales. The action of the large scales on the good parts is controlled by means of the uniform L^r-bound for the truncations of T. The contribution of the bad, mean zero part under the large scales of the operator is then controlled by means of suitably localized estimates relying on the cancellation of *constant-mean zero* type. We emphasize that the present work shares a perspective based on bilinear forms with other recent papers: [Krause and Lacey 2016; Lacey and Spencer 2017]. The notable difference is that these references, dealing with oscillatory and random discrete singular integrals, use (dilation) symmetry breaking and TT^*, rather than constant-mean zero, as the principal cancellation mechanisms, in accordance with the oscillatory nature of their objects of study.

Section 3 contains localized estimates for kernels of Dini- and Hörmander-type which, besides being of use in later arguments, allow us to reprove the optimal sparse domination results for these classes; see its last subsection for the statements. In Sections 4 and 5 we provide the necessary localized estimates for Theorems A and B respectively. The estimates of Section 4 are a delicate strengthening of the microlocal arguments of [Seeger 1996]. The proof of Theorem B, a re-elaboration along the same lines as the arguments of [Christ 1988], is carried out in Section 5. Although we find it hard to believe that these techniques can be sharpened towards the stronger localized $(1, 1)$ estimate, we have no explicit counterexample for this possibility.

Notation. As is customary, $q' = \frac{q}{q-1}$ denotes the Lebesgue dual exponent to $q \in (1, \infty)$, with the usual extension $1' = \infty,\ \infty' = 1$. We denote the center of a cube $Q \in \mathbb{R}^d$ by c_Q and its sidelength by $\ell(Q)$. We will also adopt the shorthand $s_Q = \log_2 \ell(Q)$. We write

$$\mathrm{M}_p(f)(x) = \sup_{Q \subset \mathbb{R}^d} \langle f \rangle_{p,Q} \mathbf{1}_Q(x)$$

for the p-Hardy Littlewood maximal function. The positive constants implied by the almost inequality sign $\lesssim$ may depend (exponentially) on the dimension d only and may vary from line to line without explicit mention.

2. A sparse domination principle

This section is dedicated to the statement and proof of our sparse domination principle, Theorem C.

The main structural assumptions. Our structural assumptions in Theorem C will be the following. Let $1 < r < \infty$ and Λ be an $L^r(\mathbb{R}^d) \times L^{r'}(\mathbb{R}^d)$-bounded bilinear form whose kernel $K = K(x, y)$ coincides with a function away from the diagonal $\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d : x = y\}$. More precisely, whenever $f_1 \in L^r(\mathbb{R}^d)$,

$f_2 \in L^{r'}(\mathbb{R}^d)$ are compactly and disjointly supported

$$\Lambda(f_1, f_2) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K(x,y) f_1(y)\,\mathrm{d}y\, f_2(x)\,\mathrm{d}x$$

with absolute convergence of the integral. We assume that there exists $1 < q \le \infty$ such that the kernel K of Λ admits the decomposition

$$\begin{gathered} K(x,y) = \sum_{s\in\mathbb{Z}} K_s(x,y), \quad \operatorname{supp} K_s \subset \{(x,y)\in\mathbb{R}^d\times\mathbb{R}^d : x-y \in A_s\}, \\ A_s := \{z\in\mathbb{R}^d : 2^{s-2} < |z| < 2^s\}, \quad [K]_{0,q} := \sup_{s\in\mathbb{Z}} 2^{\frac{sd}{q'}} \sup_{x\in\mathbb{R}^d} \big(\|K_s(x,x+\cdot)\|_q + \|K_s(x+\cdot,x)\|_q\big) < \infty. \end{gathered} \tag{SS}$$

Further, we assume that the truncated forms associated to the above decomposition by

$$\Lambda_\mu^\nu(h_1,h_2) := \int \sum_{\mu<s\le\nu} K_s(x,y) h_1(y) h_2(x)\,\mathrm{d}y\,\mathrm{d}x \quad \mu,\nu \in \mathbb{Z}\cup\{-\infty,\infty\} \tag{2-1}$$

satisfy

$$C_{\mathsf{T}}(r) =: \sup_{\mu<\nu} \big(\|\Lambda_\mu^\nu\|_{L^r(\mathbb{R}^d)\times L^{r'}(\mathbb{R}^d)\to\mathbb{C}}\big) < \infty. \tag{T}$$

Remark 2.1. Under the assumptions (SS) and (T), a standard limiting argument [Stein 1993, Paragraph I.7.2] yields that

$$\Lambda(f_1,f_2) = \langle m f_1, \bar{f_2}\rangle + \lim_{\nu\to\infty} \Lambda_{-\nu}^\nu(f_1,f_2)$$

for some $m \in L^\infty(\mathbb{R}^d)$, whenever $f_1 \in L^r(\mathbb{R}^d)$, $f_2 \in L^{r'}(\mathbb{R}^d)$. It is not hard to see [Lacey and Mena Arias 2017, Lemma 4.7] that

$$|\langle m f_1, f_2\rangle| \lesssim \|m\|_\infty \sup_{\mathcal{S}} \mathsf{PSF}_{\mathcal{S};1,1}(f_1,f_2)$$

so that for the purpose of our Theorem C below we may assume that $m = 0$ in the above equality. For this reason, when $\mu = -\infty$ or $\nu = \infty$ or both, we are allowed to omit the subscript or superscript in (2-1) and simply write Λ^ν or Λ_μ or Λ. Also, when $\mu \ge \nu$, the summation in (2-1) is void, so that $\Lambda_\mu^\nu \equiv 0$.

Localized spaces over stopping collections. A further condition in our abstract theorem will involve local norms associated to *stopping collections* of (dyadic) cubes. Throughout the article, by *dyadic cubes* we refer to the elements of any fixed dyadic lattice $\mathcal{D}$ in $\mathbb{R}^d$.

Let $Q \in \mathcal{D}$ be a fixed dyadic cube in $\mathbb{R}^d$. A collection $\mathcal{Q} \subset \mathcal{D}$ of dyadic cubes is a *stopping collection* with top Q if the elements of $\mathcal{Q}$ are pairwise disjoint and contained in $3Q$,

$$L, L' \in \mathcal{Q},\ L \cap L' \ne \varnothing \implies L = L', \qquad L \in \mathcal{Q} \implies L \subset 3Q, \tag{2-2}$$

and enjoy the further separation properties

$$L, L' \in \mathcal{Q},\ |s_L - s_{L'}| \ge 8 \implies 7L \cap 7L' = \varnothing, \qquad \bigcup_{L\in\mathcal{Q}:\,3L\cap 2Q\ne\varnothing} 9L \subset \bigcup_{L\in\mathcal{Q}} L =: \mathsf{sh}\,\mathcal{Q}; \tag{2-3}$$

the notation $\mathsf{sh}\,\mathcal{Q}$ for the union of the cubes in $\mathcal{Q}$ will also be used below. For $1 \le p \le \infty$, define $\mathcal{Y}_p(\mathcal{Q})$ to be the subspace of $L^p(\mathbb{R}^d)$ of functions satisfying

$$\operatorname{supp} h \subset 3Q, \quad \infty > \|h\|_{\mathcal{Y}_p(\mathcal{Q})} := \begin{cases} \max\bigl\{\|h\mathbf{1}_{\mathbb{R}^d \setminus \mathsf{sh}\,\mathcal{Q}}\|_\infty,\ \sup_{L\in\mathcal{Q}} \inf_{x\in\widehat{L}} \mathrm{M}_p h(x)\bigr\}, & p<\infty, \\ \|h\|_\infty, & p=\infty, \end{cases}$$

where $\widehat{L}$ is the (nondyadic) 2^5-fold dilate of L. We also denote by $\mathcal{X}_p(\mathcal{Q})$ the subspace of $\mathcal{Y}_p(\mathcal{Q})$ of functions satisfying

$$b = \sum_{L\in\mathcal{Q}} b_L, \quad \operatorname{supp} b_L \subset L.$$

Furthermore, we write $b \in \dot{\mathcal{X}}_p(\mathcal{Q})$ if

$$b \in \mathcal{X}_p(\mathcal{Q}), \qquad \int_L b_L = 0 \quad \forall L \in \mathcal{Q}.$$

We will use the notation $\|b\|_{\mathcal{X}_p(\mathcal{Q})}$ for $\|b\|_{\mathcal{Y}_p(\mathcal{Q})}$ when $b \in \mathcal{X}_p(\mathcal{Q})$, and similar notation for $b \in \dot{\mathcal{X}}_p(\mathcal{Q})$. When the stopping collection $\mathcal{Q}$ is clear from the context or during proofs we may omit $(\mathcal{Q})$ from the subscript and simply write $\|\cdot\|_{\mathcal{Y}_p}$ or $\|\cdot\|_{\mathcal{X}_p}$.

Remark 2.2 (Calderón–Zygmund decomposition). There is a natural Calderón–Zygmund decomposition associated to stopping collections. Observe that if $\mathcal{Q}$ is a stopping collection, then

$$\sup_{L\in\mathcal{Q}} \langle h\rangle_{p,L} \le 2^{5d}\|h\|_{\mathcal{Y}_p(\mathcal{Q})}.$$

Therefore, we may decompose $h \in \mathcal{Y}_p(\mathcal{Q})$ as

$$h = g + b, \quad b = \sum_{L\in\mathcal{Q}} b_L \in \dot{\mathcal{X}}_p(\mathcal{Q}), \quad b_L = \biggl(h - \frac{1}{|L|}\int_L h(x)\,\mathrm{d}x\biggr)\mathbf{1}_L$$

such that

$$\|g\|_{\mathcal{Y}_\infty(\mathcal{Q})} \le 2^{5d}\|h\|_{\mathcal{Y}_p(\mathcal{Q})}, \quad \|b\|_{\dot{\mathcal{X}}_p(\mathcal{Q})} \le 2^{5d+1}\|h\|_{\mathcal{Y}_p(\mathcal{Q})}.$$

These are nothing but the usual properties of the Calderón-Zygmund decomposition rewritten in our context.

The statement. Before stating our result, we introduce the notation

$$\Lambda_{\mathcal{Q},\mu,\nu}(h_1,h_2) := \Lambda_\mu^{\min\{s_\mathcal{Q},\nu\}}(h_1\mathbf{1}_Q, h_2) = \Lambda_\mu^{\min\{s_\mathcal{Q},\nu\}}(h_1\mathbf{1}_Q, h_2\mathbf{1}_{3Q}) \tag{2-4}$$

for all dyadic cubes Q; the last equality in (2-4) is a consequence of the assumptions on the support of K_s in (SS). Furthermore, given a stopping collection $\mathcal{Q}$ with top Q, we define the truncated forms

$$\Lambda_{\mathcal{Q},\mu,\nu}(h_1,h_2) := \Lambda_{\mathcal{Q},\mu,\nu}(h_1,h_2) - \sum_{\substack{L\in\mathcal{Q}\\ L\subset\widehat{Q}}} \Lambda_{L,\mu,\nu}(h_1,h_2) = \Lambda_{\mathcal{Q},\mu,\nu}(h_1\mathbf{1}_Q, h_2\mathbf{1}_{3Q}). \tag{2-5}$$

Again, the last equality is due to the support of K_s in (SS). A further consequence of assumptions (SS) and (T) is that the forms $\Lambda_{\mathcal{Q},\mu,\nu}$ satisfy uniform bounds on $\mathcal{Y}_r(\mathcal{Q}) \times \mathcal{Y}_{r'}(\mathcal{Q})$.

Lemma 2.3. *There exists a positive absolute constant ϑ such that*

$$\big|\Lambda_{\mathcal{Q},\mu,\nu}(h_1,h_2)\big| \le 2^{\vartheta d} C_{\mathsf{T}}(r)|Q|\,\|h_1\|_{\mathcal{Y}_r(\mathcal{Q})}\|h_2\|_{\mathcal{Y}_{r'}(\mathcal{Q})}$$

uniformly over all μ, ν, all dyadic cubes Q and stopping collections $\mathcal{Q}$ with top Q.

Proof. We may estimate the first term in the definition (2-5) as follows:

$$|\Lambda_{\mathcal{Q},\mu,\nu}(h_1,h_2)| \le C_{\mathsf{T}}(r)\|h_1\mathbf{1}_Q\|_r\,\|h_2\mathbf{1}_{3Q}\|_{r'} \lesssim C_{\mathsf{T}}(r)|Q|\,\|h_1\|_{\mathcal{Y}_r}\|h_2\|_{\mathcal{Y}_{r'}}. \tag{2-6}$$

Further, using the support condition in (2-4) with L in place of Q and the disjointness property (2-2) in the last step, we obtain

$$\begin{aligned}\sum_{L\in\mathcal{Q}:L\subset Q}|\Lambda_{L,\mu,\nu}(h_1,h_2)| &= \sum_{L\in\mathcal{Q}:L\subset Q}|\Lambda_{L,\mu,\nu}(h_1\mathbf{1}_L,h_2\mathbf{1}_{3L})| \le C_{\mathsf{T}}(r)\sum_{L\in\mathcal{Q}:L\subset Q}\|h_1\mathbf{1}_L\|_r\,\|h_2\mathbf{1}_{3L}\|_{r'}\\ &\lesssim C_{\mathsf{T}}(r)\|h_1\|_{\mathcal{Y}_r}\|h_2\|_{\mathcal{Y}_{r'}}\sum_{L\in\mathcal{Q}}|L| \lesssim C_{\mathsf{T}}(r)|Q|\,\|h_1\|_{\mathcal{Y}_r}\|h_2\|_{\mathcal{Y}_{r'}}.\end{aligned}$$

The proof of the lemma is thus completed by combining (2-6) with the last display. □

Our main theorem hinges upon estimates which are modified versions of the one occurring in Lemma 2.3, when one of the two arguments of $\Lambda_{\mathcal{Q},\mu,\nu}$ belongs to $\mathcal{X}$-type localized spaces.

Theorem C. *There exists a positive absolute constant Θ such that the following holds. Let Λ be a bilinear form satisfying* (SS) *and* (T) *above. Assume that there exist $1 \le p_1, p_2 < \infty$ and a positive constant C_{L} such that the estimates*

$$\begin{aligned}\big|\Lambda_{\mathcal{Q},\mu,\nu}(b,h)\big| &\le C_{\mathsf{L}}|Q|\,\|b\|_{\dot{\mathcal{X}}_{p_1}(\mathcal{Q})}\|h\|_{\mathcal{Y}_{p_2}(\mathcal{Q})},\\ \big|\Lambda_{\mathcal{Q},\mu,\nu}(h,b)\big| &\le C_{\mathsf{L}}|Q|\,\|h\|_{\mathcal{Y}_\infty(\mathcal{Q})}\|b\|_{\dot{\mathcal{X}}_{p_2}(\mathcal{Q})}\end{aligned} \tag{L}$$

hold uniformly over all $\mu,\nu\in\mathbb{Z}$, all dyadic lattices $\mathcal{D}$, all $Q\in\mathcal{D}$ and all stopping collections $\mathcal{Q}\subset\mathcal{D}$ with top Q. Then the estimate

$$\sup_{\mu,\nu\in\mathbb{Z}}\big|\Lambda^{\nu}_{\mu}(f_1,f_2)\big| \le 2^{\Theta d}[C_{\mathsf{T}}(r)+C_{\mathsf{L}}]\sup_{\mathcal{S}}\mathsf{PSF}_{\mathcal{S};p_1,p_2}(f_1,f_2) \tag{2-7}$$

holds for all $f_j\in L^{p_j}(\mathbb{R}^d)$ with compact support, $j=1,2$.

Remark 2.4. By the limiting argument of Remark 2.1, the conclusion (2-7) gives that

$$\big|\Lambda(f_1,f_2)\big| \le 2^{\Theta d}[C_{\mathsf{T}}(r)+C_{\mathsf{L}}]\sup_{\mathcal{S}}\mathsf{PSF}_{\mathcal{S};p_1,p_2}(f_1,f_2) \tag{2-8}$$

when $f_1, f_2\in L^\infty(\mathbb{R}^d)$ with compact support. If we know that Λ extends boundedly to $L^t(\mathbb{R}^d)\times L^{t'}(\mathbb{R}^d)$ for some $1<t<\infty$, another simple limiting argument using the dominated convergence theorem extends (2-8) to all $f_1\in L^t(\mathbb{R}^d)$, $f_2\in L^{t'}(\mathbb{R}^d)$. It is in this last form that Theorem C will be applied to deduce Theorems A and B.

Remark 2.5 (a comparison between sparse domination principles). Theorem C identifies rather clearly the conditions needed for sparse domination of a kernel operator T, namely the adjoint of the bilinear form Λ. Condition (L) is a localized reformulation of the *constant-mean zero* cancellation around which L^p, $p \neq 2$, Calderón–Zygmund theory revolves, and it is essentially a strengthening of the weak-L^{p_j} estimate for T ($j = 1$) and its adjoint ($j = 2$). Further, our assumption of uniform L^r-boundedness of the truncations in (T) is much tamer than requiring L^r-boundedness of the maximal truncations of T. In fact, our theorem can be applied even when no estimates for maximal truncations of T are known.

Of course the exponents p_j enter the sparse domination estimate (2-7), while the exponent r occurring in (T) does not. This is in contrast with the other sparse domination principles occurring in the literature. For instance, in [Lerner 2016, Theorem 4.2], a sparse domination of type (1-1) with exponents $(r, 1)$ is obtained for operators T whose *grand maximal function*

$$\mathcal{M}_T f(x) := \sup_{Q \ni x} \sup_{y \in Q} \big| T(f \mathbf{1}_{\mathbb{R}^d \setminus 3Q})(y) \big|$$

has the weak-L^r-bound for some $r \geq 1$. Notice that $\mathcal{M}_T$ may be as large as the maximal truncation of T.

A further comparison can be drawn with the abstract result of [Bernicot et al. 2016], which is a sparse domination principle for nonintegral singular operators. The off-diagonal estimate assumption Theorem 1.1(b) of the work above is a clear counterpart of (SS), while the maximal truncation assumption of Theorem 1.1(c) in the same work is the nonkernel analogue of the grand maximal function from [Lerner 2016]. It would be interesting to investigate whether, in the nonkernel setting of [Bernicot et al. 2016], an assumption in the vein of (L) can be used instead.

Remark 2.6 (the essence of (L)). Let $\mathcal{Q}$ be a stopping collection with top Q. When b belongs to an $\mathcal{X}_\alpha(\mathcal{Q})$-type space, the forms

$$(b, h) \mapsto \Lambda_{\mathcal{Q},\mu,\nu}(b, h), \quad (b, h) \mapsto \Lambda_{\mathcal{Q},\mu,\nu}(h, b)$$

have a much more familiar representation, which is what allows verification of assumption (L) in practice. By rephrasing the definition, when $b \in \mathcal{X}_1(\mathcal{Q})$ is supported on Q (which we can assume with no restriction) we have the equality

$$\Lambda_{\mathcal{Q},\mu,\nu}(b, h) = \sum_{j \geq 1} \int \sum_{\mu < s \leq \min\{s_Q, \nu\}} K_s(x, y) b_{s-j}(y) h(x) \, \mathrm{d}y \, \mathrm{d}x, \quad \text{where } b_s := \sum_{\substack{L \in \mathcal{Q} \\ s_L = s}} b_L. \tag{2-9}$$

This notation will be used throughout the paper; see for instance (2-10) below. Furthermore, if q is the exponent occurring in (SS), $h \in \mathcal{Y}_{q'}(\mathcal{Q})$, and $b \in \mathcal{X}_{q'}(\mathcal{Q})$, then $\Lambda_{\mathcal{Q},\mu,\nu}(h, b)$ is essentially self-adjoint up to a tolerable error term. Namely, if h is supported on Q (which we can also always assume),

$$\Lambda_{\mathcal{Q},\mu,\nu}(h, b) = \bigg(\sum_{j \geq 1} \int \sum_{\mu < s \leq \min\{s_Q, \nu\}} K_s(y, x) b^{\mathrm{in}}_{s-j}(y) h(x) \, \mathrm{d}y \, \mathrm{d}x \bigg) + V_{\mathcal{Q}}(h, b), \tag{2-10}$$

where

$$b^{\mathrm{in}} = \sum_{\substack{L \in \mathcal{Q} \\ 3L \cap 2Q \neq \varnothing}} b_L$$

is a truncation of b and thus also belongs to $\mathcal{X}_{q'}(\mathcal{Q})$ with $\|b^{\mathrm{in}}\|_{\mathcal{X}_{q'}(\mathcal{Q})} \le \|b\|_{\mathcal{X}_{q'}(\mathcal{Q})}$, and the remainder $V_{\mathcal{Q}}(h,b)$ satisfies

$$|V_{\mathcal{Q}}(h,b)| \le 2^{\vartheta d}[K]_{0,q}|Q|\,\|h\|_{\mathcal{Y}_{q'}(\mathcal{Q})}\|b\|_{\mathcal{X}_{q'}(\mathcal{Q})} \tag{2-11}$$

for a suitable positive absolute constant ϑ. The representation (2-10)–(2-11) is a simple consequence of the structure of $b \in \mathcal{X}_{q'}(\mathcal{Q})$ and of the separation properties (2-2), (2-3). We provide the necessary details for (2-10)–(2-11) in Appendix A at the end.

Proof of Theorem C. Given a form Λ satisfying the assumptions of Theorem C, $\mu < \nu \in \mathbb{Z}$ and $f_j \in L^{p_j}(\mathbb{R}^d)$, $j = 1, 2$, with compact support, we will construct a sparse collection $\mathcal{S}$ of cubes of $\mathbb{R}^d$ such that

$$\big|\Lambda_\mu^\nu(f_1, f_2)\big| \le 2^{\Theta d}\boldsymbol{C}\sum_{Q\in\mathcal{S}}|Q|\,\langle f_1\rangle_{p_1,Q}\langle f_2\rangle_{p_2,Q}, \tag{2-12}$$

where $\boldsymbol{C}$ is the expression within the square brackets in the conclusion of Theorem C. Here and below, we denote by Θ a suitably large positive absolute constant which will be chosen during the course of the proof. Within this proof, we will also denote by ϑ positive absolute constants which belong to $[2^{-8}\Theta, 2^{-7}\Theta]$ and may differ at each occurrence. As the assumptions of Theorem C are stable if we replace Λ with Λ_μ^ν, we can work under the assumption that $K_s = 0$ for all $s \notin (\mu, \nu]$ and thus drop μ, ν from the notations (2-4), (2-5).

The proof of (2-12) is iterative and is carried out in the next subsection. Here, we give the main estimate for the form $\Lambda^{s\mathcal{Q}}$ from (2-4) in terms of stopping collection norms.

Lemma 2.7. *Let Q be a fixed dyadic cube in $\mathbb{R}^d$ and $\mathcal{Q}$ be a stopping collection with top Q. Then*

$$\big|\Lambda^{s\mathcal{Q}}(h_1\mathbf{1}_Q, h_2\mathbf{1}_{3Q})\big| \le 2^{\vartheta d}\boldsymbol{C}\,|Q|\,\|h_1\|_{\mathcal{Y}_{p_1}(\mathcal{Q})}\|h_2\|_{\mathcal{Y}_{p_2}(\mathcal{Q})} + \sum_{\substack{L\in\mathcal{Q}\\ L\subset Q}}\big|\Lambda^{s_L}(h_1\mathbf{1}_L, h_2\mathbf{1}_{3L})\big|. \tag{2-13}$$

Proof. We are free to assume that $\operatorname{supp} h_1 \subset Q$ and $\operatorname{supp} h_2 \subset 3Q$ for simplicity of notation. For $j = 1, 2$, construct the Calderón–Zygmund decomposition of h_j with respect to the family $\mathcal{Q}$ as described in Remark 2.2, that is,

$$h_j = g_j + b_j, \quad b_j = \sum_{L\in\mathcal{Q}} b_{jL}, \quad b_{jL} := \bigg(h_j - \frac{1}{|L|}\int_L h_j(x)\,\mathrm{d}x\bigg)\mathbf{1}_L.$$

The Calderón–Zygmund properties in this context are, for $j = 1, 2$,

$$\|g_j\|_{\mathcal{Y}_\infty} \lesssim \|h_j\|_{\mathcal{Y}_{p_j}}, \quad \|b_j\|_{\dot{\mathcal{X}}_{p_j}} \lesssim \|h_j\|_{\mathcal{Y}_{p_j}}.$$

Using the definition (2-5), we decompose on our way to (2-13):

$$\begin{aligned}\Lambda^{s\mathcal{Q}}(h_1,h_2) &= \Lambda_{\mathcal{Q}}(h_1,h_2) + \sum_{\substack{L\in\mathcal{Q}\\ L\subset Q}}\Lambda^{s_L}(h_1\mathbf{1}_L, h_2)\\ &= \Lambda_{\mathcal{Q}}(g_1,g_2)+\Lambda_{\mathcal{Q}}(b_1,g_2)+\Lambda_{\mathcal{Q}}(g_1,b_2)+\Lambda_{\mathcal{Q}}(b_1,b_2)+\sum_{\substack{L\in\mathcal{Q}\\ L\subset Q}}\Lambda^{s_L}(h_1\mathbf{1}_L, h_2\mathbf{1}_{3L}).\end{aligned} \tag{2-14}$$

The last sum on the last right-hand side is estimated by the sum appearing on the right-hand side of (2-13). We are left with estimating the first four terms in the last line of (2-14). The leftmost is controlled by the estimate of Lemma 2.3:

$$|\Lambda_{\mathcal{Q}}(g_1, g_2)| \lesssim C_{\mathsf{T}}(r)\,|Q|\,\|g_1\|_{\mathcal{Y}_r}\|g_2\|_{\mathcal{Y}_{r'}} \lesssim \boldsymbol{C}\,|Q|\,\|h_1\|_{\mathcal{Y}_{p_1}}\|h_2\|_{\mathcal{Y}_{p_2}}.$$

The second term is handled by appealing to assumption (L), which yields

$$|\Lambda_{\mathcal{Q}}(b_1, g_2)| \le C_{\mathsf{L}}\,|Q|\,\|b_1\|_{\dot{\mathcal{X}}_{p_1}}\|g_2\|_{\mathcal{Y}_{p_2}} \lesssim \boldsymbol{C}\,|Q|\,\|h_1\|_{\mathcal{Y}_{p_1}}\|h_2\|_{\mathcal{Y}_{p_2}},$$

where the second estimate follows from the Calderón–Zygmund properties above. The third is also estimated by appealing to (L), as

$$|\Lambda_{\mathcal{Q}}(g_1, b_2)| \le C_{\mathsf{L}}\,|Q|\,\|g_1\|_{\mathcal{Y}_\infty}\|b_2\|_{\dot{\mathcal{X}}_{p_2}} \lesssim \boldsymbol{C}\,|Q|\,\|h_1\|_{\mathcal{Y}_{p_1}}\|h_2\|_{\mathcal{Y}_{p_2}}.$$

Finally, again by assumption (L),

$$|\Lambda_{\mathcal{Q}}(b_1, b_2)| \le C_{\mathsf{L}}\,|Q_0|\,\|b_1\|_{\dot{\mathcal{X}}_{p_1}}\|b_2\|_{\mathcal{Y}_{p_2}} \lesssim \boldsymbol{C}\,|Q|\,\|h_1\|_{\mathcal{Y}_{p_1}}\|h_2\|_{\mathcal{Y}_{p_2}},$$

where the final inequality follows again from the Calderón–Zygmund estimates. □

***Proof of* (2-12).** The proof is obtained by means of the iterative procedure described below.

Preliminaries: We will produce stopping collections iteratively, by suitable Whitney decompositions of unions of sets

$$E_Q = \Big\{ x \in 3Q : \max_{j=1,2} \frac{\mathrm{M}_{p_j}(f_j\,\mathbf{1}_{3Q})(x)}{\langle f_j\rangle_{p_j,3Q}} > 2^{\frac{\Theta d}{4}} \Big\} \tag{2-15}$$

associated to a cube Q and a pair of functions f_1, f_2. We notice that

$$E_Q \subset 3Q, \quad |E_Q| \le 2^{-\vartheta d}|Q|; \tag{2-16}$$

the measure estimate is a consequence of the maximal theorem, and holds provided Θ is chosen sufficiently large. In this proof, we say that two dyadic cubes L, L' are *neighbors*, and write $L \sim L'$, if

$$7L \cap 7L' \neq \varnothing, \quad |s_L - s_{L'}| < 8.$$

The separation condition (2-3) tells us that if the 7-fold dilates of two cubes L, L' belonging to the same stopping collection intersect nontrivially, then L, L' must be neighbors. We also recall the notation $\widehat{L}$ for the 2^5-fold dilate of L.

Initialize: Let $f_j \in L^{p_j}(\mathbb{R}^d)$, $j = 1, 2$, with compact support be fixed. By suitably choosing the dyadic lattice $\mathcal{D}$, we may find $Q_0 \in \mathcal{D}$ such that $\operatorname{supp} f_1 \subset Q_0$, $\operatorname{supp} f_2 \subset 3Q_0$ and s_{Q_0} is larger than the largest nonzero scale occurring in the kernel. Then set $\mathcal{S}_0 = \{Q_0\}$, $E_0 = 3Q_0$, and define referring to (2-15)

$$E_1 := E_{Q_0}, \quad \mathcal{S}_1 := \text{maximal cubes } L \in \mathcal{D} \text{ such that } 9L \subset E_1.$$

Notice that the following properties are satisfied:

$$L \in \mathcal{S}_1 \text{ are a pairwise disjoint collection,} \tag{2-17}$$

$$E_1 = \bigcup_{L\in\mathcal{S}_1} L = \bigcup_{L\in\mathcal{S}_1} 9L \subset E_0, \quad |Q_0 \setminus E_1| \geq (1-2^{-d\vartheta})|Q_0|, \tag{2-18}$$

$$L, L' \in \mathcal{S}_1,\ 7L \cap 7L' \neq \varnothing \implies L \sim L'. \tag{2-19}$$

Property (2-17) and the first part of (2-18) are by construction, while the second part of (2-18) follows from the estimate of (2-16). For (2-19) suppose instead that $7L \cap 7L'$ is not empty when $s_L \leq s_{L'} - 8$. By the relation between the sidelengths it follows that $\widehat{L} \subset 9L'$, which implies that the 9-fold dilate of the dyadic parent of L is contained in $9L'$ as well, contradicting the maximality of L. By virtue of (2-17)–(2-19), $\mathcal{Q}_1(Q_0) := \mathcal{S}_1$ is a stopping collection with top Q_0; compare with (2-2), (2-3). The first property in (2-18) guarantees that

$$\sup_{x\notin \mathsf{sh}\, \mathcal{Q}_1(Q_0)} |f_j(x)| \leq 2^{\frac{\Theta d}{4}} \langle f_j \rangle_{p_j, 3Q_0}.$$

Further, by the maximality condition on $L \in \mathcal{S}_1$, it follows that

$$\sup_{L\in\mathcal{Q}_1(Q_0)} \inf_{\widehat{L}} \mathrm{M}_{p_j}(f_j \mathbf{1}_{3Q_0}) \leq 2^{\frac{\Theta d}{4}} \langle f_j \rangle_{p_j, 3Q_0}$$

for $j = 1, 2$. The last two inequalities tell us that

$$\|f_j\|_{\mathcal{Y}_{p_j}(\mathcal{Q}_1(Q_0))} \leq 2^{\frac{\Theta d}{4}} \langle f_j \rangle_{p_j, 3Q_0}, \quad j = 1, 2.$$

Applying (2-13) to the stopping collection $\mathcal{Q}_1(Q_0)$, and $h_1 = f_1,\ h_2 = f_2$ we obtain

$$\begin{aligned} \big|\Lambda(f_1, f_2)\big| &= \big|\Lambda^{s_{Q_0}}(f_1 \mathbf{1}_{Q_0}, f_2 \mathbf{1}_{3Q_0})\big| \\ &\leq 2^{\Theta d} \boldsymbol{C} |Q_0| \langle f_1 \rangle_{p_1, 3Q_0} \langle f_2 \rangle_{p_2, 3Q_0} + \sum_{\substack{L\in\mathcal{Q}_1(Q_0)\\ L\subset Q_0}} \big|\Lambda^{s_L}(f_1 \mathbf{1}_L, f_2 \mathbf{1}_{3L})\big|. \end{aligned} \tag{2-20}$$

The obtained properties (2-17)–(2-19) and estimate (2-20) are the $\ell = 1$ case of the induction assumption in the inductive step below.

Inductive step: Suppose inductively collections $\mathcal{S}_\ell$, $0 \leq \ell \leq k$, and sets E_ℓ, $1 \leq \ell \leq k$, have been constructed, with the properties that for all $1 \leq \ell \leq k$

$$L \in \mathcal{S}_\ell \text{ are a pairwise disjoint collection,} \tag{2-21}$$

$$E_\ell = \bigcup_{L\in\mathcal{S}_\ell} L = \bigcup_{L\in\mathcal{S}_\ell} 9L \subset E_{\ell-1}, \qquad |Q \setminus E_\ell| \geq (1-2^{-\vartheta d})|Q| \quad \forall\, Q \in \mathcal{S}_{\ell-1}, \tag{2-22}$$

$$L, L' \in \mathcal{S}_\ell,\ 7L \cap 7L' \neq \varnothing \implies L \sim L'. \tag{2-23}$$

Suppose also that if $\mathcal{T}_{k-1} = \mathcal{S}_0 \cup \cdots \cup \mathcal{S}_{k-1}$, the estimate

$$\big|\Lambda(f_1, f_2)\big| \leq 2^{\Theta d} \boldsymbol{C} \sum_{R\in\mathcal{T}_{k-1}} |R| \langle f_1 \rangle_{p_1, 3R} \langle f_2 \rangle_{p_2, 3R} + \sum_{Q\in\mathcal{S}_k} \big|\Lambda^{s_Q}(f_1 \mathbf{1}_Q, f_2 \mathbf{1}_{3Q})\big| \tag{2-24}$$

has been shown to hold. At this point define

$$E_{k+1} := \bigcup_{Q\in\mathcal{S}_k} E_Q, \quad \mathcal{S}_{k+1} := \text{maximal cubes } L\in\mathcal{D} \text{ such that } 9L\subset E_{k+1},$$
$$\mathcal{Q}_{k+1}(Q) = \{L\in\mathcal{S}_{k+1} : L\subset 3Q\}, \quad Q\in\mathcal{S}_k.$$

Property (2-21), together with the first property in (2-22), as $E_Q\subset 3Q\subset E_k$, and (2-23), via the same reasoning we used for (2-19), now hold for $\ell = k+1$ as well. Let now $Q\in\mathcal{S}_k$. Property (2-23) with $\ell = k$ implies that

$$3Q\cap E_{k+1}\subset \bigcup_{Q'\in\mathcal{S}_k : Q'\sim Q} E_{Q'}.$$

Therefore, we learn that

$$|Q\cap E_{k+1}|\le |3Q\cap E_{k+1}|\le \sum_{Q'\in\mathcal{S}_k : Q'\sim Q} |E_{Q'}|\le 2^{-\vartheta d}|Q| \tag{2-25}$$

by applying for each $Q'\in\mathcal{S}_k$ with $Q'\sim Q$ the estimate of (2-16), and observing that the cardinality of $\{Q'\in\mathcal{D} : Q'\sim Q\}$ is bounded by an absolute dimensional constant, and $|Q|, |Q'|$ are comparable, again up to an absolute dimensional constant. From the above display we obtain the second part of (2-22) for $\ell = k+1$. Moreover, one observes that if $L\in\mathcal{S}_{k+1}$ with $L\cap 3Q\neq\varnothing$, then by virtue of property (2-25), L must be significantly shorter than Q and thus contained in one of the 3^d translates of the dyadic cube Q whose union covers $3Q$. Namely, we have the equality

$$\mathcal{Q}_{k+1}(Q) = \{L\in\mathcal{S}_{k+1} : L\cap 3Q\neq\varnothing\},$$

which also gives the last equality in

$$\bigcup_{L\in\mathcal{Q}_{k+1}(Q) : 3L\cap 2Q\neq\varnothing} 9L\subset \bigcup_{L\in\mathcal{S}_{k+1} : L\cap 3Q\neq\varnothing} L = \bigcup_{L\in\mathcal{Q}_{k+1}(Q)} L = \mathsf{sh}\,\mathcal{Q}_{k+1}(Q),$$

as the set in the left-hand side of the last display is contained in $3Q$ and (2-22) holds for $\ell = k+1$. Comparing with (2-2), (2-3), the discussion above gives that $\mathcal{Q}_{k+1}(Q)$ is a stopping collection with top Q such that $E_Q\subset \mathsf{sh}\,\mathcal{Q}_{k+1}(Q)$, so that

$$\sup_{x\notin \mathsf{sh}\,\mathcal{Q}_{k+1}(Q)} |f_j\,\mathbf{1}_{3Q}(x)|\le 2^{\frac{\Theta d}{4}}\langle f_j\rangle_{p_j,3Q}.$$

Furthermore, for $j=1,2$

$$\sup_{L\in\mathcal{Q}_{k+1}(Q)} \inf_{\widehat{L}} \mathrm{M}_{p_j}(f_j\,\mathbf{1}_{3Q})\le 2^{\frac{\Theta d}{4}}\langle f_j\rangle_{p_j,3Q};$$

otherwise the 9-fold dilate of the dyadic parent of some $L\in\mathcal{Q}_{k+1}(Q)$ would be contained in E_Q and thus in E_{k+1}, contradicting the maximality of such an L. Therefore

$$\|f_j\,\mathbf{1}_{3Q}\|_{\mathcal{Y}_{p_j}(\mathcal{Q}_{k+1}(Q))}\le 2^{\frac{\Theta d}{4}}\langle f_j\rangle_{p_j,3Q}, \quad j=1,2,$$

and we may apply (2-13) to each $Q \in \mathcal{S}_k$ summand in (2-24), with $h_1 = f_1,\ h_2 = f_2$ and obtain

$$\begin{aligned}\big|\Lambda^{s_Q}(f_1\mathbf{1}_Q, f_2\mathbf{1}_{3Q})\big| &\le 2^{\Theta d}\boldsymbol{C}\,|Q|\langle f_1\rangle_{p_1,3Q}\langle f_2\rangle_{p_2,3Q} + \sum_{L\in\mathcal{Q}_{k+1}(Q):L\subset Q}\big|\Lambda^{s_L}(f_1\mathbf{1}_L, f_2\mathbf{1}_{3L})\big| \\ &= 2^{\Theta d}\boldsymbol{C}\,|Q|\langle f_1\rangle_{p_1,3Q}\langle f_2\rangle_{p_2,3Q} + \sum_{L\in\mathcal{S}_{k+1}:L\subset Q}\big|\Lambda^{s_L}(f_1\mathbf{1}_L, f_2\mathbf{1}_{3L})\big|.\end{aligned}$$

As $Q \in \mathcal{S}_k$ are pairwise disjoint, see (2-21), summing over $Q \in \mathcal{S}_k$, writing $\mathcal{T}_k = \mathcal{S}_0 \cup \cdots \cup \mathcal{S}_k$ and combining the resulting estimate with (2-24), we arrive at

$$\big|\Lambda(f_1, f_2)\big| \le 2^{\Theta d}\boldsymbol{C}\sum_{Q\in\mathcal{T}_k}|Q|\langle f_1\rangle_{p_1,3Q}\langle f_2\rangle_{p_2,3Q} + \sum_{L\in\mathcal{S}_{k+1}}\big|\Lambda^{s_L}(f_1\mathbf{1}_L, f_2\mathbf{1}_{3L})\big|,$$

that is, (2-24) with k replaced by $k+1$. This, together with the previously obtained (2-21)–(2-23) for $\ell = k+1$, completes the current iteration.

Termination: A consequence of our construction is that $\sigma_k := \max\{s_Q : Q \in \mathcal{S}_k\} \le s_{Q_0} - \vartheta k$. The algorithm terminates when $k = K$, where K is such that σ_K is strictly less than the minimal nonzero scale in the kernel. For $k = K$ in (2-24) the second sum on the right-hand side vanishes identically and we have obtained the estimate (2-12) by setting $\mathcal{T} := \mathcal{T}_{K-1}$ and $\mathcal{S} := \{3Q : Q \in \mathcal{T}\}$. We see that the collection $\mathcal{T}$, and thus the collection of the dilates $\mathcal{S}$, are sparse by simply observing that the sets

$$F_Q := Q \setminus E_{k+1}, \quad Q \in \mathcal{S}_k,$$

are pairwise disjoint for $Q \in \mathcal{T}$ and have measure larger than $(1 - 2^{-d\vartheta})|Q|$, as can be seen from (2-22).

3. Localized estimates for Dini- and Hörmander-type kernels

In the first part of this section, we state and prove a family of localized estimates, of the type occurring in condition (L) of Theorem C, for kernels falling within the scope of (SS) and possessing additional smoothness properties, of Dini or Hörmander type. These estimates and their proof are a reformulation of the classical inequalities intervening in the proof of the weak-L^1-bound for Calderón–Zygmund operators (see, for example, [Stein 1993, Chapter I]). We choose to provide details as we believe the arguments to be rather explanatory of the driving philosophy behind Theorem C.

As we mentioned in the Introduction, our abstract Theorem C, coupled with the localized estimates that follow, can be employed to reprove the optimal sparse domination estimates for Calderón–Zygmund kernels of Dini and Hörmander type, thus recovering the results (among others) of [Bui et al. 2017; Hytönen et al. 2017; Lacey 2017; Lerner 2016; Volberg and Zorin-Kranich 2016]. We provide a summary of the statements of such domination theorems in the second part of this section.

Localized estimates and kernel norms. Throughout these estimates, we assume that a stopping collection $\mathcal{Q}$ with top Q as in Section 2 has been fixed, and the notations $\Lambda_{\mathcal{Q},\mu,\nu}$ refer to (2-5). It is understood that the constants implied by the almost inequality signs depend on dimension only and are in particular are uniform over the choice of $\mathcal{Q}$. We begin with the single-scale localized estimate where no cancellation is exploited.

Lemma 3.1 (trivial estimate). *Let $1 < \beta \le \infty$ and $\alpha = \beta'$. Then for all $j \ge 1$,*

$$\sum_s \int |K_s(x,y)|\,|b_{s-j}(y)|\,|h(x)|\,\mathrm{d}y\,\mathrm{d}x \lesssim [K]_{0,\beta}\,|Q|\,\|b\|_{\mathcal{X}_1}\|h\|_{\mathcal{Y}_\alpha}.$$

Proof. As $\|b_L\|_1 \lesssim |L|\,\|b\|_{\mathcal{X}_1}$ for $L \in \mathcal{Q}$, it suffices to prove that for each $L \in \mathcal{Q}$ and $s = s_L + j$,

$$\int |K_s(x,y)|\,|b_L(y)|\,|h(x)|\,\mathrm{d}y\,\mathrm{d}x \lesssim [K]_{0,\beta}\,\|b_L\|_1\|h\|_{\mathcal{Y}_\alpha}. \tag{3-1}$$

In turn, it then suffices to prove that

$$s \ge s_L \Longrightarrow \sup_{y\in L}\int |K_s(y+u,y)|\,|h(y+u)|\,\mathrm{d}u \lesssim [K]_{0,\beta}\,\|h\|_{\mathcal{Y}_\alpha},$$

which readily follows from

$$\begin{aligned}\int |K_s(y+u,y)|\,|h(y+u)|\,\mathrm{d}u &\le \|K_s(y+\cdot,y)\|_\beta\left(\int_{B(y,2^{s+10})}|h(z)|^\alpha\,\mathrm{d}z\right)^{\frac{1}{\alpha}}\\ &\lesssim [K]_{0,\beta}\big(\inf_{\widehat{L}}\mathrm{M}_\alpha h\big) \le [K]_{0,\beta}\|h\|_{\mathcal{Y}_\alpha}\end{aligned}$$

when $y \in L$. Above, we used the support condition (SS) and Hölder's inequality for the first step, and subsequently that the ball $B(y,2^{s+10}) = \{z \in \mathbb{R}^d : |z-y| < 2^{s+10}\}$ contains the dilate $\widehat{L}$. □

We introduce a further family of kernel norms in addition to the one of (SS), to which we refer for notation. For $1 < \beta \le \infty$ set

$$[K]_{1,\beta} := \sum_{j=1}^{\infty} \varpi_{j,\beta}(K), \tag{3-2}$$

where

$$\varpi_{j,\beta}(K) := \sup_{s\in\mathbb{Z}} 2^{\frac{sd}{\beta'}} \sup_{x\in\mathbb{R}^d} \sup_{\substack{h\in\mathbb{R}^d\\ \|h\|_\infty < 2^{s-j-1}}} \big(\big\|K_s(x,x+\cdot)-K_s(x+h,x+\cdot)\big\|_\beta + \big\|K_s(x+\cdot,x)-K_s(x+\cdot,x+h)\big\|_\beta\big).$$

The second localized estimate we consider uses the finiteness of $[K]_{1,\beta}$ to incorporate the constant-mean zero cancellation effect.

Lemma 3.2 (cancellation estimate). *Let $1 < \beta \le \infty$ and $\alpha = \beta'$. Then for all $\mu, \nu \in \mathbb{Z}$,*

$$\big|\Lambda_{\mathcal{Q},\mu,\nu}(b,h)\big| + \big|\Lambda_{\mathcal{Q},\mu,\nu}(h,b)\big| \lesssim \big([K]_{0,\infty} + [K]_{1,\beta}\big)|Q|\,\|b\|_{\dot{\mathcal{X}}_1}\|h\|_{\mathcal{Y}_\alpha}. \tag{3-3}$$

Proof. It will suffice to prove the estimate

$$\sum_{L\in\mathcal{Q}}\sum_{j=1}^{\infty}\left|\int K_{s_L+j}(x,y)\tilde{b}_L(y)\tilde{h}(x)\,\mathrm{d}y\,\mathrm{d}x\right| \lesssim [K]_{1,\beta}\,|Q|\,\|\tilde{b}\|_{\dot{\mathcal{X}}_1}\|\tilde{h}\|_{\mathcal{Y}_\alpha}. \tag{3-4}$$

In fact, by using the representations in (2-9), (2-10) we see that for all $\mu, \nu \in \mathbb{Z}$ and each pair $b \in \dot{\mathcal{X}}_1$, $h \in \mathcal{Y}_\alpha$, the forms $|\Lambda_{\mathcal{Q},\mu,\nu}(b,h)|$, $|\Lambda_{\mathcal{Q},\mu,\nu}(h,b)|$ are both bounded above by the left-hand side of (3-4) for suitable $\tilde{b} \in \dot{\mathcal{X}}_1$, $\tilde{h} \in \mathcal{Y}_\alpha$ whose norms are dominated by $\|b\|_{\dot{\mathcal{X}}_1}$, $\|h\|_{\mathcal{Y}_\alpha}$ respectively, up to possibly

replacing K_s with its transpose and controlling the remainder term $V_{\mathcal{Q}}(h,b)$ in the case of $\Lambda_{\mathcal{Q},\mu,\nu}(h,b)$. This remainder is estimated in (2-11) for $q=\infty$, which is acceptable for the right-hand side of (3-3).

We will obtain estimate (3-4) from the bound

$$\sum_{j=1}^{\infty}\left|\int K_{s_L+j}(x,y)\,b_L(y)\tilde{h}(x)\,\mathrm{d}y\,\mathrm{d}x\right| \lesssim [K]_{1,\beta}\,|L|\,\|\tilde{b}\|_{\dot{\mathcal{X}}_1}\|\tilde{h}\|_{\mathcal{Y}_\alpha}, \quad L\in\mathcal{Q} \tag{3-5}$$

by summing over $L\in\mathcal{Q}$ in and using their disjointness, given in (2-2). Fix $L\in\mathcal{Q}$ and $j\geq 1$. Using the cancellation of $\tilde{b}_L$ and then arguing as in the proof of (3-1) above we obtain

$$\begin{aligned}\left|\int K_{s_L+j}(x,y)\,b_L(y)\tilde{h}(x)\,\mathrm{d}y\,\mathrm{d}x\right| &\leq \|\tilde{b}_L\|_1 \sup_{y\in L}\int |K_{s_L+j}(y+u,y)-K_{s_L+j}(y+u,c_L)|\,|\tilde{h}(y+u)|\,\mathrm{d}u\\ &\lesssim \|\tilde{b}_L\|_1\,\omega_{j,\beta}(K)\big(\inf_{\widehat{L}}\mathrm{M}_\alpha\tilde{h}\big) \lesssim \omega_{j,\beta}(K)\,|L|\,\|\tilde{b}\|_{\dot{\mathcal{X}}_1}\|\tilde{h}\|_{\mathcal{Y}_\alpha},\end{aligned}$$

and (3-5) follows by summing over $j\geq 1$. □

Sparse domination of Calderón–Zygmund kernels. We now briefly mention how our abstract result, Theorem C, can be employed to recover sparse domination, and thus weighted bounds, for Calderón–Zygmund kernels with minimal smoothness assumptions. Let T be an $L^2(\mathbb{R}^d)$-bounded operator whose kernel K satisfies the usual size normalization

$$\sup_{x\neq y}|x-y|^d\,|K(x,y)|\leq 1.$$

Let ψ be a fixed Schwartz function supported in $A_1=\{x\in\mathbb{R}^d : 2^{-2}<|x|<1\}$ such that

$$\sum_{s\in\mathbb{Z}}\psi(2^{-s}x)=1, \quad x\neq 0.$$

It is immediate to see that (SS) holds, and in particular $[K]_{0,\infty}\leq C$, for the decomposition

$$K_s(x,y):=K(x,y)\,\psi\left(\frac{x-y}{2^s}\right), \quad s\in\mathbb{Z}.$$

We further assume that $[K]_{1,\beta}<\infty$ for some $1<\beta\leq\infty$, where the kernel norm has been defined in (3-2). When $\beta=\infty$, this is exactly the Dini condition [Hytönen et al. 2017; Lacey 2017; Lerner 2016]. For $\beta<\infty$, the above condition is equivalent to the assumptions of [Volberg and Zorin-Kranich 2016], where in fact a multilinear version is presented.

The assumptions of Theorem C then hold for the dual form

$$\Lambda(f_1,f_2)=\langle Tf_1,\bar{f_2}\rangle.$$

We have already observed that (SS) is verified with $q=\infty$. It is well known that the L^2-boundedness of Λ together with $[K]_{1,\beta}<\infty$ yields that the truncation forms Λ_μ^ν, see (2-1), are uniformly bounded on $L^t(\mathbb{R}^d)\times L^{t'}(\mathbb{R}^d)$ [Stein 1993, Chapter I.7] for all $1<t<\infty$; thus we have condition (T) with, for instance, $r=2$. Furthermore, Lemma 3.2 is exactly (L) for the corresponding $\Lambda_{\mathcal{Q},\mu,\nu}$, with $p_1=1$, $p_2=\alpha=\beta'$.

Applying Theorem C in the form given in Remark 2.4, we obtain the following sparse domination result, which recovers (the dual form of) the domination theorems from the above-mentioned references. We cite the same references for the sharp weighted norm inequalities that descend from this result.

Theorem D (Calderón–Zygmund theory). *Let T be as above and $1 \le \beta < \infty$. For all $1 < t < \infty$ and all pairs $f_1 \in L^t(\mathbb{R}^d)$, $f_2 \in L^{t'}(\mathbb{R}^d)$,*

$$|\langle Tf_1, f_2\rangle| \le C_\beta [K]_{1,\beta} \sup_{\mathcal{S}} \mathsf{PSF}_{\mathcal{S};1,\beta'}(f_1, f_2),$$

where C_β is a positive constant depending on β and on the dimension d only.

4. Proof of Theorem A

Let $1 < q \le \infty$ and suppose that $\Omega \in L^q(S^{d-1})$ has unit norm and vanishing integral. Set $x' = x/|x|$. We decompose for $x \ne 0$ the kernel of T_Ω in (1-2) as

$$\frac{\Omega(x')}{|x|^d} = \sum_s K_s(x), \quad K_s(x) = \Omega(x') 2^{-sd} \phi(2^{-s}x),$$

where ϕ is a suitable smooth radial function supported in $A_1 = \{2^{-2} \le |x| \le 1\}$. The main result of this section is the following proposition: again, we assume that a stopping collection $\mathcal{Q}$ with top the dyadic cube Q as in Section 2 has been fixed and the notations $\mathcal{Y}_t$ and similar refer to that fixed setting.

Proposition 4.1. *Let $\Omega \in L^q(S^{d-1})$ be of unit norm and vanishing integral. Let $\{\varepsilon_s\} \in \{-1, 0, 1\}^{\mathbb{Z}}$ be a choice of signs, $b \in \dot{\mathcal{X}}_1$ and define*

$$\mathsf{K}(b, h) := \sum_{j \ge 1} \sum_s \varepsilon_s \langle K_s * b_{s-j}, \bar{h}\rangle$$

where

$$b_s = \sum_{\substack{L \in \mathcal{Q} \\ s_L = s}} b_L.$$

There exists an absolute constant C, in particular uniform over all $\{\varepsilon_s\} \in \{-1, 0, 1\}^{\mathbb{Z}}$, such that

$$|\mathsf{K}(b, h)| \le \frac{Cp}{p-1} |Q| \, \|b\|_{\dot{\mathcal{X}}_1} \|h\|_{\mathcal{Y}_p} \begin{cases} \|\Omega\|_{L^{q,1}\log L(S^{d-1})}, & q < \infty,\ p \ge q', \\ \|\Omega\|_{L^\infty(S^{d-1})}, & q = \infty,\ p > 1. \end{cases} \tag{4-1}$$

With the above proposition in hand, we may now give the proof of Theorem A. The structural assumptions (SS), (T) of the abstract result Theorem C applied to the above decomposition of (the dual form of) T_Ω are respectively verified with $q = q$ and with $r = 2$ (this is the classical L^2-boundedness of the truncations of T_Ω [Calderón and Zygmund 1956; Grafakos and Stefanov 1999]).

We still need to verify (L) for the values $p_1 = 1$ and $p_2 = p$ for each p in the claimed range (depending on whether $q = \infty$ or not). It is immediate from the representations (2-9) that in this setting $\Lambda_{\mathcal{Q},\mu,\nu}(b, h) = \mathsf{K}(b\mathbf{1}_Q, h)$ for a suitable choice of signs $\{\varepsilon_s\}$ depending on μ, ν. So Proposition 4.1 yields the first condition in (L) with $p_1 = 1$, $p_2 = p$. On the other hand, we get from (2-10) that $\Lambda_{\mathcal{Q},\mu,\nu}(h, b)$ is equal to $\mathsf{K}(b^{\mathsf{in}}, h\mathbf{1}_Q)$, again for a suitable choice of signs $\{\varepsilon_s\}$ depending on μ, ν, up to replacing K_s

by $K_s(-\cdot)$, and up to subtracting the remainder term from (2-11), which is estimated in this case by an absolute constant times

$$|Q|\,\|h\|_{\mathcal{Y}_\infty}\|b\|_{\mathcal{Y}_{q'}} \le |Q|\,\|h\|_{\mathcal{Y}_\infty}\|b\|_{\mathcal{Y}_p},$$

which is acceptable for the right-hand side of the second condition in (L) when $p_2 = p$. These considerations and another application of Proposition 4.1 finally yield Theorem A, via our abstract result in the form described in Remark 2.4.

Proof of Proposition 4.1. Throughout this proof, C is a positive absolute dimensional constant which may vary at each occurrence without explicit mention. We assume $\{\varepsilon_s\} \in \{-1,0,1\}^{\mathbb{Z}}$ is given. For the sake of simplicity, we redefine $K_s := \varepsilon_s K_s$; it will be clear from the proof below that the signs of K_s play no role. Fix a positive integer j. For $\delta > 0$ to be fixed at the end of the argument define

$$O_j = \{\theta \in S^{d-1} : |\Omega(\theta)| > 2^{\delta j}\}, \quad \Omega_j = \Omega\,\mathbf{1}_{S^{d-1}\setminus O_j}, \quad \Delta_j = \Omega\,\mathbf{1}_{O_j}. \tag{4-2}$$

We now have the decomposition

$$K_s = H_s^j + V_s^j, \quad H_s^j = K_s\,\mathbf{1}_{\operatorname{supp}\Omega_j}, \quad V_s^j = K_s\,\mathbf{1}_{O_j}. \tag{4-3}$$

The first localized form we treat, namely the contribution of the unbounded part of Ω, is dealt with by means of a trivial estimate.

Lemma 4.2. $$\mathsf{V}^j(b,h) := \sum_s |\langle V_s^j * b_{s-j}, \bar{h}\rangle| \le C\,\|\Delta_j\|_q\,|Q|\,\|b\|_{\mathcal{X}_1}\|h\|_{\mathcal{Y}_p}, \quad p \ge q'.$$

Proof. It suffices of course to prove the estimate above with q' in place of p. This is actually a particular case of Lemma 3.1 applied with $K = \{V_s^j\}$ and $\beta = q$, as it is immediate to see that for this kernel one has $[K]_{0,q} \le C\,\|\Delta_j\|_q$. □

The contribution of the bounded part of K_s in (4-3) is more delicate, and we postpone the proof of the following lemma to the next subsection.

Lemma 4.3. *There exist absolute constants $C, c > 0$ such that for all $1 < p \le \infty$*

$$\mathsf{H}^j(b,h) := \left|\sum_s \langle H_s^j * b_{s-j}, \bar{h}\rangle\right| \le C\,2^{-cj\frac{p-1}{p}}\,\|\Omega_j\|_\infty\,|Q|\,\|b\|_{\dot{\mathcal{X}}_1}\|h\|_{\mathcal{Y}_p}.$$

We may now complete the proof of Proposition 4.1. We assume $q < \infty$. The remaining case is actually simpler as V^j is identically zero. Our decomposition (4-3) yields that

$$|\mathsf{K}(b,h)| \le \sum_{j\ge 1} |\mathsf{H}^j(b,h)| + \sum_{j\ge 1} |\mathsf{V}^j(b,h)|.$$

Choosing $\delta = c(p-1)/(2p)$ in (4-2) and using Lemma 4.3, we estimate

$$\begin{aligned}\sum_{j\ge 1} |\mathsf{H}^j(b,h)| &\le C\,|Q|\,\|b\|_{\dot{\mathcal{X}}_1}\|h\|_{\mathcal{Y}_p} \sum_{j\ge 1} 2^{-cj\frac{p-1}{p}}\,\|\Omega_j\|_\infty \\ &\le C\,|Q|\,\|b\|_{\dot{\mathcal{X}}_1}\|h\|_{\mathcal{Y}_p} \sum_{j\ge 1} 2^{-cj\frac{p-1}{2p}} \le \frac{Cp}{p-1}|Q|\,\|b\|_{\dot{\mathcal{X}}_1}\|h\|_{\mathcal{Y}_p},\end{aligned}$$

which is smaller than the right-hand side of (4-1). Using Lemma 4.2, the latter sum involving V_j is then estimated by

$$\Big(\sum_{j\ge1}\|\Delta_j\|_q\Big)|Q|\,\|b\|_{\mathcal{X}_1}\|h\|_{\mathcal{Y}_p}\le\frac{Cp}{p-1}\|\Omega\|_{L^{q,1}\log L(S^{d-1})}|Q|\,\|b\|_{\mathcal{X}_1}\|h\|_{\mathcal{Y}_p},$$

which also complies with the right-hand side of (4-1); here we have used that

$$\sum_{j\ge1}\|\Delta_j\|_q\le\sum_{j\ge1}\sum_{k\ge j}2^{\delta k}|O_k\setminus O_{k+1}|^{\frac1q}\le\sum_{k\ge1}k2^{\delta k}|O_k\setminus O_{k+1}|^{\frac1q}\le\frac{C}{\delta}\|\Omega\|_{L^{q,1}\log L(S^{d-1})}.$$

The proposition is thus proved up to establishing Lemma 4.3.

Proof of Lemma 4.3. Our first observation is actually another trivial estimate.

Lemma 4.4. *There exists $C>0$ such that $|\mathsf{H}^j(b,h)|\le C\|\Omega_j\|_\infty|Q|\,\|b\|_{\mathcal{X}_1}\|h\|_{\mathcal{Y}_1}$.*

Proof. This is an application of Lemma 3.1 to $K=\{H_s^j\}$ with $\beta=\infty$, as it is immediate to see that for this kernel one has $[K]_{0,\infty}\le C\|\Omega_j\|_\infty$. □

The second step is an estimate with decay, but involving $\mathcal{Y}_\infty$ norms.

Lemma 4.5. *There exist $C,c>0$ such that $|\mathsf{H}^j(b,h)|\le C2^{-cj}\|\Omega_j\|_\infty|Q|\,\|b\|_{\mathcal{X}_1}\|h\|_{\mathcal{Y}_\infty}$.*

Before the proof of Lemma 4.5, which is given in the next subsection, we observe that the estimate of Lemma 4.3 is obtained by Riesz–Thorin (for instance) interpolation in h of the last two lemmata.

Proof of Lemma 4.5. The techniques of this subsection are an elaboration of the arguments of [Seeger 1996]. In particular Lemma 4.6 below is a stronger version of Lemma 2.1 of that work, while Lemma 4.7 is essentially the dual form of its Lemma 2.2.

We perform a further decomposition of H_s^j. Let $\Xi=\{e_\nu\}$ be a maximal 2^{-j-10d}-separated set contained in $\operatorname{supp}\Omega_j$. We may partition $\operatorname{supp}\Omega_j$ into $\#\Xi\lesssim2^{j(d-1)}$ subsets E_ν each containing e_ν and such that $\operatorname{diam}|E_\nu|\lesssim2^{-j}$. Set

$$H_{s\nu}^j(x)=H_s^j(x)\mathbf{1}_{E_\nu}(x').$$

Also, let ψ be a smooth function on $\mathbb{R}$ with $\mathbf{1}_{[-2,2]}\le\psi\le\mathbf{1}_{[-4,4]}$. Let $\kappa\in[0,1)$ and define the multiplier operator

$$\widehat{P}_\nu^j(\xi)=\psi(2^{j(1-\kappa)}\xi'\cdot e_\nu).$$

We now have the decomposition

$$H_s^j:=\Gamma_s^j+\Upsilon_s^j,\quad \Gamma_s^j:=\sum_\nu P_\nu^j*H_{s\nu}^j,\quad \Upsilon_s^j:=H_s^j-\Gamma_s^j$$

so that H^j is the sum of the single-scale bilinear forms

$$\mathsf{G}_j(b,h)=\Big\langle\sum_s\Gamma_s^j*b_{s-j},\bar h\Big\rangle,\quad \mathsf{U}_j(b,h)=\Big\langle\sum_s\Upsilon_s^j*b_{s-j},\bar h\Big\rangle$$

satisfying the estimates below.

Lemma 4.6. *Let $\tau > 1$. Then*

$$|\mathsf{G}_j(b,h)| \le C_\tau 2^{-j\frac{(1-\kappa)}{2}} \|\Omega_j\|_\infty |Q| \|b\|_{\mathcal{X}_1} \|h\|_{\mathcal{Y}_\tau}, \quad C_\tau = \frac{C\tau}{\tau-1}.$$

Lemma 4.7. *Let $b \in \dot{\mathcal{X}}_1$. For all $\varepsilon > 0$ there exists a constant $C_{\kappa,\varepsilon}$ depending on κ, ε only such that*

$$|\mathsf{U}_j(b,h)| \le C_{\kappa,\varepsilon} 2^{-\varepsilon j} \|\Omega_j\|_\infty |Q| \|b\|_{\dot{\mathcal{X}}_1} \|h\|_{\mathcal{Y}_\infty}.$$

Notice that the combination of Lemma 4.6 with $\tau = 2$ and $\kappa = \frac{1}{2}$ and Lemma 4.7 with $\varepsilon = \frac{1}{4}$ yields the required estimate for Lemma 4.5, with $c = \frac{1}{4}$. Lemma 4.5 is thus proved up to the arguments for Lemmata 4.6 and 4.7.

Proof of Lemma 4.6. We may factor out $\|\Omega_j\|_\infty$ and assume that the angular part in the definition of Γ_j is bounded by 1. We can also assume that $H^j_{s\nu}$ and b are positive as cancellation plays no role in this argument; this is just a matter of saving space in the notation. Using interpolation and duality with t below being the dual exponent of τ, the estimate of the lemma follows if we show that for each integer $r \ge 1$ and $t = 2r$

$$\frac{1}{|Q|^{\frac{1}{t}}} \Big\| \sum_s \Gamma^j_s * b_{s-j} \Big\|_t \lesssim t 2^{-\frac{j(1-\kappa)}{2}} \|b\|_{\mathcal{X}_1} \tag{4-4}$$

with an implicit constant that does not depend on r. Setting

$$M_\nu = \sum_s P^j_\nu * H^j_{s\nu} * b_{j-s}, \quad D_\nu = \sum_s H^j_{s\nu} * b_{s-j},$$

we rewrite the left-hand side of (4-4) raised to t-th power and subsequently estimate

$$\begin{aligned} \Big\| \sum_{\nu_1,\dots,\nu_r} \prod_{k=1}^r M_{\nu_k} \Big\|_2^2 &= \Big\| \sum_{\nu_1,\dots,\nu_r} \widehat{M}_{\nu_1} * \cdots * \widehat{M}_{\nu_r} \Big\|_2^2 \lesssim 2^{rj(d-2+\kappa)} \sum_{\nu_1,\dots,\nu_r} \Big\| \prod_{k=1}^r D_{\nu_k} \Big\|_2^2 \\ &\lesssim 2^{tj(d-1)} 2^{-rj(1-\kappa)} \sup_\nu \|D_\nu\|_t^t. \end{aligned} \tag{4-5}$$

We have used Plancherel for the first equality, followed by the observation that $\widehat{P}^j_{\nu_k}(\xi)$ is uniformly bounded and nonzero only if $|\xi' - e_{\nu_k}| < 2^{-j(1-\kappa)}$. Thus there are at most $C 2^{rj(d-2+\kappa)}$ r-tuples such that the r-fold convolution is nonzero, whence the first bound. Another usage of Plancherel, the observation that there are at most $2^{rj(d-1)}$ tuples in the summation, and finally Hölder's inequality yield the second bound. We are thus done if we estimate for each fixed ν

$$\sum_{s_1 \ge \cdots \ge s_t} \int \Big(\prod_{k=1}^t H^j_{s_k \nu}(x - y_k) b_{s_k - j}(y_k) \Big) \, \mathrm{d}y_1 \cdots \mathrm{d}y_t \, \mathrm{d}x \lesssim C^t 2^{-tj(d-1)} |Q| \|b\|_{\mathcal{X}_1}^t \tag{4-6}$$

as $\|D_\nu\|_t^t$ is at most t^t times the above integral. Notice that if $\sigma \le s$ then $\operatorname{supp} H^j_{\sigma\nu}$ is contained in a box R_s centered at zero and having one long side of length $\lesssim 2^s$ and $d-1$ short sides of length 2^{s-j}. If $z \in \mathbb{R}^d$, $R_s(z) = z + R_s$ and

$$\mathcal{Q}_s(z) = \big\{ L \in \mathcal{Q} : s_L \le s - j,\ L \subset 100 R_s(z) \big\}, \quad \mathsf{b}_{R_s(z)} := \sum_{L \in \mathcal{Q}_s(z)} b_L,$$

we have, by the disjointness of $L \in \mathcal{Q}$,

$$2^{-sd}\|\mathsf{b}_{R_s(z)}\|_1 \lesssim 2^{-sd}|R_s(z)|\,\|b\|_{\mathcal{X}_1} \le C2^{-j(d-1)}\|b\|_{\mathcal{X}_1} := \alpha. \tag{4-7}$$

Also notice that for all fixed $y_1, \dots, y_t$ and for all $s_1 \ge \cdots \ge s_t$,

$$I_{s_1,\dots,s_t}(y_1,\dots,y_t) := \int \Big(\prod_{k=1}^{t} H^j_{s_k\nu}(x-y_k)\Big)\,\mathrm{d}x \le \|H^j_{s_t\nu}\|_1 \prod_{k=1}^{t-1}\|H^j_{s_k\nu}\|_\infty \lesssim 2^{-j(d-1)}2^{-d\mathsf{s}_{t-1}},$$

where, here and in what follows, we set

$$\mathsf{s}_n = \sum_{k=1}^{n} s_k, \quad n = 1,\dots,t.$$

Furthermore, $I_{s_1,\dots,s_t}(y_1,\dots,y_t)$ is nonzero only if $y_k \in 2R_{s_{k-1}}(y_{k-1})$ for $k = t, t-1, \dots, 2$. Now, writing b_{s_k} in place of b_{s_k-j} for reasons of space as j is kept fixed throughout and using (4-7) repeatedly, the sum in (4-6) is equal to

$$\begin{aligned}
&\sum_{s_1\ge\cdots\ge s_t}\int I_{s_1,\dots,s_t}(y_1,\dots,y_t)\Big(\prod_{k=1}^{t}\mathsf{b}_{s_k}(y_k)\Big)\,\mathrm{d}y_1\cdots\mathrm{d}y_t\\
&\lesssim 2^{-j(d-1)}\sum_{s_1\ge\cdots\ge s_{t-1}} 2^{-d\mathsf{s}_{t-2}}\int \mathsf{b}_{s_1}(y_1)\Big(\prod_{k=2}^{t-1}\mathsf{b}_{s_k}(y_k)\mathbf{1}_{2R_{s_{k-1}}(y_{k-1})}(y_k)\Big)\frac{\|\mathsf{b}_{R_{s_{t-1}}(y_{t-1})}\|_1}{2^{ds_{t-1}}}\,\mathrm{d}y_1\cdots\mathrm{d}y_{t-1}\\
&\lesssim \alpha 2^{-j(d-1)}\sum_{s_1\ge\cdots\ge s_{t-2}} 2^{-d\mathsf{s}_{t-3}}\int \mathsf{b}_{s_1}(y_1)\Big(\prod_{k=2}^{t-2}\mathsf{b}_{s_k}(y_k)\mathbf{1}_{2R_{s_{k-1}}(y_{k-1})}(y_k)\Big)\frac{\|\mathsf{b}_{R_{s_{t-2}}(y_{t-2})}\|_1}{2^{ds_{t-2}}}\,\mathrm{d}y_1\cdots\mathrm{d}y_{t-2}\\
&\lesssim\cdots\lesssim \alpha^{t-1}2^{-j(d-1)}|Q|\,\|b\|_{\mathcal{X}_1} \le C^t 2^{-tj(d-1)}|Q|\,\|b\|^t_{\mathcal{X}_1}
\end{aligned}$$

as claimed, and this completes the proof. □

Proof of Lemma 4.7. Again we factor out $\|\Omega_j\|_\infty$ and work under the assumption that the angular part is bounded by 1. In this proof, M is a large integer whose value may differ at each occurrence and the constants implied by the almost inequality sign are allowed to depend on M only. Let β be a smooth function supported in $A_1 = \{2^{-1} \le |\xi| \le 2\}$ and satisfying

$$\sum_{k\in\mathbb{Z}} \beta^2(2^k\xi) = 1, \quad \xi \ne 0.$$

Set $B_k = \mathcal{F}^{-1}\{\beta(2^k\cdot)\}$. Defining

$$\widehat{R}^{jk}_{s\nu}(\xi) = \beta(2^k\xi)(1-\widehat{P}^j_\nu(\xi))\widehat{H}^j_{s\nu}(\xi),$$

we recall from [Seeger 1996, equations (2.6), (2.7)] the estimate

$$\|R^{jk}_{s\nu}\|_1 \lesssim_M 2^{-j(d-1)}\min\{1, 2^{-M\kappa j}2^{-M(s-j-k)}\}.$$

Now, fix s and $L \in \mathcal{Q}$ with $\ell(L) = 2^{s-j}$ for the moment. Recalling the definition of Υ_s^j, we have the decomposition

$$|\langle \Upsilon_s^j * b_L, \bar{h}\rangle| \le \sum_\nu \sum_k |\langle R_{s\nu}^{jk} * B_k * b_L, \bar{h}\rangle|,$$

and the cancellation estimate (cf. [Seeger 1996, equation (2.5)], a simpler version of Lemma 3.2)

$$\begin{aligned} |\langle R_{s\nu}^{jk} * B_k * b_L, \bar{h}\rangle| &\lesssim \min\{1, 2^{(s-j)-k}\} \|R_{s\nu}^{jk}\|_1 \|b_L\|_1 \|h\|_\infty \\ &\lesssim 2^{-j(d-1)} \min\big\{2^{(s-j)-k}, 2^{-M\kappa j - M(s-j-k)}\big\} |L| \, \|b\|_{\dot{\mathcal{X}}_1} \|h\|_{\mathcal{Y}_\infty}. \end{aligned} \tag{4-8}$$

Note that $\#\Xi \lesssim 2^{j(d-1)}$. So for each $\varepsilon > 0$ we can use the left estimate in (4-8) for $k \ge s - j(1-\varepsilon)$ and the right estimate otherwise, and obtain

$$|\langle \Upsilon_s^j * b_L, \bar{h}\rangle| \le \sum_\nu \sum_k |\langle R_{s\nu}^{jk} * B_k * b_L, \bar{h}\rangle| \lesssim 2^{-\varepsilon j} |L| \, \|b\|_{\dot{\mathcal{X}}_1} \|h\|_{\mathcal{Y}_\infty} \tag{4-9}$$

provided that M is chosen large enough to have $2\varepsilon < M\kappa$. The proof is thus completed by summing (4-9) over $L \in \mathcal{Q}$ with $\ell(L) = 2^{s-j}$ and later over s. □

5. Proof of Theorem B

Throughout this proof, C is a positive absolute dimensional constant which may vary at each occurrence without explicit mention. Most of the arguments in this section are contained in [Christ 1988, Section 3]; we reproduce the details for clarity.

Let $\psi(x) = \cos(2\pi(|x| - \delta/4))$. From the asymptotic expansion of the inverse Fourier transform of the multiplier of B_δ [Christ 1988, Section 3], which is $\mathcal{C}^\infty$ and radial, we obtain the kernel representation

$$B_\delta(x) = \sum_{s \ge 1} \sum_\nu K_{s,\nu}(x) + L(x).$$

Here

$$K_{s,\nu}(x) = \Omega_\nu(x')\psi(x)2^{-sd}\phi(2^{-s}x),$$

with Ω_ν a finite smooth partition of unity on the unit sphere S^{d-1} with sufficiently small support which is introduced for technical reasons, and ϕ a suitable smooth radial function supported in $A_1 = \{2^{-2} \le |x| \le 1\}$, while $L(x)$ is an integrable kernel with $L(x) \le C(1+|x|)^{-(d+1)}$, so that

$$Lf(x) \le C\mathrm{M}_1 f(x),$$

which can be ignored for our purposes. We can also think of ν as fixed and omit it from the notation, and consider the kernel $K = \{K_s\}$ as above. We are going to verify that conditions in Theorem C are satisfied by (the dual form to) B_δ. First of all, condition (SS) is obvious from the above discussion as $[K]_{0,\infty} < \infty$. Second, the (T) condition follows from the well-known estimate

$$\sup_{\mu,\nu} \|\Lambda_\mu^\nu\|_{L^2(\mathbb{R}^d)\times L^2(\mathbb{R}^d)} \le C;$$

see for instance [Duoandikoetxea and Rubio de Francia 1986, Theorem E]. In order to verify condition (L), let $\mathcal{Q}$ be a stopping collection with top Q. Let $b \in \mathcal{X}_1(\mathcal{Q})$; we change a bit the notation for b_s in this context by redefining

$$b_s := \sum_{s_L = s} b_L, \quad s \geq 1, \qquad b_0 := \sum_{s_L \leq 0} b_L.$$

It is easy to see that in this context if $b \in \mathcal{X}_1$ supported on Q and $h \in \mathcal{Y}_1$, one has

$$\Lambda_{\mathcal{Q},\mu,\nu}(b,h) = \left\langle \sum_{j\geq 1} \sum_{s \geq j} \varepsilon_s K_s * b_{s-j}, \bar{h} \right\rangle$$

for a suitable choice of signs $\{\varepsilon_s\} \in \{-1,0,1\}^{\mathbb{Z}}$, and the same for $\Lambda_{\mathcal{Q},\mu,\nu}(h,b)$ up to replacing b by b^{in}, restricting h to be supported on Q, transposing K_s, and subtracting the remainder terms, which are estimated by

$$|Q|\, \|b\|_{\mathcal{X}_1} \|h\|_{\mathcal{Y}_1}.$$

Theorem B is thus obtained from the next proposition via an application of Theorem C.

Proposition 5.1. *Let $\{\varepsilon_s\} \in \{-1,0,1\}^{\mathbb{Z}}$ be a choice of signs, $b \in \mathcal{X}_1$ and define*

$$\mathsf{K}(b,h) := \left\langle \sum_{j\geq 1} \sum_{s \geq j} \varepsilon_s K_s * b_{s-j}, \bar{h} \right\rangle.$$

There exists an absolute constant C, in particular uniform over $\{\varepsilon_s\} \in \{-1,0,1\}^{\mathbb{Z}}$, such that

$$|\mathsf{K}(b,h)| \leq \frac{Cp}{p-1} |Q|\, \|b\|_{\mathcal{X}_1} \|h\|_{\mathcal{Y}_p}.$$

Notice that here we do not need to require $b \in \dot{\mathcal{X}}_1$ as per the oscillatory nature of the problem.

Proof of Proposition 5.1. Given our choice of $\{\varepsilon_s\} \in \{-1,0,1\}^{\mathbb{Z}}$, we relabel $K_s := \varepsilon_s K_s$. It will be clear from the proof that the signs ε_s play no role. We split

$$\mathsf{K}(b,h) = \sum_{j\geq 1} \mathsf{K}^j(b,h), \quad \mathsf{K}^j(b,h) := \sum_{s \geq j} \langle K_s * b_{s-j}, \bar{h} \rangle.$$

The first estimate is a trivial one.

Lemma 5.2. *There exists $C > 0$ such that $|\mathsf{K}^j(b,h)| \leq C\, |Q|\, \|b\|_{\mathcal{X}_1} \|h\|_{\mathcal{Y}_1}$.*

Proof. This follows from applying Lemma 3.1 with $\beta = \infty$ to $K = \{K_s\}$, as it is immediate to see that for this kernel one has $[K]_{0,\infty} \leq C$ as already remarked. □

The second estimate, which is essentially contained in [Christ 1988, Section 3], is the one providing decay.

Lemma 5.3. *There exists $C, c > 0$ such that $|\mathsf{K}^j(b,h)| \leq C 2^{-cj} |Q|\, \|b\|_{\mathcal{X}_1} \|h\|_{\mathcal{Y}_2}$.*

It is easy to see that interpolating the above estimates yields

$$|\mathsf{K}^j(b,h)| \le C\,2^{-j\frac{c(p-1)}{p}}|Q|\,\|b\|_{\mathcal{X}_1}\|h\|_{\mathcal{Y}_p},$$

the summation of which yields Proposition 5.1.

Proof of Lemma 5.3. Let $\widetilde{K}_s(\cdot) = \overline{K_s(-\cdot)}$. We recall from [Christ 1988, Lemma 3.1] the estimates

$$\begin{aligned} &|K_s * \widetilde{K}_s(x)| \le C\,2^{-ds}(1+|x|)^{-\delta},\\ &\|K_s * \widetilde{K}_t\|_\infty \le C\,2^{-dt}2^{-\delta s}, \quad \forall s < t-1. \end{aligned} \tag{5-1}$$

By duality, it suffices to prove that

$$\|K_j * b_0\|_2^2 + \Big\|\sum_{s>j} K_s * b_{s-j}\Big\|_2^2 \le C\,2^{-cj}|Q|\,\|b\|_{\mathcal{X}_1}^2. \tag{5-2}$$

For the first term we use the first estimate in (5-1):

$$\|K_j * b_0\|_2^2 = |\langle b_0, K_j * \widetilde{K}_j * b_0\rangle| \le \|b_0\|_1\|K_j * \widetilde{K}_j * b_0\|_\infty \le C\,2^{-\min(\delta,d)j}|Q|\,\|b\|_{\mathcal{X}_1}^2.$$

The last inequality above follows from

$$\|K_j * \widetilde{K}_j * b_0\|_\infty \le 2^{-jd}\sum_{m=0}^{j} 2^{-m\delta}\sup_{x\in\mathbb{R}^d}\|b_0\|_{L^1(B(x,C2^m))} \le C\,2^{-\min(\delta,d)j}\|b\|_{\mathcal{X}_1},$$

where $B(x, C2^m)$ denotes a ball centered at x with radius $C2^m$. For the second term, we begin by quoting from [Christ 1988, inequality (3.2)] that

$$\|K_s * b_{s-j}\|_2^2 \le C\,2^{-\delta j}\|b\|_{\mathcal{X}_1}\|b_{s-j}\|_1. \tag{5-3}$$

Observe that

$$\begin{aligned} &\Big\|\sum_{s>j} K_s{*}b_{s-j}\Big\|_2^2\\ &\le \sum_{s>j}\|K_s{*}b_{s-j}\|_2^2 + 2\sum_s|\langle K_s{*}b_{s-j}, K_{s-1}{*}b_{s-1-j}\rangle| + 2\sum_t\sum_{j<s<t-1}|\langle \widetilde{K}_t{*}K_s{*}b_{s-j}, b_{t-j}\rangle|. \end{aligned} \tag{5-4}$$

The first two terms are bounded by

$$C\,2^{-\delta j}\|b\|_{\mathcal{X}_1}\sum_s\|b_{s-j}\|_1 \le C\,2^{-\delta j}|Q|\,\|b\|_{\mathcal{X}_1}^2,$$

according to (5-3) for the first one and Cauchy–Schwarz followed by (5-3) for the second. For the third term, from the second estimate of (5-1) and support considerations one has

$$\|\widetilde{K}_t * K_s * b_{s-j}\|_\infty \le C\big(\sup_{x\in\mathbb{R}^d}\|b_{s-j}\|_{L^1(B(x,C2^t))}\big)\|\widetilde{K}_t * K_s\|_\infty \le C\,2^{-\delta s}\|b\|_{\mathcal{X}_1}.$$

Therefore, the third summand in (5-4) is dominated by

$$C\,\|b\|_{\mathcal{X}_1}\sum_{t>j}\|b_{t-j}\|_1\sum_{j<s<t-1}2^{-\delta s} \le C\,2^{-\delta j}|Q|\,\|b\|_{\mathcal{X}_1}^2,$$

and collecting all the above estimates (5-2) follows. □

Appendix A: Verification of (2-10)–(2-11)

Let $\mathcal{Q}$ be a stopping collection with top Q, $h \in \mathcal{Y}_{q'}$, $b \in \mathcal{X}_{q'}$. Clearly we can assume $\operatorname{supp} h \subset Q$. By possibly replacing K_s by zero when $s \notin (\mu, \nu]$ we can ignore the truncations μ, ν in what follows and omit them from the notation. Recall the definitions (2-4), (2-5)

$$\Lambda_{\mathcal{Q}}(h,b) = \Lambda_{\mathcal{Q}}(h,b) - \sum_{\substack{R\in\mathcal{Q}\\ R\subset Q}} \Lambda_R(h,b) = \Lambda^{s_{\mathcal{Q}}}(h,b) - \sum_{\substack{R\in\mathcal{Q}\\ R\subset Q}} \Lambda^{s_R}(h\,\mathbf{1}_R, b)$$

and the decomposition

$$b = b^{\mathsf{in}} + b^{\mathsf{out}}, \quad b^{\mathsf{in}} = \sum_{\substack{L\in\mathcal{Q}\\ 3L\cap 2Q\neq\varnothing}} b_L, \quad b^{\mathsf{out}} = \sum_{\substack{L\in\mathcal{Q}\\ 3L\cap 2Q=\varnothing}} b_L.$$

We first estimate

$$|\Lambda_{\mathcal{Q}}(h, b^{\mathsf{out}})| \lesssim [K]_{0,q}\,|Q|\,\|h\|_{\mathcal{Y}_1}\|b\|_{\mathcal{X}_{q'}}, \tag{A-1}$$

which is a single-scale estimate. In fact, since $\operatorname{dist}(R, \operatorname{supp} b^{\mathsf{out}}) \geq \ell(R)/2$ for all $R \subset Q$, by virtue of the support restriction in (SS),

$$s < s_R \implies \int K_s(x,y)h(y)\mathbf{1}_R(y)b^{\mathsf{out}}(x)\,\mathrm{d}y\,\mathrm{d}x = 0.$$

Therefore, by the same argument used in (3-1),

$$|\Lambda^{s_{\mathcal{Q}}}(h, b^{\mathsf{out}})| \leq \int |K_{s_{\mathcal{Q}}}(x,y)||h(y)||b^{\mathsf{out}}(x)|\,\mathrm{d}y\,\mathrm{d}x \lesssim [K]_{0,q}\,|Q|\,\|h\|_{\mathcal{Y}_1}\|b\|_{\mathcal{X}_{q'}}. \tag{A-2}$$

Proceeding similarly, if $R \in \mathcal{Q}$, $R \subset Q$

$$|\Lambda^{s_R}(h\,\mathbf{1}_R, b^{\mathsf{out}})| \leq \int |K_{s_R}(x,y)||h\,\mathbf{1}_R(y)||b^{\mathsf{out}}(x)|\,\mathrm{d}y\,\mathrm{d}x \lesssim [K]_{0,q}\,|R|\,\|h\|_{\mathcal{Y}_1}\|b\|_{\mathcal{X}_{q'}}.$$

and the claimed (A-1) follows by summing the last display over $R \in \mathcal{Q}, R \subset Q$, which are pairwise disjoint, and combining the result with (A-2). The representation (2-10) will then be a simple consequence of the equality

$$\Lambda_{\mathcal{Q}}(h, b^{\mathsf{in}}) = \Big(\Lambda^{s_{\mathcal{Q}}}(h, b^{\mathsf{in}}) - \sum_{L\in\mathcal{Q}:3L\cap 2Q\neq\varnothing} \Lambda^{s_L}(h, b_L)\Big) + V_{\mathcal{Q}}(h,b), \tag{A-3}$$

where the remainder $V_{\mathcal{Q}}$ satisfies

$$|V_{\mathcal{Q}}(h,b)| \lesssim [K]_{0,q}\,|Q|\,\|h\|_{\mathcal{Y}_{q'}}\|b\|_{\mathcal{X}_{q'}}. \tag{A-4}$$

We turn to the proof of (A-3). We will use below without explicit mention that whenever $L, R \in \mathcal{Q}$ with $3R \cap 3L \neq \varnothing$, we have $|s_L - s_R| < 8$, a consequence of the separation property (2-3). First of all, the restriction on the support (SS) gives that

$$\sum_{R\in\mathcal{Q}} \Lambda^{s_R}(h\,\mathbf{1}_R, b^{\mathsf{in}}) = \sum_{R\in\mathcal{Q}} \sum_{\substack{L\in\mathcal{Q}\\ 3L\cap 3R\neq\varnothing\\ 3L\cap 2Q\neq\varnothing}} \Lambda^{s_R}(h\,\mathbf{1}_R, b_L), \tag{A-5}$$

as $\Lambda^{s_R}(h\,\mathbf{1}_R, b_L) = 0$ unless $3L \cap 3R$ is nonempty. As there are at most 16 s-scales in each difference $\Lambda^{s_L} - \Lambda^{s_R}$, using the trivial estimate (3-1) with $\beta = q$ for each such scale yields

$$\begin{aligned}\sum_{R\in\mathcal{Q}} \sum_{\substack{L\in\mathcal{Q}\\ 3L\cap 3R\neq\varnothing\\ 3L\cap 2Q\neq\varnothing}} &\big|\Lambda^{s_L}(h\,\mathbf{1}_R, b_L) - \Lambda^{s_R}(h\,\mathbf{1}_R, b_L)\big| \\ &\lesssim [K]_{0,q}\|h\|_{\mathcal{Y}_{q'}} \sum_{R\in\mathcal{Q}} \sum_{\substack{L\in\mathcal{Q}\\ 3L\cap 3R\neq\varnothing\\ 3L\cap 2Q\neq\varnothing}} \|b_L\|_1 \\ &\lesssim [K]_{0,q}\|h\|_{\mathcal{Y}_{q'}}\|b\|_{\mathcal{X}_1} \sum_{R\in\mathcal{Q}} |R| \lesssim [K]_{0,q}\,|Q|\,\|h\|_{\mathcal{Y}_{q'}}\|b\|_{\mathcal{X}_1}. \qquad \text{(A-6)}\end{aligned}$$

Recalling the second property of stopping collections in (2-3), we have the decomposition

$$h = h^{\mathsf{in}} + h^{\mathsf{out}}, \quad h^{\mathsf{in}} := h\,\mathbf{1}_{\bigcup_{R\in\mathcal{Q}} R}, \quad \operatorname{supp} h^{\mathsf{out}} \cap \Big(\bigcup_{\substack{L\in\mathcal{Q}\\ 3L\cap 2Q\neq\varnothing}} 9L \Big) = \varnothing.$$

Therefore, up to including the error term of (A-6) in (A-4), (A-5) can be rewritten as

$$\begin{aligned}\sum_{R\in\mathcal{Q}} \sum_{\substack{L\in\mathcal{Q}\\ 3L\cap 3R\neq\varnothing\\ 3L\cap 2Q\neq\varnothing}} \Lambda^{s_L}(h\,\mathbf{1}_R, b_L) &= \sum_{\substack{L\in\mathcal{Q}\\ 3L\cap 2Q\neq\varnothing}} \Lambda^{s_L}(h^{\mathsf{in}}, b_L) - \sum_{\substack{L\in\mathcal{Q}\\ 3L\cap 2Q\neq\varnothing}} \Lambda^{s_L}(\tilde{h}_L, b_L), \\ \tilde{h}_L &= \sum_{\substack{R\in\mathcal{Q}\\ 3L\cap 3R=\varnothing}} h\,\mathbf{1}_R, \quad \operatorname{supp}\tilde{h}_L \subset \mathbb{R}^d \setminus 3L. \end{aligned} \qquad \text{(A-7)}$$

We note that all the terms in the second sum on the right-hand side of the first line of (A-7) vanish due to the support restriction on K_s, as all the scales appearing are less than or equal to s_L and $\operatorname{supp} b_L \subset L$. The reasoning beginning with decomposition (A-5) leads thus to the equality, up to tolerable error terms,

$$\sum_{R\in\mathcal{Q}} \Lambda^{s_R}(h\,\mathbf{1}_R, b^{\mathsf{in}}) = \sum_{\substack{L\in\mathcal{Q}\\ 3L\cap 2Q\neq\varnothing}} \Lambda^{s_L}(h, b_L) - \sum_{\substack{L\in\mathcal{Q}\\ 3L\cap 2Q\neq\varnothing}} \Lambda^{s_L}(h^{\mathsf{out}}, b_L). \qquad \text{(A-8)}$$

Finally the second term on the right-hand side of (A-8) also vanishes, by virtue of the restriction on the support of h^{out}, which does not intersect $9L$ for any L in the sum. Therefore, (A-8) is actually the equality

$$\sum_{\substack{R\in\mathcal{Q}\\ R\subset Q}} \Lambda^{s_R}(h\,\mathbf{1}_R, b^{\mathsf{in}}) = \sum_{R\in\mathcal{Q}} \Lambda^{s_R}(h\,\mathbf{1}_R, b^{\mathsf{in}}) = \sum_{\substack{L\in\mathcal{Q}\\ 3L\cap 2Q\neq\varnothing}} \Lambda^{s_L}(h, b_L) + V_{\mathcal{Q}}(h, b),$$

where $V_{\mathcal{Q}}(h, b)$ satisfies (A-4); the first equality in the above display is due to $\operatorname{supp} h \subset Q$. This equality clearly implies the sought after (A-3).

Appendix B: Sparse domination implies weak L^1 estimate

We show that if a sublinear operator T satisfies the sparse estimate (1-1) for $p_1 = 1$, $p_2 = r$ for some $1 \le r < \infty$ then T is of weak type $(1, 1)$. In particular, as mentioned in the Introduction, together with

Theorem A, this yields the weak L^1 estimate of T_Ω, which is the main result of [Seeger 1996]. The proof that follows is a simplified version of the arguments in [Culiuc et al. 2016a, Appendix A]; we are sure these arguments are well known but were unable to locate a precise reference.

Theorem E. *Suppose that the sublinear operator T has the following property: there exists $C > 0$ and $1 \le r < \infty$ such that for every f_1, f_2 bounded with compact support there exists a sparse collection $\mathcal{S}$ such that*

$$|\langle Tf_1, f_2\rangle| \le C \sum_{Q\in\mathcal{S}} |Q|\,\langle f_1\rangle_{1,Q}\langle f_2\rangle_{r,Q}. \tag{B-1}$$

Then $T : L^1(\mathbb{R}^d) \to L^{1,\infty}(\mathbb{R}^d)$ boundedly.

Proof. By standard arguments it suffices to verify that

$$\sup_{\|f_1\|_1=1} \sup_{G\subset\mathbb{R}^d} \inf_{\substack{G'\subset G\\ |G|\le 2|G'|}} \sup_{|f_2|\le \mathbf{1}_{G'}} |\langle Tf_1, f_2\rangle| \le C,$$

where f_1, f_2 are bounded and compactly supported and G has finite measure. Given such f_1 with $\|f_1\|_1 = 1$ and G of finite measure, define the sets

$$H := \{x\in\mathbb{R}^d : \mathrm{M}_1 f_1(x) > C|G|^{-1}\}, \quad \widetilde{H} := \bigcup_{Q\in\mathcal{Q}} 3Q, \quad \mathcal{Q} = \{\text{max. dyad. cube } Q : |Q\cap H| \ge 2^{-5}|Q|\}.$$

It is easy to see that $|\widetilde{H}| \le 2^{-10}|G|$ for suitable choice of C. Therefore the set $G' : G\setminus\widetilde{H}$ satisfies $|G| \le 2|G'|$. We make the preliminary observation that

$$\sup_{x\in H^c} \mathrm{M}_1 f_1(x) \le C|G|^{-1},$$

so that by interpolation

$$\|\mathrm{M}_1 f_1\|_{L^{p'}(H^c)} \le \big(\sup_{x\in H^c} \mathrm{M}_1 f_1(x)\big)^{1-\frac{1}{p'}} \|\mathrm{M}_1 f_1\|_{1,\infty}^{\frac{1}{p'}} \le C|G|^{-\left(1-\frac{1}{p'}\right)}, \tag{B-2}$$

where $p' > 1$ is chosen such that $p > r$. Fixing now any f_2 restricted to G', we apply the domination estimate, yielding the existence of a sparse collection $\mathcal{S}$ for which we have the estimate

$$|\langle Tf_1, f_2\rangle| \le C \sum_{Q\in\mathcal{S}} |Q|\,\langle f_1\rangle_{1,Q}\langle f_2\rangle_{r,Q}.$$

We claim that

$$|Q\cap H| \le 2^{-5}|Q| \quad \forall Q\in\mathcal{S}. \tag{B-3}$$

This is because if (B-3) fails for Q, we know Q must be contained in $3Q'$ for some $Q'\in\mathcal{Q}$. But the support of f_2 is contained in $\widetilde{H}^c$, which does not intersect $3Q'$, whence $\langle f_2\rangle_{r,Q} = 0$. Relation (B-3) has the consequence that if $\{E_Q : Q\in\mathcal{S}\}$ denote the distinguished pairwise disjoint subsets of $Q\in\mathcal{S}$

with $|E_Q| \geq 2^{-2}|Q|$, the sets $\widetilde{E}_Q := E_Q \cap H^c$ are also pairwise disjoint and $|\widetilde{E}_Q| \geq 2^{-3}|Q|$. Therefore, since the union of $\widetilde{E}_Q$ is contained in H^c, by standard arguments we arrive at

$$|\langle Tf_1, f_2\rangle| \leq C \sum_{Q\in\mathcal{S}} |Q| \langle f_1\rangle_{1,Q} \langle f_2\rangle_{r,Q} \leq C \sum_{Q\in\mathcal{S}} |\widetilde{E}_Q| \langle f_1\rangle_{1,Q} \langle f_2\rangle_{r,Q} \leq C \int_{H^c} \mathrm{M}_1 f(x) \mathrm{M}_r f_2(x)\,\mathrm{d}x$$
$$\leq C \|\mathrm{M}_1 f_1\|_{L^{p'}(H^c)} \|\mathrm{M}_r f_2\|_{L^p(\mathbb{R}^d)} \leq C|G|^{-\left(1-\frac{1}{p'}\right)} |G|^{\frac{1}{p}} \leq C,$$

using (B-2) in the last step. □

Acknowledgments

This work was initiated while Conde-Alonso was visiting the Department of Mathematics at the University of Virginia during the Fall semester of 2016; the kind hospitality of the Department is gratefully acknowledged.

The authors have greatly benefited from their participation in the online seminar series Sparse Domination of Singular Integral Operators hosted by Georgia Tech in the Fall 2016 semester. In particular, they would like to thank Ben Krause, Michael Lacey, and Dario Mena Arias for their inspiring seminar talks on their current and forthcoming works related to sparse domination in the arithmetic setting.

The authors are grateful to Alexander Barron and Jill Pipher for stimulating discussions on the subject of this article which led to a simplification of one of the assumptions in Theorem C, and to David Cruz-Uribe, Javier Duoandikoetxea and Carlos Pérez for providing useful references and remarks concerning the weighted theory of rough singular integrals.

References

[Benea et al. 2017] C. Benea, F. Bernicot, and T. Luque, "Sparse bilinear forms for Bochner Riesz multipliers and applications", *Trans. London Math. Soc.* **4**:1 (2017), 110–128.

[Bernicot et al. 2016] F. Bernicot, D. Frey, and S. Petermichl, "Sharp weighted norm estimates beyond Calderón–Zygmund theory", *Anal. PDE* **9**:5 (2016), 1079–1113. MR Zbl

[Bui et al. 2017] T. A. Bui, J. M. Conde-Alonso, X. T. Duong, and M. Hormozi, "A note on weighted bounds for singular operators with nonsmooth kernels", *Studia Math.* **236**:3 (2017), 245–269. MR Zbl

[Calderón and Zygmund 1956] A. P. Calderón and A. Zygmund, "On singular integrals", *Amer. J. Math.* **78** (1956), 289–309. MR Zbl

[Carro and Domingo-Salazar 2016] M. Carro and C. Domingo-Salazar, "Weighted weak-type (1,1) estimates for radial Fourier multipliers via extrapolation theory", preprint, 2016, http://www.maia.ub.edu/~domingo/publis/PaperBochnerRiesz_F.pdf. To appear in *J. Anal. Math.*

[Christ 1988] M. Christ, "Weak type (1, 1) bounds for rough operators", *Ann. of Math.* (2) **128**:1 (1988), 19–42. MR Zbl

[Christ and Rubio de Francia 1988] M. Christ and J. L. Rubio de Francia, "Weak type (1, 1) bounds for rough operators, II", *Invent. Math.* **93**:1 (1988), 225–237. MR Zbl

[Conde-Alonso and Rey 2016] J. M. Conde-Alonso and G. Rey, "A pointwise estimate for positive dyadic shifts and some applications", *Math. Ann.* **365**:3-4 (2016), 1111–1135. MR Zbl

[Cruz-Uribe et al. 2011] D. V. Cruz-Uribe, J. M. Martell, and C. Pérez, *Weights, extrapolation and the theory of Rubio de Francia*, Operator Theory: Advances and Applications **215**, Springer, 2011. MR Zbl

[Culiuc et al. 2016a] A. Culiuc, F. D. Plinio, and Y. Ou, "Domination of multilinear singular integrals by positive sparse forms", preprint, 2016. arXiv

[Culiuc et al. 2016b] A. Culiuc, F. D. Plinio, and Y. Ou, "Uniform sparse domination of singular integrals via dyadic shifts", preprint, 2016. To appear in *Math. Res. Lett.* arXiv

[Di Plinio and Lerner 2014] F. Di Plinio and A. K. Lerner, "On weighted norm inequalities for the Carleson and Walsh–Carleson operator", *J. Lond. Math. Soc.* (2) **90**:3 (2014), 654–674. MR Zbl

[Duoandikoetxea 1993] J. Duoandikoetxea, "Weighted norm inequalities for homogeneous singular integrals", *Trans. Amer. Math. Soc.* **336**:2 (1993), 869–880. MR Zbl

[Duoandikoetxea and Rubio de Francia 1986] J. Duoandikoetxea and J. L. Rubio de Francia, "Maximal and singular integral operators via Fourier transform estimates", *Invent. Math.* **84**:3 (1986), 541–561. MR Zbl

[Grafakos and Stefanov 1999] L. Grafakos and A. Stefanov, "Convolution Calderón–Zygmund singular integral operators with rough kernels", pp. 119–143 in *Analysis of divergence* (Orono, ME, 1997), edited by W. O. Bray, Birkhäuser, Boston, 1999. MR Zbl

[Hytönen et al. 2012] T. Hytönen, C. Pérez, and E. Rela, "Sharp reverse Hölder property for A_∞ weights on spaces of homogeneous type", *J. Funct. Anal.* **263**:12 (2012), 3883–3899. MR Zbl

[Hytönen et al. 2017] T. P. Hytönen, L. Roncal, and O. Tapiola, "Quantitative weighted estimates for rough homogeneous singular integrals", *Israel J. Math.* **218**:1 (2017), 133–164. MR

[Krause and Lacey 2016] B. Krause and M. T. Lacey, "Sparse bounds for random discrete Carleson theorems", preprint, 2016. arXiv

[Lacey 2017] M. T. Lacey, "An elementary proof of the A_2 bound", *Israel J. Math.* **217**:1 (2017), 181–195. MR

[Lacey and Mena Arias 2017] M. T. Lacey and D. Mena Arias, "The sparse $T(1)$ theorem", *Houston J. Math.* **43**:1 (2017), 111–127.

[Lacey and Spencer 2017] M. T. Lacey and S. Spencer, "Sparse bounds for oscillatory and random singular integrals", *New York J. Math.* **23** (2017), 119–131. Zbl

[Lerner 2013] A. K. Lerner, "A simple proof of the A_2 conjecture", *Int. Math. Res. Not.* **2013**:14 (2013), 3159–3170. MR Zbl

[Lerner 2016] A. K. Lerner, "On pointwise estimates involving sparse operators", *New York J. Math.* **22** (2016), 341–349. MR Zbl

[Lerner and Nazarov 2015] A. K. Lerner and F. Nazarov, "Intuitive dyadic calculus: the basics", preprint, 2015. arXiv

[Luque et al. 2015] T. Luque, C. Pérez, and E. Rela, "Optimal exponents in weighted estimates without examples", *Math. Res. Lett.* **22**:1 (2015), 183–201. MR Zbl

[Moen 2012] K. Moen, "Sharp weighted bounds without testing or extrapolation", *Arch. Math.* (*Basel*) **99**:5 (2012), 457–466. MR Zbl

[Pérez et al. 2016] C. Pérez, I. Rivera-Rios, and L. Roncal, "A_1 theory of weights for rough homogeneous singular integrals and commutators", preprint, 2016. arXiv

[Seeger 1996] A. Seeger, "Singular integral operators with rough convolution kernels", *J. Amer. Math. Soc.* **9**:1 (1996), 95–105. MR Zbl

[Shi and Sun 1992] X. L. Shi and Q. Y. Sun, "Weighted norm inequalities for Bochner–Riesz operators and singular integral operators", *Proc. Amer. Math. Soc.* **116**:3 (1992), 665–673. MR Zbl

[Stein 1993] E. M. Stein, *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*, Princeton Mathematical Series **43**, Princeton University Press, 1993. MR Zbl

[Vargas 1996] A. M. Vargas, "Weighted weak type $(1, 1)$ bounds for rough operators", *J. London Math. Soc.* (2) **54**:2 (1996), 297–310. MR Zbl

[Volberg and Zorin-Kranich 2016] A. Volberg and P. Zorin-Kranich, "Sparse domination on non-homogeneous spaces with an application to A_p weights", preprint, 2016. arXiv

[Watson 1990] D. K. Watson, "Weighted estimates for singular integrals via Fourier transform estimates", *Duke Math. J.* **60**:2 (1990), 389–399. MR Zbl

Received 19 Jan 2017. Accepted 24 Apr 2017.

JOSÉ M. CONDE-ALONSO: jconde@mat.uab.cat
Departament de Matemàtiques, Facultat de Ciències, Universitat Autònoma de Barcelona, 08193 Barcelona, Spain

AMALIA CULIUC: amalia@math.gatech.edu
School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, United States

FRANCESCO DI PLINIO: francesco.diplinio@virginia.edu
Department of Mathematics, University of Virginia, Kerchof Hall, Box 400137, Charlottesville, VA 22904, United States

YUMENG OU: yumengou@mit.edu
Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, United States

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at msp.org/apde.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in APDE are usually in English, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with details about how your graphics were generated.

White space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Analysis & PDE

Volume 10 No. 5 2017

2157-5045(2017)10:5;1-T