Analysis & PDE

Volume 2 No. 2 2009

mathematical sciences publishers

Analysis & PDE
pjm.math.berkeley.edu/apde

See inside back cover or pjm.math.berkeley.edu/apde for submission instructions.

Regular subscription rate for 2009: $180.00 a year ($120.00 electronic only).
Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840, USA.

Analysis & PDE, at Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

PUBLISHED BY
mathematical sciences publishers
http://www.mathscipub.org
A NON-PROFIT CORPORATION
Typeset in LaTeX

ANALYSIS AND PDE
Vol. 2, No. 2, 2009

LOWER ESTIMATES ON MICROSTATES FREE ENTROPY DIMENSION

DIMITRI SHLYAKHTENKO

By proving that certain free stochastic differential equations with analytic coefficients have stationary solutions, we give a lower estimate on the microstates free entropy dimension of certain n-tuples $X_1, \dots, X_n$. In particular, we show that $\delta_0(X_1, \dots, X_n) \geq \dim_{M\bar{\otimes}M^o} V$, where $M = W^*(X_1, \dots, X_n)$ and $V = \{(\partial(X_1), \dots, \partial(X_n)) : \partial \in \mathscr{C}\}$ is the set of values of derivations $A = \mathbb{C}[X_1, \dots X_n] \to A \otimes A$ with the property that $\partial^*\partial(A) \subset A$. We show that for q sufficiently small (depending on n) and $X_1, \dots, X_n$ a q-semicircular family, $\delta_0(X_1, \dots, X_n) > 1$. In particular, for small q, q-deformed free group factors have no Cartan subalgebras. An essential tool in our analysis is a free analog of an inequality between Wasserstein distance and Fisher information introduced by Otto and Villani (and also studied in the free case by Biane and Voiculescu).

1. Introduction

We present in this paper a general technique for proving lower estimates for Voiculescu's microstates free entropy dimension δ_0. The free entropy dimension δ_0 was introduced in [Voiculescu 1994; 1996] and is a number associated to an n-tuple of self-adjoint elements $X_1, \dots, X_n$ in a tracial von Neumann algebra. This quantity has been used by various authors [Voiculescu 1996; Ge 1998; Ge and Shen 2002; Ştefan 2005; Jung 2007] to prove a number of very important results in von Neumann algebras. These results often take the form: *If* $\delta_0(X_1, \dots, X_n) > 1$, *then* $M = W^*(X_1, \dots, X_n)$ *cannot have certain decomposition properties* (for example, is non-Γ, has no Cartan subalgebras, is not a nontrivial tensor product and so on). For this reason, it is important to know if some given von Neumann algebra has a set of generators with the property that $\delta_0 > 1$. We prove that this is the case (for small values of q) for the "q-deformed free group factors" of [Bożejko and Speicher 1991].

Theorem 1. *For a fixed* N *and all* $|q| < (4N^3+2)^{-1}$, *the* q*-semicircular family* $X_1, \dots, X_N$ *satisfies* $\delta_0(X_1, \dots, X_N) > 1$ *and* $\delta_0(X_1, \dots, X_N) \geq N(1 - q^2N(1-q^2N)^{-1})$.

The theorem applies for $|q| \leq 0.029$ if $N = 2$. Combined with the available results on free entropy dimension, we obtain that, in this range of values of q, the algebras $\Gamma_q(\mathbb{R}^N) = W^*(X_1, \dots, X_N)$ have no Cartan subalgebras (or, more generally, that $\Gamma_q(\mathbb{R}^N)$, when viewed as a bimodule over any of its abelian subalgebras, contain a coarse subbimodule). Theorem 1 also implies that these algebras are prime (this was already proved in [Shlyakhtenko 2004] using the techniques of [Ozawa 2004]).

The free entropy dimension δ_0 is closely related to L^2 Betti numbers [Connes and Shlyakhtenko 2005; Mineyev and Shlyakhtenko 2005] — more precisely, with Murray–von Neumann dimensions of

MSC2000: 46L54.
Keywords: free stochastic calculus, free probability, von Neumann algebras, q-semicircular elements.
Research supported by NSF grant DMS-0555680.

spaces of certain derivations. For example, the nonmicrostates free entropy dimension δ^* (which is the nonmicrostates "relative" of δ_0) is in many cases equal to L^2 Betti numbers of the underlying (nonclosed) algebra [Mineyev and Shlyakhtenko 2005; Shlyakhtenko 2006]. It is known that $\delta_0 \leq \delta^*$ and thus it is important to find lower estimates for δ_0 in terms of dimensions of spaces of derivations. To this end we prove.

Theorem 2. *Let (A, τ) be a finitely-generated algebra with a positive trace τ and generators $X_1, \dots, X_N$, and let $\mathrm{Der}_c(A; A \otimes A)$ denote the space of derivations from A to $A \otimes A$ which are L^2 closable and such that $\partial^*\partial(X_j) \in A$. Consider the A,A-bimodule*

$$V = \{(\delta(X_1), \dots, \delta(X_n)) : \delta \in \mathrm{Der}_c(A; A \otimes A)\} \subset (A \otimes A)^N.$$

Finally, assume that $M = W^*(A, \tau)$ can be embedded into the ultrapower of the hyperfinite II_1 factor. Then

$$\delta_0(X_1, \dots, X_n) \geq \dim_{M \bar{\otimes} M^o} \overline{V}^{L^2(A \otimes A, \tau \otimes \tau)^N}.$$

We actually prove Theorem 2 under a less restrictive assumption: we require that $\delta(X_j)$ and $\delta^*\delta(X_j)$ be "analytic" as functions of $X_1, \dots, X_N$; more precisely, there should exist noncommutative power series Ξ_j and ξ_j with sufficiently large multiradii of convergence so that $\delta(X_j) = \Xi_j(X_1, \dots, X_N)$ and $\delta^*\delta(X_j) = \xi_j(X_1, \dots, X_N)$; see Theorem 16 below for a precise statement.

This theorem is a rich source of lower estimates for δ_0. For example, if $T \in A \otimes A$, then

$$\delta : X \mapsto [X, T] = XT - TX$$

is a derivation in $\mathrm{Der}_c(A; A \otimes A)$. If $W^*(A)$ is diffuse, then the map

$$L^2(A \otimes A) \ni T \mapsto ([T, X_1], \dots, [T, X_N]) \to L^2(A \otimes A)^N$$

is injective and thus the dimension over $M \bar{\otimes} M^o$ of its image is the same as the dimension of $L^2(A \otimes A)$, that is, 1. Hence $\dim_{M \bar{\otimes} M^o} \overline{V} \geq 1$ and so $\delta_0(X_1, \dots, X_n) \geq 1$ if $W^*(A)$ is R^ω embeddable ("hyperfinite monotonicity" in [Jung 2003b]).

If the two tuples $X_1, \dots, X_m$ and $X_{m+1}, \dots X_N$ are freely independent and each generates a diffuse von Neumann algebra, then for $T \in A \otimes A$ the derivation δ defined by $\delta(X_j) = [X_j, T]$ for $1 \leq j \leq m$ and $\delta(X_j) = 0$ for $m + 1 \leq j \leq N$ is also in $\mathrm{Der}_c(A)$. Then one easily gets that $\dim_{M \bar{\otimes} M^o} \overline{V} > 1$ (indeed, V contains vectors of the form $([T, X_1], \dots, [T, X_m], 0, \dots, 0)$, $T \in L^2(A \otimes A)$, and so its closure is strictly larger than the closure of the set of all vectors $([T, X_1], \dots, [T, X_N])$, $T \in L^2(A \otimes A)$). Thus $\delta_0(X_1, \dots, X_N) > 1$ if $W^*(A)$ is R^ω embeddable.

If $X_1, \dots, X_N$ are such that their conjugate variables [Voiculescu 1998] are polynomials, then the difference quotient derivations are in Der_c and thus $V = (A \otimes A)^N$, and so $\delta_0 = N$ (if $W^*(A)$ is R^ω embeddable).

In the case that $X_1, \dots, X_N$ are generators of the group algebra $\mathbb{C}\Gamma$ of a discrete group Γ,

$$\delta^*(X_1, \dots, X_N) = \beta_1^{(2)}(\Gamma) - \beta_0^{(2)}(\Gamma) + 1,$$

where $\beta_j^{(2)}$ are the L^2 Betti numbers of Γ (see [Lück 2002] for a definition). It is therefore natural to ask whether the same holds true for δ_0 instead of δ^* for some class of groups. If this is true, then knowing that

$\beta_1^{(2)}(\Gamma) \neq 0$ implies that $\delta_0 > 1$ and thus the group algebra has a variety of properties that we explained above (see also [Peterson 2009]).

It is clearly necessary for the equality $\delta_0 = \beta_1^{(2)} - \beta_0^{(2)} + 1$ that Γ can be embedded into the ultrapower of the hyperfinite II_1 factor (because otherwise δ_0 would be $-\infty$). In particular, one is tempted to conjecture that equality holds at least in the case when Γ is residually finite.

Theorem 2 implies a result like the one in [Brown et al. 2008]:

Theorem 3. *Assume that Γ is embeddable into the unitary group of the ultrapower of the hyperfinite II_1 factor. Then*

$$\delta_0(\Gamma) \geq \dim_{L(\Gamma)} \overline{\{c : \Gamma \to \mathbb{C}\Gamma \text{ } cocycle\}}.$$

In particular, if Γ belongs to the class of groups containing all groups with $\beta_1^{(2)} = 0$ and closed under amalgamated free products over finite subgroups, passage to finite index subgroups and finite extensions, then

$$\delta_0(\Gamma) = \beta_1^{(2)}(\Gamma) - \beta_0^{(2)}(\Gamma) + 1.$$

Let us now describe the main idea of the present paper. Our main result states that if the von Neumann algebra $M = W^*(X_1, \dots, X_n)$ can be embedded into the ultrapower of the hyperfinite II_1 factor, then

$$\delta_0(X_1, \dots, X_n) \geq \dim_{M \bar{\otimes} M^o} V, \tag{1-1}$$

where $V = \overline{\left\{(\partial(X_1), \dots, \partial(X_n)) : \partial \in \mathscr{C}\right\}}^{L^2}$ and $\mathscr{C}$ is some class of derivations from the algebra of non-commutative polynomials $\mathbb{C}[X_1, \dots, X_n]$ to $L^2(M) \bar{\otimes} L^2(M^o)$, which will be made precise later.

The quantity $\delta_0(X_1, \dots, X_n)$ is, very roughly, a kind of Minkowski dimension ("relative" to R^ω) of the set $\mathscr{V}$ of embeddings of M into R^ω, the ultrapower of the hyperfinite II_1 factor (indeed, the set of such embeddings can be identified with the set of images under the embedding of the generators $X_1, \dots, X_n$, that is, with the set of microstates for $X_1, \dots, X_n$). On the other hand, $\dim_{M \bar{\otimes} M^o} V$ is a linear dimension (relative to $M \bar{\otimes} M^o$) of a certain vector space. If we could find an interpretation for V as a subspace of a "tangent space" to $\mathscr{V}$, then the inequality (1-1) takes the form of the inequality linking the Minkowski dimension of a manifold with the linear dimension of its tangent space. One natural proof of such an inequality would involve proving that a linear homomorphism of the tangent space to a manifold at some point can be exponentiated to a local diffeomorphism of a neighborhood of that point.

Thus an essential step in proving a lower inequality on free entropy dimension is to find an analog of such an exponential map.

This leads to the idea, given a matrix $Q_{ij} \in \left(L^2(M) \bar{\otimes} L^2(M^o)\right)^n$ of values of derivations (so that $Q_{ij} = \partial_j(X_i)$ for some n-tuple of derivation ∂_j belonging to our class $\mathscr{C}$), to try to associate to Q a one-parameter deformation α_t of a given embedding $\alpha = \alpha_0$ of M into R^ω. It turns out that there are two (related) ways to do this.

The first approach comes from the idea that we (at least in principle) know how to exponentiate derivations from an algebra to itself (the result should be a one-parameter automorphism group of the algebra). We thus try to extend $\partial = \partial_1 \oplus \cdots \oplus \partial_n$ to a derivation of a larger algebra $\mathscr{A} = \mathbb{C}[X_1, \dots, X_n, S_1, \dots, S_n]$, where $S_1, \dots, S_n$ are free from $X_1, \dots, X_n$ and form a free semicircular family. The key point is that the closure in $L^2(\mathscr{A})$ of $\operatorname{span}(MS_1M + \cdots + MS_nM)$ is isomorphic to $[L^2(M) \otimes L^2(M)]^n$. The inverse of this isomorphism takes an n-tuple $a = (a_1 \otimes b_1, \dots, a_n \otimes b_n)$ to $\sum a_j S_j b_j$, which we denote by

$a\#S$. We now define a new derivation $\tilde{\partial}$ of $\mathscr{A}$ with values in $L^2(\mathscr{A})$ by $\tilde{\partial}(X_j)=\partial(X_j)\#S$. To be able to exponentiate $\tilde{\partial}$, we need to make sure that it is antihermitian as an unbounded operator on $L^2(\mathscr{A})$, which naturally leads to the equation $\tilde{\partial}(S_j)=-\partial^*(\zeta_j)$, where $\zeta_j=(0,\dots,1\otimes 1,\dots,0)$ (j-th entry nonzero). One can check that if ζ_j is in the domain of ∂^* for all j, then $\tilde{\partial}$ is a closable operator which has an antihermitian extension, and so it can be exponentiated to a one-parameter group of automorphisms α_t of $L^2(\mathscr{A})$. Unfortunately, unless we know more about the derivation ∂ (such as, for example, assuming that $\tilde{\partial}(\mathscr{A})\subset\mathscr{A}$), we cannot prove that α_t takes $W^*(\mathscr{A})$ to $W^*(\mathscr{A})$. However, if this is the case, then we do get a one-parameter family of embeddings $\alpha_t|_M : M\to M * L(\mathbb{F}(n))\subset R^\omega$. We explain this approach in more detail in the Appendix.

The second approach was suggested to us by A. Guionnet, to whom we are indebted for generously allowing us to publish it. The idea involves considering the free stochastic differential equation

$$dX_j(t)=\sum_i Q_{ij}(X_1(t),\dots,X_n(t))\#dS_i-\tfrac{1}{2}\xi_j(X_1(t),\dots,X_n(t)),\qquad X_j(0)=X_j, \tag{1-2}$$

where $\partial(X_j)=(Q_{1j},\dots,Q_{nj})\in\big(L^2(M)\mathbin{\bar{\otimes}}L^2(M^o)\big)^n$ and $\xi_j(X_1,\dots,X_n)=\partial^*\partial(X_j)$. One difficulty in even phrasing the problem is that it is not quite clear what is meant by Q_{ij} and ξ_j applied to their arguments (in the classical case, this would mean a function applied to the random variable $X(t)$). However, if this equation can be formulated and has a stationary solution $X(t)$ (namely one for which the law does not depend on t), then the map $\alpha_t : X_j\mapsto X_j(t^2)$ determines a one-parameter family of embeddings of the von Neumann algebra M into some other von Neumann algebra $\mathscr{M}$ (generated by all $X(t):t\geq 0$). This can be carried out successfully if Q and ξ are sufficiently nice; this is this is the case, for example, when $X_1,\dots,X_n$ are q-semicircular variables, in which case Q and ξ can be taken to be analytic noncommutative power series.

Let us assume now that ∂ takes $\mathscr{B}=\mathbb{C}[X_1,\dots,X_n]$ to $\mathscr{B}\otimes\mathscr{B}^o$ and also $\partial^*(1\otimes 1)\in\mathscr{B}$ (this is the case, for example, if $X_1,\dots,X_n$ have polynomial conjugate variables [Voiculescu 1998]). Then both approaches work to actually give one a stronger statement: one gets a one-parameter family of embeddings $\alpha_t : M\to R^\omega$ so that $\|\alpha_t(X_j)-(X_j+t\sum_i Q_{ij}\#S_i)\|_2=O(t^2)$. Let us assume for the moment that $Q_{ij}=\delta_{ij}1\otimes 1$, so that our estimate reads

$$\|\alpha_t(X_j)-(X_j+tS_j)\|_2=O(t^2). \tag{1-3}$$

An estimate of this kind was used as a crucial step by Otto and Villani in their work on the classical transportation cost inequality [Otto and Villani 2000, §4 Lemma 2]; a free version (for $n=1$) is the key ingredient in the proof of free transportation cost inequality and free Wasserstein distance given in [Biane and Voiculescu 2001]. Indeed, since the law of $\alpha_t(X_j)$ is the same as X_j, one obtains after working out the error bounds an estimate on the noncommutative Wasserstein distance between the laws $\mu_{X_1,\dots,X_n}$ and $\mu_{X_1+tS_1,\dots,X_n+tS_n}$:

$$d_W(\mu_{X_1,\dots,X_n},\mu_{X_1+tS_1,\dots,X_n+tS_n})\leq\tfrac{1}{2}\Phi(X_1,\dots,X_n)^{1/2}t+O(t^2).$$

We now point out that this estimate is of direct relevance to a lower estimate on δ_0. Indeed, suppose that some n-tuple of $k\times k$ matrices $x_1,\dots,x_n$ has as its law approximately the law of $X_1,\dots,X_n$ (that is, $(x_1,\dots,x_n)\in\Gamma(X_1,\dots,X_n;k,l,\varepsilon)$ in the notation of [Voiculescu 1994]). Then (1-3) implies that by approximating $\alpha_t(X_j)$ with polynomials in $X_1,\dots,X_n,S_1,\dots,S_n$, one can find another n-tuple

$x'_1, \dots, x'_n$ with almost the same law as $X_1, \dots, X_n$, and so that $\|x'_j - (x_j + ts_j)\| \le Ct^2$ (here $s_1, \dots, s_n$ are some matrices whose law is approximately that of $S_1, \dots, S_n$, and which are approximately free from $x_1, \dots, x_n$). But this means that if one moves along a line starting at $x_1, \dots, x_n$ in the direction of $s_1, \dots, s_n$, then the distance to the set $\Gamma(X_1, \dots, X_n; k, l, \varepsilon)$ grows quadratically. Thus this line is *tangent* to the set $\Gamma(X_1, \dots, X_n; k, l, \varepsilon)$. From this one can derive estimates relating the packing numbers of $\Gamma(X_1, \dots, X_n; k, l, \varepsilon)$ and $\Gamma(X_1 + tS_1, \dots, X_n + tS_n; k, l, \varepsilon)$ which can be converted into a lower estimate on δ_0.

In conclusion, it is worth pointing out that the main obstacle that we face in trying to extend the estimate (1-1) to larger classes of derivations is the question of existence of stationary solutions of (1-2) for more general classes of functions Q and ξ (and not, surprisingly enough, the "usual" difficulties in dealing with sets of microstates).

2. Existence of stationary solutions

2.1. *Free SDEs with analytic coefficients.* The main result of this section states that a free stochastic differential equation of the form

$$dX_t = \Xi \# dS_t - \tfrac{1}{2}\xi_t dt$$

where X_t is an N-tuple of random variables has a stationary solution, as long as the coefficients Ξ and ξ are analytic (that is, they are noncommutative power series with sufficient radii of convergence).

2.1.1. *Estimates on certain operators appearing in free Ito calculus.* Let f be a noncommutative power series in N variables. We denote by $c_f(n)$ the maximal modulus of a coefficient of a monomial of degree n in f. Thus if $f = \sum f_{i_1 \dots i_n} X_{i_1} \cdots X_{i_n}$, then $c_f(n) = \max_{i_1 \dots i_n} |f_{i_1 \dots i_n}|$. We also write

$$\phi_f(z) = \sum c_f(n) z^n.$$

Then $\phi_f(z)$ is a formal power series in z. If ρ is the radius of convergence of ϕ_f, we'll say that $R = \rho/N$ is the multiradius of convergence of f.

We also write

$$\|f\|_\rho = \sum_{n \ge 0} c_f(n) N^n \rho^n \in [0, +\infty].$$

Note that $\|f\|_\rho = \sup_{|z| \le N\rho} |\phi_f(z)|$ (since all of the coefficients in the power series $\phi_f(z)$ are real and positive).

We denote by $\mathcal{F}(R)$ the collection of all power series f for which the multiradius of convergence is at least R. In other words, we require $\|f\|_\rho < \infty$ for all $\rho < R$.

Note that $\mathcal{F}_R$ is a complete topological vector space when endowed with the topology such that $T_i \to T$ if and only if $\|T_i - T\|_\rho \to 0$ for all $\rho < R$.

Let Ψ be a noncommutative power series in N variables having the form

$$\sum f_{i_1, \dots, i_k; j_1, \dots, j_l} Y_{i_1} \cdots Y_{i_k} \otimes Y_{j_1} \cdots Y_{j_l}.$$

We call Ψ a formal noncommutative power series with values in $\mathbb{C}\langle Y_1, \dots, Y_N \rangle^{\otimes 2}$. We write $c_\Psi(m, n)$ the maximal modulus of a coefficient of a monomial of the form $Y_{i_1} \cdots Y_{i_m} \otimes Y_{j_1} \cdots Y_{j_n}$ in Ψ. We let

$\phi_\Psi(z,w)=\sum_{n,m} c_{\psi(m,n)} z^m w^n$. We put

$$\|\Psi\|_\rho = \sup_{|z|,|w|\le N\rho} |\phi_\Psi(z,w)| = \phi_\Psi(N\rho,N\rho) = \sum_{n\ge 0}\Bigg(\sum_{k+l=n} c_\Psi(k,l)\Bigg) N^n\rho^n \in [0,+\infty].$$

We denote by $\mathcal{F}'(R)$ the collection of all noncommutative power series for which $\|\Psi\|_\rho<\infty$ for all $\rho<R$.

It will be convenient to use the following notation. Let $\phi(z_1,\dots,z_n)$, $\psi(z_1,\dots,z_n)$ be two formal power series (in commuting variables). We say that $\phi\prec\psi$ if all coefficients in ϕ, ψ are real and positive, and for each $k_1,\dots,k_n$, the coefficient of $z_1^{k_1}\cdots z_n^{k_n}$ in ϕ is less than or equal to the corresponding coefficient in ψ.

If $\mathcal{M}$ is a unital Banach algebra, $Y_1,\dots,Y_N\in\mathcal{M}$ and $\|Y_j\|<\rho$ for all j, then $\|g(Y_1,\dots,Y_n)\|\le\|g\|_\rho$ whenever g is in any one of the spaces $\mathcal{F}(R)$, or $\mathcal{F}'(R)$ (here the norm $\|g(Y_1,\dots,Y_n)\|$ denotes the norm on $\mathcal{M}$ or on the projective tensor product $\mathcal{M}^{\otimes 2}$, as appropriate).

We now collect some facts about power series:

- Let $f,g\in\mathcal{F}(R)$. Then $\phi_{fg}\prec\phi_f\phi_g$. In particular, $fg\in\mathcal{F}(R)$ and $\|fg\|_\rho\le\|f\|_\rho\|g\|_\rho$.
- Let $f=\sum f_{i_1\dots i_n}X_{i_1}\cdots X_{i_n}\in\mathcal{F}(R)$ and denote by $\mathcal{D}_{ij}f$ the formal power series

$$\mathcal{D}_{ij}f=\sum_{i_1\dots i_n}\sum_{k<l}\delta_{i_k=i}\delta_{i_l=j}f_{i_1\dots i_n}X_{i_{k+1}}\cdots X_{i_{l-1}}\otimes X_{i_{l+1}}\cdots X_{i_n}X_{i_1}\cdots X_{i_{k-1}}.$$

  Since a monomial $X_{i_1}\cdots X_{i_k}\otimes X_{j_1}\cdots X_{j_r}$ could arise in the expression for $\mathcal{D}_{ij}f$ in at most $r+1$ ways, $c_{\mathcal{D}_j f}(a,b)\le(b+1)c_f(a+b+2)$. Denote by $\hat\phi_f$ the power series

$$\hat\phi_f(z,w)=\sum_{n,m}(n+1)c_f(n+m+2)z^m w^n.$$

  Then $\phi_{\mathcal{D}_j f}\prec\hat\phi_f$. Since $\hat\phi_f(z,z)\prec\phi_f''(z)$, we conclude that

$$\|\mathcal{D}_{ij}f\|_\rho\le\sup_{|z|\le N\rho}|\phi_f''(z)|$$

  and in particular $\mathcal{D}_{ij}f\in\mathcal{F}'(R)$.
- Let $\Theta=\sum\Theta_{i_1\dots i_n;j_1\dots j_m}X_{i_1}\cdots X_{i_n}\otimes X_{j_1}\cdots X_{j_m}\in\mathcal{F}'(R)$, and let

$$\Psi=\sum\Psi_{i_1\dots i_n;j_1\dots j_m}X_{i_1}\dots X_{i_n}\otimes X_{j_1}\dots X_{j_m}\in\mathcal{F}'.$$

  Consider

$$\Psi\#_{in}\Theta=\sum\Psi_{t_1\dots t_a,s_1,\dots,s_b}\Theta_{i_1\dots i_n;j_1\dots j_m}X_{i_1}\cdots X_{i_n}\ X_{t_1}\cdots X_{t_a}\otimes X_{s_1}\cdots X_{s_b}\ X_{j_1}\cdots X_{j_m}.$$

  (In the simple case that $\Psi=A\otimes B$ and $\Theta=P\otimes Q$, where A, B, P, Q are monomials, we have $\Psi\#_{in}\Theta=PA\otimes BQ$, that is, $\#_{in}$ is the "inside" multiplication on $\mathcal{F}'(R)$). Then

$$c_{\Psi\#_{in}\Theta}(n,m)\le\sum_{k+l=n}\sum_{r+s=m}c_\Psi(k,r)c_\Theta(l,s),$$

and hence the coefficient of $z^n w^m$ in $\phi_{\Psi\#_{in}\Theta}(z,w)$ is dominated by the coefficient of $z^n w^m$ in $\phi_\Psi(z,w)\phi_\Theta(z,w)$. Consequently, $\phi_{\Psi\#_{in}\Theta} \prec \phi_\Psi\phi_\Theta$ and

$$\|\Psi\#_{in}\Theta\|_\rho \le \|\Psi\|_\rho\|\Theta\|_\rho.$$

In particular, $\Psi\#_{in}\Theta \in \mathcal{F}'(R)$. Similar estimates and conclusion of course hold for the "outside" multiplication $\Psi\#_{out}\Theta$, defined by

$$\Psi\#_{out}\Theta = \sum \Psi_{s_1,\dots,s_b;t_1\dots t_a}\Theta_{i_1\dots i_n;j_1\dots j_m} X_{t_1}\cdots X_{t_a}\ X_{i_1}\cdots X_{i_n} \otimes X_{j_1}\cdots X_{j_m}\ X_{s_1}\cdots X_{s_b}.$$

In that case we get $\phi_{\Psi\#_{out}\Theta}(z,w) \prec \phi_\Psi(w,z)\phi_\Theta(z,w)$ and $\|\Psi\#_{out}\Theta\|_\rho \le \|\Psi\|_\rho\|\Theta\|_\rho$.

- Let τ be a linear functional on the algebra of noncommutative polynomials in n variables, so that $|\tau(X_{i_1}\cdots X_{i_n})| \le R_0^n$ for all n. Given $\Theta = \sum \Theta_{i_1\dots i_n;j_1\dots j_m} X_{i_1}\cdots X_{i_n}\otimes X_{j_1}\cdots X_{j_m} \in \mathcal{F}'(R)$, assume that $R_0 < R$ and consider the formal sum

$$(1\otimes\tau)(\Theta) = \sum_{n,i_1,\dots,i_n}\Bigg(\sum_{m,j_1,\dots,j_m}\Theta_{i_1\dots i_n;j_1\dots j_m}\tau(X_{j_1}\cdots X_{j_m})\Bigg)X_{i_1}\cdots X_{i_n}.$$

More precisely, we consider the formal power series in which the coefficient of $X_{i_1}\cdots X_{i_n}$ is given by the sum

$$\sum_{m,j_1,\dots,j_m}\Theta_{i_1\dots i_n;j_1\dots j_m}\tau(X_{j_1}\cdots X_{j_m}).$$

But since $|\tau(X_{j_1}\cdots X_{j_m})| \le R_0^m$, this sum is bounded by the coefficient of z^n in the power series expansion of $\phi(z, NR_0)$ (as a function of z), and is convergent. Thus $\phi_{(1\otimes\tau)(\Theta)}(z) \prec \phi_\Theta(z, NR_0)$ and we readily see that $(1\otimes\tau)(\Theta)$ is well-defined, belongs to $\mathcal{F}(R)$, and moreover

$$\|(1\otimes\tau)(\Theta)\|_\rho \le \|\Theta\|_\rho,$$

whenever $\rho > R_0$.

- Let $f = \sum f_{i_1\dots i_n} X_{i_1}\cdots X_{i_n} \in \mathcal{F}(R)$ and consider the j-th cyclic partial derivative [Voiculescu 1999; 2002b]

$$\mathcal{D}_j f = \sum_{i_1\dots i_n}\sum_{l=1}^{n}\delta_{i_l=j}X_{i_{l+1}}\dots X_{i_n}\cdot X_{i_1}\cdots X_{i_{l-1}}.$$

Then we see that $\phi_{\mathcal{D}_j f} \prec (\phi_f)'$ and $\mathcal{D}_j f \in \mathcal{F}(R)$.

We now combine these estimates:

Lemma 4. *Let τ as above be a linear functional on the space of noncommutative polynomials in N variables satisfying $\tau(X_{i_1}\cdots X_{i_n}) \le R_0^n$. Let $R > R_0$ and assume that $\xi_j \in \mathcal{F}(R)$, $j = 1,\dots,N$, $\Psi = (\Psi_{ij}) \in M_{N\times N}\mathcal{F}'(R)$. For $f \in \mathcal{F}(R)$ let*

$$\mathcal{L}^{(\tau)}(f) = (1\otimes\tau)\Bigg(\sum_{ijk}\Psi_{jk}\#_{in}(\Psi_{ki}\#_{out}(\mathcal{D}_{ij}f))\Bigg) - \sum_j \tfrac{1}{2}\xi_j\mathcal{D}_j f.$$

Then $\mathcal{L}_j^{(\tau)}(f) \in \mathcal{F}(R)$ *and moreover for any* $R_0 < \rho < R$,

$$\|\mathcal{L}^{(\tau)}(f)\|_\rho \le \sum_{ijk} \|\Psi_{jk}\|_\rho \|\Psi_{ki}\|_\rho \cdot \sup_{|z|\le N\rho} |\phi_f''(z)| + \tfrac{1}{2}\sum_j \|\xi_j\|_\rho \sup_{|z|\le N\rho} |\phi_f'|,$$

$$\phi_{\mathcal{L}^{(\tau)}(f)}(z) \prec \sum_{ijk} \phi_{\Psi_{jk}}(z, NR_0)\phi_{\Psi_{ki}}(NR_0, z)\hat{\phi}_f(z, NR_0) + \tfrac{1}{2}\sum_j \phi_{\xi_j}(z)\phi_f'(z),$$

where $\hat{\phi}_f(z, w) = \sum_{n,m}(n+1)c_f(n+m+2)z^m w^n$.

For ϕ a power series in $z, w_1, \dots, w_k$ with multiradius of convergence bigger than ρ and all coefficients of monomials nonnegative, let $\phi_{w_1,\dots,w_k}(z) = \phi(z, w_1, \dots, w_k)$. Set

$$Q\phi(z, w_1, \dots w_{k+1}) = \widehat{\phi_{w_1,\dots,w_k}}(z, w_{k+1}) \quad \text{and} \quad D\phi(z, w_1, \dots, w_k) = \partial_z^2\phi(z, w_1, \dots, w_k).$$

We note that $\hat{\phi}(z, z) \prec \phi''(z)$, and that Q, D and $\mathfrak{D}$ are monotone for the ordering $\prec$. It follows that if κ_j, λ_j are some power series with radius of convergence bigger than ρ and positive coefficients, then for any $a_1, b_1, \dots, a_k, b_k \ge 0$ and any $R < \rho$,

$$\left[Q^{a_1}\kappa_1(z)D^{b_1}\lambda_1(z)\mu_1 Q^{a_2}\kappa_2(z)D^{b_2}\lambda_2(z)\cdots D^{b_k}\lambda_k\right]\Big|_{z=w_1=\cdots=w_{\sum b_k}=R}$$
$$\le \left[D^{a_1}\kappa_1(z)D^{b_1}\lambda_1(z)D^{a_2}\kappa_2(z)D^{b_2}\lambda_2(z)\cdots D^{b_k}\lambda_k\right]\Big|_{z=w_1=\cdots=w_{\sum b_k}=R}.$$

Now define

$$\hat{\mathcal{L}}\phi(z) = \sum_{ijk} \phi_{\psi_{jk}}(z, NR_0)\phi_{\Psi_{ki}}(NR_0, z)\phi''(z) + \tfrac{1}{2}\sum_j \phi_{\xi_j}(z)\phi'(z).$$

Then we have obtained the inequality

$$\phi_{\mathcal{L}^n f}(NR_0) \le \hat{\mathcal{L}}^n\phi_f(NR_0),$$

which we record as:

Lemma 5. *Let* $\hat{\mathcal{L}}\phi(z) = \sum_{ijk}\phi_{\psi_{jk}}(z, NR_0)\phi_{\Psi_{ki}}(NR_0, z)\phi''(z) + \frac{1}{2}\sum_j \phi_{\xi_j}(z)\phi(z)$ *and let* τ *be a trace so that for any monomial* P, $|\tau(P)| < R_0^n$, $n = \deg P$. *Then*

$$|\tau(\mathcal{L}^n f)| \le \hat{\mathcal{L}}^n\phi_f(NR_0).$$

2.1.2. *Analyticity of* $\partial^*\partial(X_j)$. Let us now assume that $\Xi = (\Xi_1, \dots, \Xi_N) \in \mathcal{F}'(R)$. Let $(X_1, \dots, X_N)$ be an N-tuple of self-adjoint operators in a tracial von Neumann algebra (M, τ) and assume that $\|X_j\| < R$ for all j. Let $\partial : L^2(M) \to L^2(M) \bar{\otimes} L^2(M)$ be the derivation densely defined on polynomials in $X_1, \dots, X_N$ by $\partial(X_j) = \Xi_j(X_1, \dots, X_N)$. We assume that $1 \otimes 1$ belongs to the domain of ∂^* and that there exists some $\zeta \in \mathcal{F}(R)$ so that $\partial^*(1 \otimes 1) = \zeta(X_1, \dots, X_N)$.

Lemma 6. *With the assumptions above, there exist* $\xi_j \in \mathcal{F}(R)$, $j = 1, \dots, N$, *so that*

$$\xi_j(X_1, \dots, X_N) = \partial^*\partial(X_j).$$

Proof. It follows from [Voiculescu 1998; Shlyakhtenko 1998] that under these assumptions, ∂ is closable. Moreover, for any a, b polynomials in $X_1, \dots, X_N$, $a \otimes b$ belongs to the domain of ∂^* and

$$\partial^*(a \otimes b) = a\zeta b + (1 \otimes \tau)[\partial(a)]b + a(\tau \otimes 1)[\partial(b)],$$

where $\zeta = \zeta(X_1, \dots, X_N) = \partial^*(1 \otimes 1)$.

Consider now formal power series in N variables having the form

$$\Theta = \sum \Theta_{i_1,\dots,i_k;j_1,\dots,j_l;t_1,\dots,t_r} Y_{i_1} \cdots Y_{i_k} \otimes Y_{j_1} \cdots Y_{j_l} \otimes Y_{t_1} \cdots Y_{t_r}.$$

We write $\phi_\Theta(z, w, v)$ for the power series whose coefficient of $z^m w^n v^k$ is equal to the maximum

$$\max\{|\Theta_{i_1,\dots,i_m;j_1,\dots,j_n;t_1,\dots,t_k}| : i_1, \dots, i_m, j_1, \dots, j_n, t_1, \dots, t_n \in \{1, \dots, N\}\}.$$

We denote by $\mathcal{F}''(R)$ the collection of all such power series for which ϕ_Θ has a multiradius of convergence at least NR.

Let $\mathcal{D}_1^{(s)} : \mathcal{F}'(R) \to \mathcal{F}''(R)$ be given by

$$\mathcal{D}_1^{(s)} \sum f_{i_1,\dots,i_k;j_1,\dots,j_l} Y_{i_1} \cdots Y_{i_k} \otimes Y_{j_1} \cdots Y_{j_l} = \sum f_{i_1,\dots,i_k;j_1,\dots,j_l} \sum_p \delta_{i_p=s} Y_{i_1} \cdots Y_{i_{p-1}} \otimes Y_{i_{p+1}} \cdots Y_{i_k} \otimes Y_{j_1} \cdots Y_{j_l}.$$

Then clearly $\phi_{\mathcal{D}_1^{(s)}(\Psi)}(z, z, w) \prec \partial_z \phi_\Psi(z, w)$ so that $\mathcal{D}_1^{(s)}\Psi$ indeed lies in $\mathcal{F}''(R)$ if $\Psi \in \mathcal{F}'(R)$.

Similarly, if we define for $\Psi \in \mathcal{F}'(R)$, $\Theta \in \mathcal{F}''(R)$

$$\Psi \#_{in}^{(1)} \Theta = \sum \Psi_{t_1\dots t_a,s_1,\dots,s_b} \Theta_{i_1\dots i_n;j_1\dots j_m;k_1\dots k_p} Y_{i_1} \cdots Y_{i_n}\ Y_{t_1} \cdots Y_{t_a} \otimes Y_{s_1} \cdots Y_{s_b}\ Y_{j_1} \cdots Y_{j_m} \otimes Y_{k_1} \cdots Y_{k_p},$$

then $\phi_{\Psi \#_{in}^{(1)} \Theta}(z, v, w) \prec \phi_\Psi(z, v) \phi_\Theta(z, v, w)$ and in particular $\Psi \#_{in}^{(1)} \Theta \in \mathcal{F}''(R)$. (Note that $\#_{in}^{(1)}$ corresponds to "multiplying around" the first tensor sign in Θ).

Finally, if τ is any linear functional so that $\tau(P) < R_0^{\deg P}$ for any monomial P and we put

$$M_2(\Psi) = \sum \Psi_{i_1,\dots,i_n;j_1,\dots,j_m;k_1,\dots k_p} Y_{i_1} \cdots Y_{i_n} \tau(Y_{j_1} \cdots Y_{j_m} Y_{k_1} \cdots Y_{k_p}),$$

then $\phi_{M_2(\Psi)}(z) \leq \phi_\Psi(z, NR_0, NR_0)$ and in particular $M_2(\Psi) \in \mathcal{F}(R)$ once $\Psi \in \mathcal{F}''(R)$ and $R_0 < R$. In the foregoing, we'll use the trace τ of M as our functional.

So if we put

$$T_1 \Theta = M_2\Big(\sum_s \Xi_s \#_{in}^{(1)} \mathcal{D}_1^{(s)}\Big),$$

then T_1 maps $\mathcal{F}'(R)$ into $\mathcal{F}(R)$.

Note that in the case that $\Theta = A \otimes B$, where A, B are monomials, $T_1 \Theta = (1 \otimes \tau)(\partial(A)) \cdot B$.

One can similarly define $T_2 : \mathcal{F}'(R) \to \mathcal{F}(R)$; it will have the property that $T_2 \Theta = A(\tau \otimes 1)(\partial(B))$.

Lastly, let $\zeta \in \mathcal{F}(R)$ and let $m : \mathcal{F}'(R) \to \mathcal{F}(R)$ be given by

$$m(\Theta) = \sum \Theta_{i_1,\dots,i_n;j_1,\dots,j_m} \zeta_{p_1,\dots,p_r} Y_{i_1} \cdots Y_{i_n} Y_{p_1,\dots p_r} Y_{j_1} \cdots Y_{j_m}.$$

Once again, $\phi_{m(\Theta)}(z) \prec \phi_\Theta(z, z) \phi_\zeta(z)$.

Let $Q(\Xi) = T_1(\Xi) + T_2(\Xi) + m(\Xi)$. We claim that $\xi = (Q(\Xi))(X_1, \dots, X_N) = \partial^*(\Xi(X_1, \dots, X_N))$.

Note that if Ξ_n is a partial sum of Ξ (say obtained as the sum of all monomials in Ξ having degree at most n), then $Q(\Xi_n)(X_1, \dots, X_N) = \partial^*(\Xi_n(X_1, \dots, X_N))$. Moreover, as $n \to \infty$, we have that $\Xi_n(X_1, \dots, X_N) \to \Xi(X_1, \dots, X_N)$ in L^2 and also $Q(\Xi_n)(X_1, \dots, X_N) \to Q(\Xi)(X_1, \dots, X_N)$ in L^2

(this can be seen by observing first that the coefficients of $Q_n(\Xi)$ converge to the coefficients of $Q(\Xi)$ and then approximating $Q(\Xi)$ and $Q(\Xi_n)$ by their partial sums).

Since ∂^* is closed, the claimed equality follows. □

2.1.3. *Existence of solutions.* Recall that a process $X_1^{(t)}, \dots, X_N^{(t)} \in (M, \tau)$ is called stationary if its law does not depend on t; that is, for any polynomial f in N noncommuting variables, $\tau\big(f(X_1^{(t)}, \dots, X_N^{(t)})\big)$ is constant.

Lemma 7. *Let $X_1^{(0)}, \dots, X_N^{(0)}$ be an N-tuple of noncommutative random variables, $R_0 > \max_j \|X_j^{(0)}\|$ and $R > R_0$. Let $\xi_j \in \mathcal{F}(R)$, $\Psi = (\Psi_{ij}) \in M_{N\times N}(\mathcal{F}'(R))$ so that $\Psi_{ij}(Z_1, \dots, Z_N)^* = \Psi_{ji}(Z_1, \dots, Z_N)$ for any self-adjoint $Z_1, \dots, Z_N$.*

Consider the free stochastic differential equation

$$dX_i(t) = \Psi(X_1(t), \dots, X_N(t))\#(dS_t^{(1)}, \dots, dS_t^{(N)}) - \tfrac{1}{2}\xi_i\big(X_1(t), \dots, X_N(t)\big)\,dt \tag{2-1}$$

with the initial condition $X_j(0) = X_j^{(0)}$, $j = 1, \dots, n$. Here $dS_t^{(1)}, \dots, dS_t^{(N)}$ is free Brownian motion, and for $Q_{kl} = \sum a_i^{kl} \otimes b_i^{kl} \in M \hat{\otimes} M$, and $Q = (Q_{kl}) \in M_{N\times N}(M \hat{\otimes} M)$, we write

$$Q\#(W_1, \dots, W_N) = \Bigg(\sum_{ki} a_i^{1k} W_k b_i^{1k}, \dots, \sum a_i^{Nk} W b_i^{Nk}\Bigg).$$

Let $A = W^*(X_1^{(0)}, \dots, X_N^{(0)})$ and let $\partial_j : L^2(A) \to L^2(A \bar{\otimes} A)$ be derivations densely defined on polynomials in $X_1^{(0)}, \dots, X_N^{(0)}$ and determined by

$$\partial_j(X_i^{(0)}) = \Xi_{ji}(X_1^{(0)}, \dots, X_N^{(0)}).$$

Assume that for all j, $\partial_i X_j^{(0)} \in$ domain ∂_i^* and that

$$\xi_j(X_1^{(0)}, \dots, X_N^{(0)}) = \sum_i \partial_i^* \partial_i(X_j^{(0)}).$$

Then there exists a $t_0 > 0$ and a stationary solution $X_j(t)$, $0 \le t < t_0$. This stationary solution satisfies $X_j(t) \in W^*\big(X_1, \dots, X_N, \{S_j(s) : 0 \le s \le t\}_{j=1}^N\big)$.

We note that in view of Lemma 6, we may instead assume that $1 \otimes 1 \in$ domain ∂_j^* and

$$\partial_j^*(1 \otimes 1) = \zeta_j(X_1^{(0)}, \dots, X_N^{(0)})$$

for some $\zeta_1, \dots, \zeta_N \in \mathcal{F}(R)$, since this assumption guarantees the existence of $\xi_j \in \mathcal{F}(R)$ satisfying the hypothesis of Lemma 7.

Proof. We note that, because Ψ and ξ are analytic, they are Lipschitz in their arguments.

Thus it follows from the standard Picard argument (see [Biane and Speicher 1998]) that a solution (with given initial conditions) exists, at least for all values of t lying in some small interval $[0, t_0)$, $t_0 > 0$. Now choose t_0 so that $\|X_j(t)\|_\infty \le R_0 < R$ for all $0 \le t < t_0$ (this is possible, since the solution to the SDE is locally norm-bounded).

Next, we note that if we adopt the notations of Lemma 4 and define for $f \in \mathscr{F}(R)$

$$\mathscr{L}^{(\tau)}(f) = \sum_{ijk}(1 \otimes \tau)(\Psi_{jk}\#_{in}(\Psi_{ki}\#_{out}(\mathscr{D}_{ij}f))) - \tfrac{1}{2}\sum_j \xi_j \mathscr{D}_j f,$$

then we have that $\mathscr{L}^{(\tau_t)} f \in \mathscr{F}(R)$ (here τ_t refers to the trace on $\mathbb{C}\langle X_1(t), \dots, X_n(t)\rangle$ obtained by restricting the trace from the von Neumann algebra containing the process X_t for small values of t, that is, $\tau_t(P) = \tau(P(X_1(t), \dots, X_n(t))))$. Ito calculus shows that for any $f \in \mathscr{F}(R)$,

$$\frac{d}{dt}\tau(f(X_1(t), \dots, X_N(t)))\Big|_{t=s} = \tau_s((\mathscr{L}^{(\tau_s)} f)(X_1(s), \dots, X_N(s))).$$

In particular, replacing f with $\mathscr{L}^{(\tau_t)} f$ and iterating gives us the equality

$$\frac{d^n}{dt^n}\tau(f(X_1(t), \dots, X_N(t)))\Big|_{t=s} = \tau_s(((\mathscr{L}^{(\tau_s)})^n f)(X_1(s), \dots, X_N(s))).$$

Since $\xi_j(X_1(0), \dots, X_n(0)) = \sum_i \partial_i^* \partial_i(X_j(0))$,

$$\tau\big(\mathscr{L}^{(\tau_0)}\big(f(X_1(0), \dots, X_N(0))\big)\big) = 0$$

for any $f \in \mathscr{F}(R)$. Applying this to f replaced with $\mathscr{L}^{(\tau_0)} f$ and iterating allows us to conclude that

$$\frac{d^n}{dt^n}\tau(f(X_1(t), \dots, X_N(t)))\Big|_{t=0} = 0, \qquad n \geq 1.$$

Let as before

$$\hat{\mathscr{L}}\phi(z) = \sum_{ijk}\phi_{\Psi_{ik}}(z, NR_0)\phi_{\Psi_{jk}}(NR_0, z)\phi''(z) + \tfrac{1}{2}\sum_j \phi_{\xi_j}(z)\phi'(z) = \alpha(z)^2\phi''(z) + \beta(z)\phi'(z),$$

where $\beta(z)$ is complex-valued function and $\alpha(z)$ is a complex vector-valued analytic function, both defined on $\{z : |z| < NR\}$. Moreover, α and β are real for $z \in \mathbb{R}$.

Consider the partial differential equation

$$\partial_t\phi(x, t) = \hat{\mathscr{L}}\phi(x, t), \qquad \phi(x, 0) = \phi_f(x), \quad x \in \mathbb{R}.$$

The solution $\phi(x, t)$ will be real-analytic in time (at least for small values of t), because the equation is elliptic. By Lemma 5, we conclude that

$$\Big|\partial_t^n \tau(f(X_1(t), \dots, X_N(t)))\big|_{t=s}\Big| \leq \Big|(\hat{\mathscr{L}}^n\phi)(NR_0, s)\Big| = \Big|\partial_t^n\phi(NR_0, t)\big|_{t=s}\Big|.$$

Hence, since all derivatives of $\tau(f(X_1(t)), \dots, f(X_N(t)))$ vanish at zero,

$$\begin{aligned}\big|\tau(f(X_1(s), \dots, X_N(s))) - \tau(f(X_1(0), \dots, X_N(0)))\big| \\ &= \left|\int_0^s \cdots \int_0^s \partial_t^n \tau(f(X_1(t), \dots, f(X_N(t))))\Big|_{t=r}(dr)^n\right| \\ &\leq C\int_0^s \cdots \int_0^s (\partial_t^n\phi(NR_0, t))\Big|_{t=r} \leq \phi(NR_0, s) - P_n(NR_0, s),\end{aligned}$$

where P_n is the n-th Taylor polynomial of ϕ at zero. Since ϕ is real-analytic in s, the right-hand side of the equation goes to zero as $n \to \infty$ for s in some interval including zero. Thus the function $s \to \tau(f(X_1(s), \dots, X_N(s)))$ is constant and so the solution is stationary. $\square$

We note that once the Equation (2-1) has a stationary solution on a small interval $[0, t_0)$, then it of course has a stationary solution for all time (since the same lemma applied to $X_{t_0/2}$ guarantees existence of the solution for up to $3t_0/2$ and so on). However, we will not need this here.

3. Otto–Villani type estimates

The main result of this section is an estimate on the noncommutative Biane–Voiculescu–Wasserstein distance between the law of an N-tuple of variables $X = X_1, \dots, X_N$ and the law of the N-tuple $X + \sqrt{t}\, Q\#S$, where $S = (S_1, \dots, S_N)$ is a free semicircular family, $Q \in M_{N\times N}\big(L^2(W^*(X_1, \dots, X_N)^{\otimes 2})\big)$ is a matrix, and for $Q_{ij} = \sum_k A_{ij}^{(k)} \otimes B_{ij}^{(k)}$, we denote by $Q\#S$ the N-tuple $(Y_1, \dots, Y_N)$ with

$$Y_i = \sum_j \sum_k A_{ij}^{(k)} S_j B_{ij}^{(k)}.$$

The sum defining Y_i is L^2 convergent; in fact, the L^2 norm of Y_i is the same as the L^2 norm of the element

$$\sum_j \sum_k A_{ij}^{(k)} \otimes B_{ij}^{(k)}.$$

The estimate on Wasserstein distance (Proposition 8) is obtained under the assumptions that a certain derivation, defined by $\partial(X_i) = (Q_{i1}, \dots, Q_{iN}) \in \big(L^2(W^*(X_1, \dots, X_N)^{\otimes 2})\big)^N$ is closable and satisfies certain further analyticity conditions (see below for more precise statements). Under such assumptions, the estimate states that

$$d_W(X, X + \sqrt{t}\, Q\#S) \le Ct.$$

The main use of this estimate will be to give a lower bound for the microstates free entropy dimension of $X_1, \dots, X_N$ (see Section 5).

3.1. *An Otto–Villani type estimate on Wasserstein distance via free SDEs.*

Proposition 8. *Let $\Xi \in M_{N\times N}(\mathcal{F}'(R))$, $M = W^*(X_1, \dots, X_N)$ and let $\partial_j : L^2(M) \to L^2(M \mathbin{\bar{\otimes}} M)$ be derivations densely defined on polynomials in $X_1, \dots, X_N$ and determined by*

$$\partial_j(X_i) = \Xi_{ji}(X_1, \dots, X_N).$$

Assume that for all j, $1 \otimes 1 \in \operatorname{domain} \partial_i^*$ and that there exist $\zeta_1, \dots, \zeta_N \in \mathcal{F}(R)$ so that

$$\zeta_j(X_1, \dots, X_N) = \partial_j(1 \otimes 1), \quad j = 1, 2, \dots, N.$$

Then there exists a II_1 factor $\mathcal{M} \cong M * L(\mathbb{F}_\infty)$ and a $t_0 > 0$ so that for all $0 \le t < t_0$ there exists an embedding $\alpha_t : M = W^*(X_1, \dots, X_N) \to \mathcal{M}$ and a free $(0, 1)$-semicircular family $S_1, \dots, S_N \in \mathcal{M}$, free from M and satisfying the inequality

$$\big\|\alpha_t(X_j) - (X_j + \sqrt{t}\,\Xi(X_1, \dots, X_N)\#S)\big\|_2 \le Ct, \tag{3-1}$$

where C is a fixed constant. Furthermore, $\alpha_t(X_j) \in W^*(X_1, \dots, X_N, S_1, \dots, S_N, \{S'_j\}_{j=1}^{\infty})$, where $\{S'_j\}_{j=1}^{\infty}$ are a free semicircular family, free from $(X_1, \dots, X_N, S_1, \dots, S_N)$.

If A can be embedded into R^{ω}, so can $\mathcal{M}$.

In particular, the noncommutative Wasserstein distance of Biane–Voiculescu satisfies

$$d_W\big((X_j)_{j=1}^N, \ (X_j + \sqrt{t}\,\Xi(X_1, \dots, X_N)\#S)_{j=1}^N \big) \le Ct.$$

Proof. By Lemma 6, we can find $\xi_1, \dots, \xi_N \in \mathcal{F}(R)$ so that $\xi_j(X_1, \dots, X_N) = \partial^*\partial(X_j)$.

Let $\mathcal{M} = W^*\big(X_1, \dots, X_N, \{S_1(s), \dots, S_N(s) : 0 \le s \le t\}\big)$, where $S_j(t)$ is a free semicircular Brownian motion. Let $X_j(t)$ be a stationary solution to the SDE (2-1) (see Lemma 7). The map that takes a polynomial in $X_1, \dots, X_N$ to a polynomial in $X_1(t), \dots, X_N(t)$ preserves traces and so extends to an embedding $\alpha_t : M \to \mathcal{M}$. By the free Burkholder–Gundy inequality [Biane and Speicher 1998], it follows that for $0 \le t < t_0 < 1$

$$\|X_j(t) - X_j(0)\| \le C_1\sqrt{t} + C_2 t \le C_3\sqrt{t},$$

where $C_1 = \sup_{t<t_0} \|\Xi(X_1(t), \dots, X_N(t)\| < \infty$, $C_2 = \max_j \sup_{t<t_0} \|\xi_j(X_1, \dots, X_N(t)\|$.

Furthermore,

$$\begin{aligned} X_j(t) - X_j(0) &= \int_0^t \Xi(X_1(s), \dots, X_N(s))\#dS_j(s) - \int_0^t \xi_j(X_1(s), \dots, X_{Nn}(s))\,ds \\ &= \int_0^t \Xi(X_1(0), \dots, X_N(0))\#dS_j(s) \\ &\quad - \int_0^t \big[\Xi(X_1(0), \dots X_N(0)) - \Xi(X_1(s), \dots, X_N(s))\big]\#dS_j(s) \\ &\quad - \int_0^t \xi_j(X_1(s), \dots, X_N(s))\,ds. \end{aligned}$$

By the Lipschitz property of the coefficients of the SDE (2-1), we see that

$$\|\Xi(X_1(s), \dots, X_N(s)) - \Xi(X_1(0), \dots, X_N(0))\| \le K \max_j \|X_j(s) - X_j(0)\| \le K'\sqrt{s}.$$

Combining this with the estimate $\|\xi_j(X_1(t), \dots, X_N(t))\| < K''$ we may apply the free Burkholder–Gundy inequality to deduce that

$$\|X_j(t) - (X_j(0) + \Xi(X_1(0), \dots, X_N(0))\#S_j(t))\| \le \left|\int_0^t (K'\sqrt{s})^2 ds\right|^{1/2} + \left\|\int_0^t K''\,ds\right\| \le Ct.$$

Thus it is enough to notice that $\|\cdot\|_2 \le \|\cdot\|$ and to take $S_j = \frac{1}{\sqrt{t}} S_j(t)$, which is a $(0, 1)$ semicircular element.

If M is R^{ω}-embeddable, we may choose $\mathcal{M}$ to be R^{ω}-embeddable as well, since it can be chosen to be a free product of M and a free group factor.

Finally, note that $X_j(t) \in W^*\big(X_1, \dots, X_N, \{S_j(s) : 0 \le s \le t\}_{j=1}^N\big)$ by construction. But the algebra $W^*\big(\{S_j(s) : 0 \le s \le t\}\big)$ can be viewed as the algebra of the Free Gaussian functor applied to the space $L^2[0, 1]$, in such a way that $S_j(s) = S([0, s])$. Then $W^*\big(\{S_j(s) : 0 \le s \le t\}\big) \subset W^*\big(S_1, \dots, S_N, \{S'_k\}_{k\in I(j)}\big)$, where $\{S'_k : k \in I(j)\}$ are free semicircular elements corresponding to the completion of the singleton set $\{t^{-1/2}\chi_{[0,t]}\}$ to an orthonomal basis of $L^2[0, 1]$.

The estimate for the Wasserstein distance now follows if we note that the law of $(\alpha_t(X_j))_{j=1}^N$ is the same as that of $(X_j)_{j=1}^N$; thus $(X_j(t))_{j=1}^N \cup (X_j + \sqrt{t}\,\Xi\#S)_{j=1}^N$ is a particular $2N$-tuple with marginal distributions the same as those of $(X_j)_{j=1}^N$ and $(X_j + \sqrt{t}\,\Xi\#S)_{j=1}^N$, so that the estimate (3-1) becomes an estimate on the Wasserstein distance. □

Remark 9. Although we do not need this in the rest of the paper, we note that the estimate in Proposition 8 also holds in the operator norm.

We should mention that an estimate similar to the one in Proposition 8 was obtained by Biane and Voiculescu [2001] in the case $N = 1$ under the much less restrictive assumptions that $\Xi = 1 \otimes 1$ and $1 \otimes 1 \in \operatorname{domain} \partial^*$ (that is, the free Fisher information $\Phi^*(X)$ is finite). Setting $\Xi_{ij} = \delta_{ij} 1 \otimes 1$ we have proved an analog of their estimate (in the N-variable case), but under the very restrictive assumption that the conjugate variables $\partial^*(\Xi)$ are analytic functions in $X_1, \dots, X_N$. The main technical difficulty in removing this restriction lies in the question of existence of a stationary solution to (2-1) in the case of very general drifts ξ.

4. Applications to q-semicircular families

4.1. *Estimates on certain operators related to q-semicircular families.*

4.1.1. *Background on q-semicircular elements.* Let $H_{\mathbb{R}}$ be a finite-dimensional real Hilbert space, H its complexification $H = H_{\mathbb{R}} \otimes_{\mathbb{R}} \mathbb{C}$, and let $F_q(H)$ be the q-deformed Fock space on H [Bożejko and Speicher 1991]. Thus

$$F_q(H) = \mathbb{C}\Omega \oplus \bigoplus_{n \geq 1} H^{\otimes n},$$

with the inner product given by

$$\langle \xi_1 \otimes \cdots \otimes \xi_n, \zeta_1 \otimes \cdots \otimes \zeta_m \rangle = \delta_{n=m} \sum_{\pi \in S_n} q^{i(\pi)} \prod_{j=1}^{n} \langle \xi_j, \zeta_{\pi(j)} \rangle,$$

where $i(\pi) = \#\{(i, j) : i < j \text{ and } \pi(i) > \pi(j)\}$.

We write HS for the space of Hilbert–Schmidt operators on $F_q(H)$. We denote by $\Xi \in HS$ the operator

$$\Xi = \sum q^n P_n,$$

where P_n is the orthogonal projection onto the subspace $H^{\otimes n} \subset F_q(H)$.

For $h \in H$, let $l(h) : F_q(h) \to F_q(H)$ be the creation operator, $l(h)(\xi_1 \otimes \cdots \otimes \xi_n) = h \otimes \xi_1 \otimes \cdots \otimes \xi_n$, and for $h \in H_{\mathbb{R}}$, let $s(h) = l(h) + l(h)^*$. We denote by M the von Neumann algebra $W^*(s(h) : h \in H_{\mathbb{R}})$. It is known [Ricard 2005; Śniady 2004] that M is a II_1 factor and that $\tau = \langle \cdot \Omega, \Omega \rangle$ is a faithful tracial state on M. Moreover, $F_q(H) = L^2(M, \tau)$ and $HS = L^2(M, \tau) \mathbin{\bar{\otimes}} L^2(M, \tau)$.

Fix an orthonormal basis $\{h_i\}_{i=1}^N \subset H_{\mathbb{R}}$ and let $X_i = s(h_i)$. Thus $M = W^*(X_1, \dots, X_N)$, $N = \dim H_{\mathbb{R}}$.

Lemma 10 [Shlyakhtenko 2004]. *For $j = 1, \dots, N$, let $\partial_j : \mathbb{C}[X_1, \dots, X_N] \to HS$ be the derivation given by $\partial_j(X_i) = \delta_{i=j}\Xi$. Let $\partial : \mathbb{C}[X_1, \dots, X_N] \to HS^N$ be given by $\partial = \partial_1 \oplus \cdots \oplus \partial_N$ and regard ∂ as an unbounded operator densely defined on $L^2(M)$. Then:*

(i) *∂ is closable.*

(ii) *If we denote by* Z_j *the vector* $0 \oplus \cdots \oplus P_\Omega \oplus \cdots \oplus 0 \in HS^N$ *(nonzero entry in* j*-th place,* P_Ω *is the orthogonal projection onto* $\mathbb{C}\Omega \in F_q(H)$*), then* Z_j *is in the domain of* ∂^* *and* $\partial^*(Z_j) = h_j$.

As a consequence of (ii), if we let ∂ be as in the lemma, we have $\xi_j = \partial^*(Z_j) \in \mathbb{C}[X_1, \cdots, X_N] \subset \mathcal{F}(R)$ for any R.

4.1.2. Ξ *as an analytic function of* $X_1, \ldots, X_n$. We now claim that for small values of q, the element $\Xi \in L^2(M)^{\otimes 2}$ defined in Lemma 10 can be thought of as an analytic function of the variables $X_1, \ldots, X_N$. Recall that $h_i \in H$ is a fixed orthonormal basis and $X_j = s(h_j)$, $j = 1, \ldots, N$ thus form a q-semicircular family.

Lemma 11. *Let* $W_{i_1,\ldots,i_n}$ *be noncommutative polynomials so that*

$$W_{i_1,\ldots,i_n}(X_1, \ldots, X_N)\Omega = h_{i_1} \otimes \cdots \otimes h_{i_n}.$$

Then the degree of $W_{i_1,\ldots,i_n}$ *is* n, *and the maximal absolute value* $c_k^{(n)}$ *of a coefficient of a monomial* $X_{j_1} \cdots X_{j_k}$, $k \le n$, *in* $W_{i_1,\ldots,i_n}$ *satisfies*

$$c_k^{(n)} \le 2^{n-k}\left(\frac{1}{1-|q|}\right)^{n-k}.$$

Furthermore, $\|W_{i_1,\ldots,i_n}\|^2_{L^2(M)} \le 2^n(1-|q|)^{-n}$.

Proof. Clearly, $c_n^{(n)} = 1$. Moreover (compare [Effros and Popa 2003])

$$W_{i_1,\ldots,i_n} = X_{i_1} W_{i_2,\ldots,i_n} - \sum_{j \ge 2} q^{j-2} \delta_{i_1 = i_j} W_{i_2,\ldots,\hat{i}_j,\ldots,i_n}$$

(where $\hat{\ }$ denotes omission). So the degree of $W_{i_1,\ldots,i_n}$ is n and the coefficient c_n of a monomial of degree k in $W_{i_1,\ldots,i_n}$ is at most the sum of a coefficient of a degree $k-1$ monomial in $W_{i_2,\ldots,i_n}$ and $\sum_{j\ge 2} q^{j-2}|k_j|$, where k_j is a coefficient of a degree k monomial in $W_{i_2,\ldots,\hat{i}_j,\ldots,i_n}$. By induction, we see that

$$\begin{aligned}
c_k^{(n)} &\le c_{k-1}^{(n-1)} + \sum_{j\ge 2}^{n} |q|^{j-2} c_k^{(n-2)} \\
&\le 2^{n-k-2}\left(\frac{1}{1-|q|}\right)^{n-k} + 2^{n-k-2}\left(\frac{1}{1-|q|}\right)^{n-k-2} \sum_{j\ge 0} |q|^j \\
&= 2^{n-k-2}\left[\left(\frac{1}{1-|q|}\right)^{n-k} + \left(\frac{1}{1-|q|}\right)^{n-k-2} \frac{1}{1-|q|}\right] \\
&\le 2^{n-k-2} \cdot 2\left(\frac{1}{1-|q|}\right)^{n-k} \le 2^{n-k}\left(\frac{1}{1-|q|}\right)^{n-k}.
\end{aligned}$$

as claimed.

The upper estimate on $\|W_{i_1,\ldots,i_n}\|^2_{L^2(M)}$ follows in a similar way. □

Lemma 12. *Let $\{\xi_k : k \in K\}$ be a finite set of vectors in an inner product space V. Let Γ be the matrix $\Gamma_{k,l} = \langle \xi_k, \xi_l \rangle$. Assume that Γ is invertible and let $B = \Gamma^{-1/2}$. Then the vectors*

$$\zeta_l = \sum_k B_{k,l} \xi_k$$

form an orthonormal basis for the span of $\{\xi_k : k \in K\}$. Moreover, if λ denotes the smallest eigenvalue of Γ, then $|B_{k,l}| \leq \lambda^{-1/2}$ for each k, l.

Proof. We have, using the fact that B is symmetric and $B\Gamma B = I$: $\langle \zeta_l, \zeta_{l'} \rangle = \langle \sum_{k,k'} B_{k,l} \xi_k, B_{k',l'} \xi_{k'} \rangle = \sum_{k,k'} B_{k,l} B_{k',l'} \Gamma_{k,k'} = (B\Gamma B)_{l,l'} = \delta_{l=l'}$. □

Lemma 13. *There exist noncommutative polynomials $p_{i_1,\dots,i_n}$ in $X_1, \dots, X_N$ so that the vectors*

$$\{p_{i_1,\dots,i_n}(X_1, \dots, X_n)\Omega\}_{i_1,\dots,i_n=1}^N$$

are orthonormal and have the same span as $\{W_{i_1,\dots,i_n}\}_{i_1,\dots,i_n=1}^N$.

Moreover, these can be chosen so that $p_{i_1,\dots,i_n}$ is a polynomial of degree at most n and the coefficient of each degree k monomial in p is at most $(1-2|q|)^{-n/2}(2N)^n(1-|q|)^k 2^{-k}$.

Proof. Consider the inner product matrix

$$\Gamma_n = [\langle W_{i_1,\dots,i_n}, W_{j_1,\dots,j_n} \rangle]_{i_1,\dots,i_n,\ j_1,\dots,j_n=1}^N.$$

Dykema and Nica [1993, Lemma 3.1] proved that one has the following recursive formula for Γ_n. Consider an N^n-dimensional vector space W with orthonormal basis $e_{i_1,\dots,i_n}$, $i_1, \dots, i_n \in \{1, \dots, N\}$, and consider the unitary representation π_n of the symmetric group S_n given by $\sigma \cdot e_{i_1,\dots,i_n} = e_{i_{\sigma(1)},\dots,i_{\sigma(n)}}$. Denote by $(1 \to j)$ the action (via π_n) of the permutation that sends 1 to j, k to $k-1$ for $2 \leq k \leq j$, and l to l for $l > j$ on W. Let $M_n = \sum_{j=1}^n q^{j-1}(1 \to j) \in \operatorname{End}(W)$. Then Γ_1 is the identity $N \times N$ matrix, and

$$\Gamma_n = (1 \otimes \Gamma_{n-1}) M_n,$$

where $1 \otimes \Gamma_n$ acts on the basis element $e_{j_1,\dots,j_n}$ by sending it to $\sum_{k_2,\dots,k_n} (\Gamma_{n-1})_{j_2,\dots,j_n,\ k_2,\dots,k_n} e_{j_1,k_2,\dots,k_n}$ and Γ acts on the basis elements by sending $e_{j_1,\dots,j_n}$ to $\sum_{k_1,\dots,k_n} (\Gamma_n)_{j_1,\dots,j_n,\ k_1,\dots,k_n} e_{k_1,\dots,k_n}$. They then proceeded to prove that the operator M_n is invertible and derive a bound for its inverse in the course of proving [Dykema and Nica 1993, Lemma 4.1]. Combining this bound and the recursive formula for Γ_n yields the following lower estimate for the smallest eigenvalue of Γ_n:

$$\begin{aligned} c_n &= \left(\frac{1}{1-|q|} \prod_{k=1}^{\infty} \frac{1-|q|^k}{1+|q|^k} \right)^n = \left(\frac{1}{1-|q|} \sum_{k=-\infty}^{\infty} (-1)^k |q|^{k^2} \right)^n \\ &\geq \left(\frac{1}{1-|q|} \left(1 - \sum_{k \geq 0} |q|^{k^2} \right) \right)^n \geq \frac{1}{(1-|q|)^{n/2}} \left(1 - \sum_{k \geq 1} |q|^k \right)^n \\ &\geq \left(\frac{1}{1-|q|} \left(1 - \frac{|q|}{1-|q|} \right) \right)^n = \left(\frac{1-2|q|}{(1-|q|)^2} \right)^n. \end{aligned}$$

Thus if we set $B = \Gamma_n^{-1/2}$, then all entries of B are bounded from above by $c_n^{-1/2}$. Thus if we apply the previous lemma with $K = \{1, \dots, N\}^n$ to the vectors $\xi_{i_1,\dots,i_n} = W_{i_1,\dots,i_n}\Omega$, we obtain that the vectors

$$\zeta_i = \sum_{j \in K} B_{j,i}\xi_j, \quad i \in K$$

form an orthonormal basis for the subspace of the Fock space spanned by tensors of length n.

Now for $i = (i_1, \dots, i_n) \in K$, let

$$p_i(X_1, \dots, X_N) = \sum_{j \in K} B_{j,i} W_j(X_1, \dots, X_N).$$

Then $\zeta_i = p_i(X_1, \dots, X_N)\Omega$ are orthonormal and (because the vacuum vector is separating), the polynomials $\{p_i : i \in K\}$ have the same span as $\{W_i : i \in K\}$.

Furthermore, if a is the coefficient of a degree k monomial r in p_i, then a is a sum of at most N^n terms, each of the form (the coefficient of r in W_j)$B_{j,i}$. Using Lemma 11, we therefore obtain the estimate

$$|a| \le N^n c_n^{-1/2} 2^{n-k}(1-|q|)^{-(n-k)} = \left(\frac{2N}{(1-2|q|)^{1/2}}\right)^n 2^{-k}(1-|q|)^k. \qquad \square$$

We now use the terminology of Section 2.1.1 in dealing with noncommutative power series.

Let $R_0 = 2(1-|q|)^{-1} \ge 2(1-q)^{-1} \ge \|X_j\|$. Then if $\alpha > 1$, $p = p_{i_1,\dots,i_n}$ is as in Lemma 13, and ϕ_p is as in Section 2.1.1, then the coefficient of z^k, $k \le n$ in ϕ_p is bounded by

$$\left(\frac{2N}{(1-2|q|)^{1/2}}\right)^n R_0^{-k} \le \left(\frac{2N\alpha}{(1-2|q|)^{1/2}}\right)^n (\alpha N R_0)^{-k}.$$

In particular for any $\rho < \alpha R_0$,

$$\|p_{i_1,\dots,i_n}\|_\rho \le \left(\frac{2N\alpha}{(1-2|q|)^{1/2}}\right)^n \sum_{k=0}^{n} (\alpha N R_0)^{-k} N^k \rho^k \le \left(\frac{2N\alpha}{(1-2|q|)^{1/2}}\right)^n \frac{1}{1-\rho/(\alpha R_0)}.$$

Lemma 14. *Let q be such that $|q| < (4N^3+2)^{-1}$. Then*:

(a) *The formula*

$$\Xi(Y_1, \dots, Y_N) = \sum_n q^n \sum_{i_1,\dots,i_n} p_{i_1,\dots,i_n}(Y_1, \dots, Y_N) \otimes p_{i_1,\dots,i_n}(Y_1, \dots, Y_N)$$

defines a noncommutative power series with values in $\mathbb{C}\langle Y_1, \dots, Y_N\rangle^{\otimes 2}$ with radius of convergence strictly bigger than the norm of a q-semicircular element, $\|X_j\| \le 2(1-q)^{-1}$.

(b) *If $X_1, \dots, X_N$ are q-semicircular elements and Ξ is as in Lemma 10, then $\Xi = \Xi(X_1, \dots, X_N)$ (convergence in Hilbert–Schmidt norm, identifying HS with $L^2(M) \bar{\otimes} L^2(M)$).*

Proof. Clearly,

$$\|p_{i_1,\dots,i_n} \otimes p_{i_1,\dots,i_n}\|_\rho \le \|p_{i_1,\dots,i_n}\|_\rho^2 \le \left(\frac{2N\alpha}{(1-2|q|)^{1/2}}\right)^{2n} \frac{1}{(1-\rho/(\alpha R_0))^2} = K_\rho \left(\frac{4N^2\alpha^2}{1-2|q|}\right)^n$$

for any $\rho < \alpha R_0$, where $R_0 = 2(1-|q|)^{-1} \ge \|X_j\|$.

Thus
$$\|\Xi\|_\rho \le K_\rho \sum_n \left(\frac{4N^2\alpha^2}{1-2|q|}\right)^n |q|^n N^n \le K_\rho \sum_n \left(\frac{4N^3\alpha|q|}{1-2|q|}\right)^n,$$
which is finite as long as $\rho < \alpha R_0$ and the fraction in the sum in the right is less than 1. Thus as long as $4N^3|q| < 1-2|q|$, that is, $|q| < (4N^3+2)^{-1}$, we can choose some $\alpha > 1$ so that the series defining Ξ has a radius of convergence of at least $\alpha R_0 > \|X_j\|$.

For part (b), we note that because $\|\cdot\|_{L^2(M)} \le \|\cdot\|_M$ and because of the definition of the projective tensor product, we see that
$$\|\cdot\|_{HS} \le \|\cdot\|_{M\hat{\otimes}M}$$
on $M\hat{\otimes}M$. Thus convergence in the projective norm on $M\hat{\otimes}M$ guarantees convergence in Hilbert–Schmidt norm. Furthermore, by definition of orthogonal projection onto a space,
$$\Xi = \sum q^n P_n,$$
where $P_n = \sum_{i_1,\dots,i_n} p_{i_1,\dots,i_n} \otimes p_{i_1,\dots,i_n} = \Xi^{(n)}(X_1,\dots,X_N)$ are the partial sums of $\Xi(X_1,\dots,X_N)$ (here we again identify HS and $L^2 \bar{\otimes} L^2$). Hence $\Xi = \Xi(X_1,\dots,X_N)$. □

5. An estimate on free entropy dimension

We now show how an estimate of the form (1-3) can be used to prove a lower bound for the free entropy dimension δ_0.

Recall from [Voiculescu 1996; 1994] that if $X_1,\dots,X_n \in (M,\tau)$ is an n-tuple of self-adjoint elements, then the set of microstates $\Gamma_R(X_1,\dots,X_n; l,k,\varepsilon)$ is defined by
$$\begin{aligned}\Gamma_R(X_1,\dots,X_n; l,k,\varepsilon) = \Big\{(x_1,\dots,x_n) \in (M_{k\times k}^{sa})^n : \|x_j\| < R \text{ and}\\ \big|\tau(p(X_1,\dots,X_n)) - (1/k)\operatorname{Tr}(p(x_1,\dots,x_n))\big| < \varepsilon\\ \text{for any monomial } p \text{ of degree } \le l\Big\}.\end{aligned}$$
If R is omitted, the value $R=\infty$ is understood. The free entropy is defined by
$$\chi(X_1,\dots,X_n) = \sup_R \inf_{l,\varepsilon} \limsup_{k\to\infty} \left(\frac{1}{k^2}\log \operatorname{Vol}(\Gamma_R(X_1,\dots,X_n; l,k,\varepsilon)) + \frac{n}{2}\log k\right).$$

The set of microstates for $X_1,\dots,X_n$ in the presence of $Y_1,\dots,Y_m$ is defined by
$$\begin{aligned}\Gamma_R(X_1,\dots,X_n : Y_1,\dots,Y_m; l,k,\varepsilon) = \big\{(x_1,\dots,x_n) : \exists(y_1,\dots,y_m)\\ \text{s.t. } (x_1,\dots,x_n,y_1,\dots,y_m) \in \Gamma_R(X_1,\dots,X_n,Y_1,\dots,Y_m; l,k,\varepsilon)\big\}.\end{aligned}$$
The corresponding free entropy in the presence is then defined as by
$$\begin{aligned}&\chi(X_1,\dots,X_n : Y_1,\dots,Y_m)\\ &\quad= \sup_R \inf_{l,\varepsilon} \limsup_{k\to\infty} \left(\frac{1}{k^2}\log\operatorname{Vol}(\Gamma_R(X_1,\dots,X_n : Y_1,\dots,Y_m; l,k,\varepsilon)) + \frac{n}{2}\log k\right).\end{aligned}$$

The $\sup_R$ is attained [Belinschi and Bercovici 2003]; in fact, once $R > \max_{i,j}\{\|X_i\|, \|Y_j\|\}$, we have

$$\chi(X_1, \dots, X_n : Y_1, \dots, Y_m) = \chi_R(X_1, \dots, X_n : Y_1, \dots, Y_m).$$

Finally, the free entropy dimension δ_0 is defined by

$$\delta_0(X_1, \dots, X_n) = n + \limsup_{t\to 0} \frac{\chi(X_1 + \sqrt{t}S_1, \dots, X_n + \sqrt{t}S_n : S_1, \dots, S_n)}{|\log t|},$$

where $S_1, \dots, S_n$ are a free semicircular family, free from $X_1, \dots, X_n$. Equivalently [Jung 2003a] one sets

$$\mathbb{K}_\delta(X_1, \dots, X_n) = \inf_{\varepsilon,l} \limsup_{k\to\infty} \frac{1}{k^2} \log K_\delta(\Gamma_\infty(X_1, \dots, X_n; k, l, \varepsilon)),$$

where $K_\delta(X)$ is the covering number of a set X (the minimal number of δ-balls needed to cover X). Then

$$\delta_0(X_1, \dots, X_n) = \limsup_{t\to 0} \frac{\mathbb{K}_t(X_1, \dots, X_n)}{|\log t|}.$$

Lemma 15. *Assume that $X_1, \dots, X_n \in (M, \tau)$, $T_{jk} \in W^*(X_1, \dots, X_n) \mathbin{\bar\otimes} W^*(X_1, \dots, X_n)^{op}$ are given. Set $S_j^T = \sum_k T_{jk}\#S_k$. Let $\eta = \dim_{M\bar\otimes M^o}\Big(\overline{\operatorname{span} MS_1^T M + \cdots + MS_n^T M}^{L^2(M\bar\otimes M^o)}\Big)$.*

Then there exists a constant K depending only on T so that for all $R > 0$, $\alpha > 0$, $t > 0$, there are $\varepsilon' > 0$, $l' > 0$, and $k' > 0$ so that for all $0 < \varepsilon < \varepsilon'$, $k > k'$, and $l > l'$, and any $(x_1, \dots, x_n) \in \Gamma(X_1, \dots, X_n; k, l, \varepsilon)$ the set

$$\begin{aligned}\Gamma_R(tS_1^{I-T}, \dots, tS_n^{I-T} \mid (x_1, \dots, x_n) : S_1, \dots, S_n; k, l, \varepsilon) = \\ \big\{(y_1, \dots, y_n) : \exists (s_1, \dots, s_n) \text{ s.t. } (y_1, \dots, y_n, x_1, \dots, x_n, s_1, \dots, s_n) \in \\ \Gamma_R(tS_1^{I-T}, \dots, tS_n^{I-T}, X_1, \dots, X_n, S_1, \dots, S_n; k, l, \varepsilon)\big\}\end{aligned}$$

can be covered by $(K/t)^{(n-\eta+\alpha)k^2}$ balls of radius t^2.

Proof. By considering the diffeomorphism of $(M^{sa}_{k\times k})^n$ given by $(a_1, \dots, a_n) \mapsto ((1/t)a_1, \dots, (1/t)a_n)$, we may reduce the statement to showing that the set

$$\Gamma_R(S_1^{I-T}, \dots, S_n^{I-T} \mid (x_1, \dots, x_n) : S_1, \dots S_n; k, l, \varepsilon)$$

can be covered by $(C/t)^{(n-\eta+\alpha)k^2}$ balls of radius t.

Note that η is the Murray–von Neumann dimension over $M \mathbin{\bar\otimes} M^o$ of the image of the map

$$(\zeta_1, \dots, \zeta_n) \mapsto (\zeta_1^T, \dots, \zeta_n^T),$$

where $\zeta_j \in L^2(M) \mathbin{\bar\otimes} L^2(M)$, $M = W^*(X_1, \dots, X_n)$. Thus if we view T as a matrix in $M_{n\times n}(M \mathbin{\bar\otimes} M^o)$, then $\tau \otimes \tau \otimes \operatorname{Tr}(E_{\{0\}}((I-T)^*(I-T))) = \eta$ (here E_X denotes the spectral projection corresponding to the set $X \subset \mathbb{R}$).

Fix $\alpha > 0$.

Then there exists $Q \in M_{n\times n}(\mathbb{C}[X_1, \dots, X_n]^{\otimes 2})$ depending only on t so that $\|Q_{ij} - (I-T)_{ij}\|_2 < t/(2n)$ (here we view Q as a matrix whose entries are noncommutative functions in n indeterminates; the entries of Q are in the space $\mathscr{F}'(\infty)$ in the notation of Section 2.1.1).

Set $S_j^Q = \sum_k Q_{jk} \# S_k$. Then

$$\left\| S_j^{Q(X_1,\dots,X_n)} - S_j^{I-T} \right\| < \frac{t}{2}.$$

In particular, $\left\| S_j^{Q(X_1,\dots,X_n)} - S_j^{I-T} \right\|_2 < t/2$. We may moreover choose Q (again, depending only on t) so that

$$\tau \otimes \tau \otimes \mathrm{Tr}(E_{[0,t/2[}(Q^*Q)^{1/2}(X_1,\dots,X_n)) \geq \tau \otimes \tau \otimes \mathrm{Tr}(E_{\{0\}}(I-T)^*(I-T)) = \eta - \tfrac{1}{2}\alpha.$$

Thus for l sufficiently large and $\varepsilon > 0$ sufficiently small, we have that if

$$(y_1,\dots,y_n) \in \Gamma_R(S_1^{I-T},\dots,S_n^{I-T} | (x_1,\dots,x_n) : S_1,\dots,S_n; k,l,\varepsilon),$$

then there exist $s_1,\dots,s_n$ such that

$$(s_1,\dots,s_n,x_1,\dots,x_n) \in \Gamma_R(S_1,\dots,S_n,X_1,\dots,X_n; k,l,\varepsilon)$$

and

$$\| s_j^{Q(x_1,\dots,x_n)} - y_j \|_2 < t.$$

By approximating the characteristic function $\chi_{[0,t/2]}$ with polynomials on the interval $[0, \| Q(x_1,\dots,x_n) \|]$ (which is compact, since $\|x_j\| < R$), we may moreover assume that l is large enough and ε is small enough that

$$\frac{1}{k^2} \mathrm{Tr} \otimes \mathrm{Tr} \otimes \mathrm{Tr}\big(E_{[0,t/2]}(Q^*Q)^{1/2}(x_1,\dots,x_n)\big) \geq \eta - \alpha.$$

Denote by ϕ the map

$$(s_1,\dots,s_n) \mapsto \big(s_1^{Q(x_1,\dots,x_n)},\dots,s_n^{Q(x_1,\dots,x_n)}\big).$$

Let $R_1 = \max_j \|S_j^{I-T}\|_2 + 1$. Assume that $\varepsilon < 1$. Then $\phi : (M_{k\times k}^{sa})^n \to (M_{k\times k}^{sa})^n$ is a linear map, and since $\|s_j\|_2^2 \leq 1 + \varepsilon < 2$, we have the inclusion

$$\Gamma_R(S_1^{I-T},\dots,S_n^{I-T} | (x_1,\dots,x_n) : S_1,\dots,S_n; k,l,\varepsilon) \subset N_t(\phi(B(2)) \cap B(R_1)),$$

where $B(R)$ the a ball of radius R in $(M_{k\times k}^{sa})^n$ (endowed with the L^2 norm) and N_t denotes a t-neighborhood.

The matrix of ϕ is precisely the matrix $Q(x_1,\dots,x_n) \in M_{n\times n}(M_{k\times k})^{\otimes 2}$.

Let β be such that $\beta n k^2$ eigenvalues of $(\phi^*\phi)^{1/2}$ are less than R_0. Then the t-covering number of $\phi(B(2)) \cap B(R_1)$ is at most

$$\left(\frac{R_1}{t}\right)^{(1-\beta)nk^2} \left(\frac{2R_0}{t}\right)^{\beta n k^2}.$$

Let $R_0 = t/2$, so $\beta = (\eta - \alpha)/n$. We conclude that the t-covering number of

$$\Gamma_R\big(S_1^{I-T},\dots,S_n^{I-T} \bigm| (x_1,\dots,x_n) : S_1,\dots,S_n; k,l,\varepsilon\big)$$

is at most $(K/t)^{(n-\eta+\alpha)k^2}$, for some constant K depending only on R_1, which itself depends only on T.

$\square$

Theorem 16. *Assume that $X_1, \dots, X_n \in (M, \tau)$, $S_1, \dots, S_n$, $\{S_j : j \in J\}$ is a free semicircular family, free from M, $T_{jk} \in W^*(X_1, \dots, X_n) \bar{\otimes} W^*(X_1, \dots, X_n)^{op}$ are given, and that for each $t > 0$ there exist $Y_j^{(t)} \in W^*(X_1, \dots, X_n, S_1, \dots, S_n, \{S'_j\}_{j \in J})$ so that:*

- *the joint law of $(Y_1^{(t)}, \dots, Y_n^{(t)})$ is the same as that of $(X_1, \dots, X_n)$,*
- *if we set $S_j^T = \sum_k T_{jk} \# S_k$ and $Z_j^{(t)} = X_j + t S_j^T$, then for some $t_0 > 0$ and some constant $C < \infty$ independent of t, we have $\|Z_j^{(t)} - Y_j^{(t)}\|_2 \le C t^2$ for all $t < t_0$.*

Let $M = W^*(X_1, \dots, X_n)$ and let

$$\eta = \dim_{M \bar{\otimes} M^o} \left(\overline{\operatorname{span} M S_1^T M + \dots + M S_n^T M}^{L^2} \right).$$

Assume finally that $W^*(X_1, \dots, X_n)$ embeds into R^ω. Then $\delta_0(X_1, \dots, X_n) \ge \eta$.

Proof. Let $T : (M \bar{\otimes} M^o)^n \to (M \bar{\otimes} M^o)^n$ be the linear map given by

$$T(Y_1, \dots, Y_n) = \left(\sum_k T_{1k} \# Y_k, \dots, \sum_k T_{nk} \# Y_k \right)$$

(here, as before, we identify $(M \bar{\otimes} M^o)^n$ with the linear span of $M S_1 M + \dots + M S_n M$ via the map $(T_1, \dots, T_n) \mapsto (S^{T_1}, \dots, S^{T_n})$). Then η is the Murray–von Neumann dimension of the image of T, and consequently

$$\eta = n - \dim_{M \bar{\otimes} M^o} \ker T.$$

Let t be fixed.

Since $Y_j^{(t)}$ can be approximated by noncommutative polynomials in $X_1, \dots, X_n$, $S_1, \dots, S_n$ and $\{S'_j : j \in J\}$, for any k_0, ε_0, l_0 sufficiently large we may find $k > k_0$, $l > l_0$, $\varepsilon < \varepsilon_0$ and $J_0 \subset J$ finite so that whenever

$$(z_1, \dots, z_n) \in \Gamma_R\big(X_1 + t S_1^T, \dots, X_n + t S_n^T : S_1, \dots, S_n, \{S'_j\}_{j \in J_0}; k, l, \varepsilon\big),$$

there exists

$$(y_1, \dots, y_n) \in \Gamma_R(X_1, \dots, X_n; k, l_0, \varepsilon_0)$$

so that

$$\|y_j - z_j\|_2 \le C t^2. \tag{5-1}$$

For a set $X \subset (M_{k \times k}^{sa})^n$ we'll write K_r for its covering number by balls of radius r.

Consider a covering of $\Gamma_R(X_1 + t S_1, \dots, X_n + t S_n : S, \dots, S_n, \{S'_j\}_{j \in J_0}; k, l, \varepsilon)$ by balls of radius $(C + 2) t^2$ constructed as follows.

First, let $(B_\alpha)_{\alpha \in I}$ be a covering of $\Gamma_R\big(X_1 + t S_1^T, \dots, X_n + t S_n^T : S_1, \dots, S_n, \{S'_j\}_{j \in J_0}; k, l_0, \varepsilon_0\big)$ by balls of radius $(C + 1) t^2$. Because of (5-1), we may assume that

$$|I| \le K_{t^2}(\Gamma_R(X_1, \dots, X_n; k, l, \varepsilon)).$$

Next, for each $\alpha \in I$, let $(x_1^{(\alpha)}, \dots, x_n^{(\alpha)}) \in B_\alpha$ be the center of B_α. Consider a covering $(C_\beta^{(\alpha)} : \beta \in J_\alpha)$ of $\Gamma_R\big(t S_1^{I-T}, \dots, t S_n^{I-T} \mid (x_1^{(\alpha)}, \dots, x_n^{(\alpha)}) : S_1, \dots, S_n; k, l, \varepsilon\big)$ by balls of radius t^2. By Lemma 15, this covering can be chosen to contain $|J_\alpha| \le (K/t)^{n-\eta'}$ balls, for any $\eta' < \eta$. Thus the sets

$$(B_\alpha + C_\beta^{(\alpha)} : \alpha \in I, \beta \in J_\alpha),$$

each of which is contained in a ball of radius at most $(C+2)t^2$, cover the set

$$\Gamma_R\big(X_1 + tS_1, \ldots, X_n + tS_n : S_1, \ldots, S_n; k, l_0, \varepsilon_0\big).$$

The cardinality of this covering is at most

$$f(t^2, k) \le |I| \cdot \sup_\alpha |J_\alpha| \le K_{t^2}(\Gamma_R(X_1, \ldots, X_n; k, l, \varepsilon) \cdot (Kt)^{\eta'-n}.$$

It follows that if we denote by $V(R, d)$ the volume of a ball of radius R in $\mathbb{R}^d$, we find that

$$\mathrm{Vol}(\Gamma_R(X_1 + tS_1, \ldots, X_n + tS_n : S_1, \ldots, S_n, \{S'_j\}_{j \in J_0})) \le f(t^2, k) \cdot V((C+2)t^2, nk^2),$$

so that if we denote by $\mathbb{K}_{t^2}(X_1, \ldots, X_n)$ the expression

$$\inf_{\varepsilon, l} \limsup_{k \to \infty} \frac{1}{k^2} \log K_{t^2}(\Gamma(X_1, \ldots, X_n; k, l, \varepsilon))$$

and set $C' = \log(C+2)$, we obtain the inequality

$$\begin{aligned}
\inf_{\epsilon, l} \limsup_{k \to \infty} \frac{1}{k^2} \log \mathrm{Vol}\Gamma_R(X_1 + tS_1, \ldots, X_n + tS_n : S_1, \ldots, S_n, \{S'_j\}_{j \in J_0}; k, l, \varepsilon) \\
\le \limsup\nolimits_{k \to \infty} \log f(t^2, k) + 2n \log t + \log(C+2) \\
\le \mathbb{K}_{t^2}(X_1, \ldots, X_n) + (\eta' - n) \log Kt + 2n \log t + C' \\
= \mathbb{K}_{t^2}(X_1, \ldots, X_n) + (\eta' + n) \log t + (\eta' - n) \log K + C'.
\end{aligned}$$

By the freeness of $\{S'_j\}_{j \in J}$ and $\{S_1, \ldots, S_n, X_1, \ldots, X_n\}$, the $\limsup$ on the right-hand side remains the same if we take $J_0 = \varnothing$. Thus

$$\chi_R(X_1 + tS_1, \ldots, X_n + tS_n : S_1, \ldots, S_n) \le \mathbb{K}_{t^2}(X_1, \ldots, X_n) + (\eta' + n) \log t + C''.$$

If we divide both sides by $|\log t|$ and add n to both sides of the resulting inequality, we obtain

$$\begin{aligned}
n + \frac{\chi_R(X_1 + tS_1, \ldots, X_n + tS_n : S_1, \ldots, S_n)}{|\log t|} &\le \frac{\mathbb{K}_{t^2}(X_1, \ldots, X_n)}{|\log t|} + (\eta' + n)\frac{\log t}{|\log t|} + n \\
&= 2\frac{\mathbb{K}_{t^2}(X_1, \ldots, X_n)}{|\log t^2|} + (\eta' + n)\frac{\log t}{|\log t|} + n.
\end{aligned}$$

Taking $\sup_R$ and $\limsup_{t \to 0}$ and noticing that $\log t < 0$ for $t < 1$, we get the inequality

$$\delta_0(X_1, \ldots, X_n) \le 2\delta_0(X_1, \ldots, X_n) - (\eta + n) + n = 2\delta_0(X_1, \ldots, X_n) - \eta'.$$

Solving this inequality for $\delta_0(X_1, \ldots, X_n)$ gives finally

$$\delta_0(X_1, \ldots, X_n) \ge \eta'.$$

Since $\eta' < \eta$ was arbitrary, we conclude that $\delta_0(X_1, \ldots, X_n) \ge \eta$ as claimed. $\square$

Corollary 17. *Let (A, τ) be a finitely-generated algebra with a positive trace τ and generators $X_1, \dots, X_n$, and let $\mathrm{Der}_a(A; A \otimes A)$ denote the space of derivations from A to $L^2(A \otimes A, \tau \otimes \tau)$ which are L^2 closable and so that for some $\Xi_j \in \mathscr{F}'(R)$, $\xi \in \mathscr{F}(R)$, $R > \max_j \|X_j\|$, $\partial^*(1 \otimes 1) = \xi(X_1, \dots, X_n)$ and $\partial(X_j) = \Xi_j(X_1, \dots, X_n)$. Consider the A,A-bimodule*

$$V = \{(\delta(X_1), \dots, \delta(X_n)) : \delta \in \mathrm{Der}_a(A; A \otimes A)\} \subset L^2(A \otimes A, \tau \otimes \tau)^n.$$

Assume finally that $M = W^*(A, \tau) \subset R^\omega$. Then

$$\delta_0(X_1, \dots, X_n) \geq \dim_{M \bar{\otimes} M^o} \overline{V}^{L^2(A \otimes A, \tau \otimes \tau)^n}.$$

Proof. Let $P : L^2(A \otimes A, \tau \otimes \tau)^n \to \overline{V}$ be the orthogonal projection, and set $v_j = P(0, \dots, 1 \otimes 1, \dots, 0)$ with $1 \otimes 1$ in the j-th position. Let $v_j^{(k)} = (v_{1j}^{(k)}, \dots, v_{nj}^{(k)}) \in L^2(A \otimes A)^n$ be vectors approximating v_j and having the property that the derivations defined by $\delta(X_j) = v_{ij}^{(k)}$ lie in Der_a. Then

$$\eta_k = \dim_{M \bar{\otimes} M^o} \overline{\operatorname{span} A v_1^{(k)} A + \cdots + A v_n^{(k)} A} \to \dim_{M \bar{\otimes} M^o} \overline{V}$$

as $k \to \infty$. Now for each k, the derivations $\delta_j : A \to L^2(A \otimes A)$ so that $\delta_j(X_i) = v_{ij}^{(k)}$ belong to Der_a. Applying Lemma 6 and Proposition 8 to $T_{ij} = v_{ij}^{(k)}$ and combining the conclusion with Theorem 16 gives

$$\delta_0(X_1, \dots, X_n) \geq \eta_k.$$

Taking $k \to \infty$ we get

$$\delta_0(X_1, \dots, X_n) \geq \dim_{M \bar{\otimes} M^o} V,$$

as claimed. □

Corollary 18. *For a fixed N, and all $|q| < (4N^3 + 2)^{-1}$, the q-semicircular family $X_1, \dots, X_N$ satisfies*

$$\delta_0(X_1, \dots, X_N) > 1 \text{ and } \delta_0(X_1, \dots, X_N) \geq N\left(1 - \frac{q^2 N}{1 - q^2 N}\right).$$

In particular, $M = W^*(X_1, \dots, X_N)$ has no Cartan subalgebra. Moreover, for any abelian subalgebra $\mathscr{A} \subset M$, $L^2(M)$, as an $\mathscr{A}$,$\mathscr{A}$-bimodule, contains a copy of the coarse correspondence.

Proof. Let ∂_i be a derivation as in Lemma 10; thus $\partial_i(X_j) = \delta_{i=j}\Xi$, as defined in Lemma 10. Then for $|q| < (4N^3 + 2)^{-1}$, Lemma 14 shows that $\partial_i \in \mathrm{Der}_a$. Then Theorem 16 implies that

$$\delta_0(X_1, \dots, X_N) \geq \dim_{M \bar{\otimes} M^o} \overline{\sum M \Xi_i M},$$

$M = W^*(X_1, \dots, X_n)$. It is known [Shlyakhtenko 2004] that for $q^2 < 1/N$ (which is the case if we make the assumptions about q as in the hypothesis of the corollary), this dimension is strictly bigger than 1 and is no less than $N(1 - q^2 N(1 - q^2 N)^{-1})$.

The facts about M follow from [Voiculescu 1996]. □

For $N = 2$, $(4N^3 + 2)^{-1} = 1/34$. Thus the theorem applies for $0 \leq q \leq 1/34 = 0.029\dots$. Our estimate also shows that $\delta_0(X_1, \dots, X_N) \to N$ as $q \to 0$.

Corollary 19. *Let Γ be a discrete group generated by $g_1, \ldots, g_n$, and let $V \subset C^1(\Gamma, \ell^2\Gamma)$ be the subset consisting of cocycles valued in $\mathbb{C}\Gamma \subset \ell^2\Gamma$. If the group von Neumann algebra of Γ can be embedded into the ultrapower of the hyperfinite II_1 factor (for example, if the group is sofic), then*

$$\delta_0(\mathbb{C}\Gamma) \geq \dim_{L(\Gamma)} \overline{V}.$$

Proof. Any such cocycle gives rise to a derivation into $\mathbb{C}\Gamma^{\otimes 2}$ by the formula

$$\partial(\gamma) = c(\gamma) \otimes \gamma^{-1}.$$

Then $\partial^*\partial(\gamma) = \|c(\gamma)\|_2^2 \gamma \in \mathbb{C}\Gamma$. Moreover, the bimodule generated by values of these derivations on any generators of $\mathbb{C}\Gamma$ has the same dimension over $L(\Gamma)\bar{\otimes} L(\Gamma)$ as $\dim_{L(\Gamma)} \bar{V}$. $\square$

For certain R^ω embeddable groups (for example, free groups, amenable groups, residually finite groups with property T, more generally embeddable groups with first L^2 Betti number $\beta_1^{(2)} = 0$, as well as groups obtained from these by taking amalgamated free products over finite subgroups and passing to finite index subgroups and finite extensions), V is actually dense in the set of ℓ^2 1-cocycles. Indeed, this is the case if all ℓ^2 derivations are inner (that is, $\beta_1^{(2)}(\Gamma) = 0$). Moreover, it follows from the Mayer–Vietoris exact sequence that amalgamated free products over finite subgroups retain the property that V is dense in the space of ℓ^2 cocycles. Moreover, this property is also clearly preserved by passing to finite-index subgroups and finite extensions. So it follows that for such groups Γ, $\delta_0(\Gamma) = \beta_1^{(2)}(\Gamma) + \beta_0^{(2)}(\Gamma) - 1$ (compare [Brown et al. 2008]).

It is likely that the techniques of the present paper could be extended to prove the following:

Conjecture 20. *Let Γ be a group generated by $g_1, \ldots, g_n$ and assume that $L(\Gamma)$ can be embedded into R^ω. Let $V \subset \ell^2(\Gamma)^n$ be the subspace $\{(c(g_1), \ldots, c(g_n)) : c : \Gamma \to \ell^2(\Gamma)$ 1-cocycle$\}$. Let $P_V : \ell^2(\Gamma)^n \to V$ be the orthogonal projection, so that $P_V \in M_{n\times n}(R(\Gamma))$, where $R(\Gamma)$ is the von Neumann algebra generated by the right regular representation of the group.*

Let $\mathcal{A} \subset R(\Gamma)$ be the closure of $\mathbb{C}\Gamma \subset R(\Gamma)$ under holomorphic functional calculus, and let $P_a \in \mathcal{A}$ be any projection so that $P_a \leq P_V$. Then $\delta_0(\Gamma) \geq \mathrm{Tr}_{M_{n\times n}} \otimes \tau_{R(\Gamma)}(P_a)$.

Note that with the notations of the Conjecture, $\mathrm{Tr}_{M_{n\times n}} \otimes \tau_{R(\Gamma)}(P_V) = \beta_1^{(2)}(\Gamma) - \beta_0^{(2)}(\Gamma) + 1 = \delta^*(\Gamma)$.

It should be noted that the restriction on the values of the cocycles ($\mathbb{C}\Gamma$ rather than $\ell^2\Gamma$) comes from the difficulty in the extending the results of Proposition 8 to the case of nonanalytic Ξ (though the term $\partial^*\partial(\gamma)$ continues to be a polynomial even in the case that the cocycle is valued in $\ell^2(\Gamma)$ rather than $\mathbb{C}\Gamma$).

Appendix: Otto–Villani type estimates via exponentiation of derivations

Let $M = W^*(X_1, \ldots, X_N)$, where $X_1, \ldots, X_N$ are self-adjoint.

Let us denote by ζ_j the vector $(0, \ldots, 0, 1 \otimes 1, 0, \ldots, 0) \in [L^2(M, \tau)^{\otimes 2}]^N$ (the only nonzero entry is in the j-th position). One can realize a free semicircular family of cardinality N on the space

$$H = L^2(M, \tau) \oplus \bigoplus_{k\geq 1} \left[(L^2(M, \tau) \bar{\otimes} L^2(M, \tau))^{\oplus N}\right]^{\otimes_M k}.$$

using creation and annihilation operators $S_i = L_i + L_i^*$, where

$$L_i \xi = \zeta_i \otimes_M \xi.$$

Then for $\zeta \in W^*(M) \bar{\otimes} W^*(M)$, the notation S_ζ makes sense with $S_{\zeta_i} = S_i$, $aS_\zeta b + b^* S_\zeta a^* = S_{a\zeta b + b^*\zeta a^*}$ and $\|S_\zeta\|_2 = \|\zeta\|_2$.

Let $A = \mathrm{Alg}(X_1, \dots, X_N)$. For $a, b \in A \otimes A$ and $j = 1, \dots, N$ write

$$(a \otimes b)\#S = aSb.$$

Proposition 21. *Let $\partial : A \to V_0 = [W^*(M, \tau) \bar{\otimes} W^*(M, \tau)]^{\oplus N} \subset V = [L^2(M, \tau) \bar{\otimes} L^2(M, \tau)]^{\oplus N}$ be a derivation. We assume that for each j, ζ_j is in the domain of $\partial^* : V \to L^2(M, \tau)$ and that $\partial(a^*) = (\partial(a))^*$, where $* : L^2(M) \bar{\otimes} L^2(M)$ is the involution $(a \otimes b)^* = b^* \otimes a^*$. Let $S_1, S_2, \dots$ be semicircular elements, free from M.*

Assume that $\partial(A) \subset (A \otimes A)^{\oplus N}$ and also that $\partial^*(1 \otimes 1) \in A$.

Then there exists a one-parameter group α_t of automorphisms of $M * W^*(S_1, \dots, S_N) \cong M * L(\mathbb{F}_N)$ so that $A \cup \{S_j : 1 \le j \le N\}$ are analytic for α_t and

$$\frac{d}{dt}\alpha_t(a)\Big|_{t=0} = S_{\partial(a)} \quad \text{for all } a \in A, \qquad \frac{d}{dt}\alpha_t(S_j)\Big|_{t=0} = -\partial^*(\zeta_j) \quad \text{for } j = 1, 2, \dots.$$

In particular,

$$\alpha_t(a) \cdot 1 = \Big(a - \frac{t^2}{2}\partial^*(\partial(a))\Big) \;+\; t\partial(a) \;-\; \frac{t^2}{2}(1 \otimes \partial + \partial \otimes 1)(\partial(a)) \in H.$$

Proof. Let B be the algebra generated by A and $S_1, \dots, S_N$ in $\mathcal{M} = W^*(A, \tau) * L(\mathbb{F}_N)$.

Let $P_j : V \to L^2(A \otimes A)$ be the j-th coordinate projection, and let $\partial_j : A \to A \otimes A$ be given by $\partial_j = P_j \circ \partial$.

Let $V_1, \dots, V_N \in B$ be given by

$$V_j = \sum_k \partial_k(X_j)\#S_k = S_{\partial(X_j)}, \qquad j = 1, \dots, N.$$

Let $V_{N+1}, \dots, V_{2N} \in B$ be given by

$$V_{N+k} = -\partial_k^*(1 \otimes 1) = -\partial^*(\zeta_k), \qquad k = 1, \dots, N.$$

Then $(V_1, \dots, V_{2N}) \in B \subset L^2(B, \tau)$ is a noncommutative vector field in the sense of [Voiculescu 2002a]. It is routine to check that this vector field is orthogonal to the cyclic gradient space.

We now use [Voiculescu 2002a] to deduce that there exists a one-parameter automorphism group α_t of $\mathcal{M} = W^*(B, \tau)$ such that

$$\frac{d}{dt}\alpha_t(X_j)\Big|_{t=0} = V_j \quad \text{for } j = 1, \dots, N, \qquad \frac{d}{dt}\alpha_t(S_k)\Big|_{t=0} = V_{N+k} \quad \text{for } k = 1, \dots, N,$$

and moreover that all elements in B are analytic for α_t. In particular, we see that

$$\begin{aligned} \frac{d}{dt}\alpha_t(X_j)\Big|_{t=0} &= S_{\partial(X_j)}, \\ \frac{d^2}{dt^2}\alpha_t(X_j)\Big|_{t=0} \cdot 1 = \delta(S_{\partial(X_j)}) &= -\partial^*(\partial(X_j)) - (1 \otimes \partial + \partial \otimes 1)(\partial(X_j)), \end{aligned}$$

as claimed. $\square$

Example 22. We give three examples in which the automorphisms α_t can be explicitly constructed. The first is the case that $X_1, \dots, X_N$ is a free semicircular system and $\partial(X_j) = (0, \dots, 1 \otimes 1, \dots 0)$ (that is, $\partial = \oplus \partial_j$, where ∂_j are the difference quotient derivations of [Voiculescu 1998]). In this case, the automorphism α_t is given by

$$\alpha_t(X_j) = (\cos t)X_j + (\sin t)S_j, \qquad \alpha_t(S_j) = -(\sin t)X_j + (\cos t)S_j.$$

Another case is that of a general N-tuple $X_1, \dots, X_N$ and ∂ an inner derivation given by $\partial(X) = [X, T]$, for $[T_j]_{j=1}^N = [-T_j^*]_{j=1}^N \in [M \bar{\otimes} M^o]^N$. Put $z = \sum T_j \# S_j$. Then α_t is an inner automorphism given by $\alpha_t(Y) = \exp(izt) Y \exp(-izt)$. Lastly, assume that $M = M_1 * M_2$ and the derivations ∂_j are determined by $\partial_j|_{M_1} = 0$, $\partial_j|_{M_2}(x) = [x, T_j]$ for some $T_j \in M \bar{\otimes} M^o$. Then again put $z = \sum T_j \# S_j$. The automorphism α_t is then given by $\alpha_t(Y) = \exp(izt) Y \exp(-izt)$. In particular, $\alpha_t|_{M_1} = \mathrm{id}$ and $\alpha_t|_{M_2}$ is given by conjugation by unitaries $\exp(izt)$ which are free from M_1 and M_2.

Proposition 21 can be used to give another proof to the Otto–Villani type estimates (Proposition 8) in the case of polynomial coefficients, using the following standard lemma:

Lemma 23. *Let $\beta_t : (M, \tau) \to (M, \tau)$ be a one-parameter group of automorphisms so that $\tau \circ \beta_t = \tau$. Let $X \in M$ be an element so that $t \mapsto \beta_t(X)$ is twice-differentiable. Finally let*

$$Z = \frac{d}{dt} \beta_t(X)\Big|_{t=0}, \qquad \xi = \frac{d^2}{dt^2} \beta_t(X)\Big|_{t=0}.$$

Then, for all t,

$$\|\beta_t(X) - (X + tZ)\|_2 \le \frac{t^2}{2} \|\xi\|_2.$$

Corollary 24. *Assume that $X_1, \dots, X_N \in A$ and $\partial_1, \dots \partial_N : A \to A \otimes A$ are derivations, so that $\partial_j^*(1 \otimes 1) \in A$. Then we have the following estimate for the free Wasserstein distance*:

$$d_W\Big((X_1, \dots, X_N), (X_1 + \sqrt{t} \sum_k \partial_k(X_1) \# S_k, \dots, X_N + \sqrt{t} \sum_k \partial_k(X_N) \# S_k)\Big) \le Ct,$$

where C is the constant given by

$$C = \tfrac{1}{2} \Bigg(\sum_j \|\partial^* \partial(X_j)\|^2_{L^2(A)} + \|(1 \otimes \partial + \partial \otimes 1)(\partial(X_j))\|^2_{[L^2(A) \otimes L^2(A) \otimes L^2(A)]^{N^2}} \Bigg)^{1/2},$$

where $\partial : A \to [L^2(A) \otimes L^2(A)]^N$ is the derivation $\partial = \partial_1 \oplus \cdots \oplus \partial_N$.

In the specific case of the difference quotient derivations determined by $\partial_k(X_j) = \delta_{kj} 1 \otimes 1$, we have

$$d_W\Big((X_1, \dots, X_N), (X_1 + \sqrt{t} S_1, \dots, X_N + \sqrt{t} S_N)\Big) \le \frac{t}{2} \Phi^*(X_1, \dots, X_N)^{1/2}.$$

Proof. Let α_t be the one-parameter group of automorphisms as in Proposition 21. We note that

$$\Bigg(\sum_j \|\alpha_{\sqrt{t}}(X_j) - (X_j + \sqrt{t} \sum_k \partial_k(X_j) \# S_k)\|_2^2 \Bigg)^{1/2} \le Ct$$

in view of Lemma 23 and the expression for $\alpha_t''(X_j)$. On the other hand, $(\alpha_{\sqrt{t}}(X_1),\dots,\alpha_{\sqrt{t}}(X_N))$ has the same law as $(X_1,\dots,X_N)$, since $\alpha_{\sqrt{t}}$ is a $*$-homomorphism. It follows that

$$\begin{aligned} d_W\big(X_1,\dots,X_N,(X_1+\sqrt{t}\sum_k \partial_k(X_1)\#S_k,\dots,X_N+\sqrt{t}\sum_k \partial_k(X_N)\#S_k)\big) \\ = d_W\big(\alpha_{\sqrt{t}}(X_1),\dots,\alpha_{\sqrt{t}}(X_N),(X_1+\sqrt{t}\sum_k \partial_k(X_1)\#S_k,\dots,X_N+\sqrt{t}\sum_k \partial_k(X_N)\#S_k)\big) \le Ct. \end{aligned}$$

In the case of the difference quotient derivations, we have:

$$\sum_k \partial_k(X_j)\#S_k = S_j, \quad (1\otimes\partial+\partial\otimes 1)(\partial(X_j)) = (1\otimes\partial+\partial\otimes 1)(1\otimes 1)=0, \quad \partial^*\partial(X_j)=\partial_j^*(1\otimes 1).$$

Thus

$$C=\tfrac{1}{2}\Big(\sum_j \|\partial_j^*(1\otimes 1)\|_2^2\Big)^{1/2} = \tfrac{1}{2}\Phi^*(X_1,\dots,X_N)^{1/2}$$

as claimed. □

Acknowledgment

The author is grateful to A. Guionnet for suggesting the idea of using stationary solutions to free SDEs as an alternative form of "exponentiating" derivations, and to (patiently) explaining to him about stochastic differential equations. The author also thanks D.-V. Voiculescu for a number of comments and suggestions. I am also very grateful to the referees for a number of suggestions and improvements to the paper.

References

[Belinschi and Bercovici 2003] S. T. Belinschi and H. Bercovici, "A property of free entropy", *Pacific J. Math.* **211**:1 (2003), 35–40. MR 2004i:46097 Zbl 1058.46043

[Biane and Speicher 1998] P. Biane and R. Speicher, "Stochastic calculus with respect to free Brownian motion and analysis on Wigner space", *Probab. Theory Related Fields* **112**:3 (1998), 373–409. MR 99i:60108 Zbl 0919.60056

[Biane and Voiculescu 2001] P. Biane and D. Voiculescu, "A free probability analogue of the Wasserstein metric on the trace-state space", *Geom. Funct. Anal.* **11**:6 (2001), 1125–1138. MR 2003d:46087 Zbl 1020.46020

[Bożejko and Speicher 1991] M. Bożejko and R. Speicher, "An example of a generalized Brownian motion", *Comm. Math. Phys.* **137**:3 (1991), 519–531. MR 92m:46096 Zbl 0722.60033

[Brown et al. 2008] N. P. Brown, K. J. Dykema, and K. Jung, "Free entropy dimension in amalgamated free products", *Proc. Lond. Math. Soc.* (3) **97**:2 (2008), 339–367. MR MR2439665 Zbl 1158.46045 arXiv math.OA/0609080

[Connes and Shlyakhtenko 2005] A. Connes and D. Shlyakhtenko, "L^2-homology for von Neumann algebras", *J. Reine Angew. Math.* **586** (2005), 125–168. MR 2007b:46104 Zbl 1083.46034 arXiv math.OA/0309343

[Dykema and Nica 1993] K. Dykema and A. Nica, "On the Fock representation of the q-commutation relations", *J. Reine Angew. Math.* **440** (1993), 201–212. MR 94e:46117 Zbl 0767.46038

[Effros and Popa 2003] E. G. Effros and M. Popa, "Feynman diagrams and Wick products associated with q-Fock space", *Proc. Natl. Acad. Sci. USA* **100**:15 (2003), 8629–8633. MR 2004h:81099 Zbl 1064.81045

[Ge 1998] L. Ge, "Applications of free entropy to finite von Neumann algebras, II", *Ann. of Math.* (2) **147**:1 (1998), 143–157. MR 99c:46068 Zbl 0924.46050

[Ge and Shen 2002] L. Ge and J. Shen, "On free entropy dimension of finite von Neumann algebras", *Geom. Funct. Anal.* **12**:3 (2002), 546–566. MR 2003f:46109 Zbl 1040.46043

[Jung 2003a] K. Jung, "A free entropy dimension lemma", *Pacific J. Math.* **211**:2 (2003), 265–271. MR 2004k:46115 Zbl 1058.46044 arXiv math.OA/0207149

[Jung 2003b] K. Jung, "The free entropy dimension of hyperfinite von Neumann algebras", *Trans. Amer. Math. Soc.* **355**:12 (2003), 5053–5089. MR 2004f:46080 Zbl 1028.46096

[Jung 2007] K. Jung, "Strongly 1-bounded von Neumann algebras", *Geom. Funct. Anal.* **17**:4 (2007), 1180–1200. MR 2009i:46115 Zbl 1146.46034

[Lück 2002] W. Lück, *L^2-invariants: theory and applications to geometry and K-theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics. **44**, Springer, Berlin, 2002. MR 2003m:58033 Zbl 1009.55001

[Mineyev and Shlyakhtenko 2005] I. Mineyev and D. Shlyakhtenko, "Non-microstates free entropy dimension for groups", *Geom. Funct. Anal.* **15**:2 (2005), 476–490. MR 2006e:46073 Zbl 1094.46039

[Otto and Villani 2000] F. Otto and C. Villani, "Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality", *J. Funct. Anal.* **173**:2 (2000), 361–400. MR 2001k:58076 Zbl 0985.58019

[Ozawa 2004] N. Ozawa, "Solid von Neumann algebras", *Acta Math.* **192**:1 (2004), 111–117. MR 2005e:46115 Zbl 1072.46040

[Peterson 2009] J. Peterson, "L^2-rigidity in von Neumann algebras", *Invent. Math.* **175**:2 (2009), 417–433. Zbl 05530522

[Ricard 2005] É. Ricard, "Factoriality of q-Gaussian von Neumann algebras", *Comm. Math. Phys.* **257**:3 (2005), 659–665. MR 2006e:46069 Zbl 1079.81038

[Shlyakhtenko 1998] D. Shlyakhtenko, "Some applications of freeness with amalgamation", *J. Reine Angew. Math.* **500** (1998), 191–212. MR 99j:46079 Zbl 0926.46046

[Shlyakhtenko 2004] D. Shlyakhtenko, "Some estimates for non-microstates free entropy dimension with applications to q-semicircular families", *Int. Math. Res. Not.* **51** (2004), 2757–2772. MR 2006e:46074 Zbl 1075.46055

[Shlyakhtenko 2006] D. Shlyakhtenko, "Remarks on free entropy dimension", pp. 249–257 in *Operator Algebras: The Abel Symposium 2004*, Abel Symp. **1**, Springer, Berlin, 2006. MR 2007i:46065 Zbl 1118.46058 arXiv math.OA/0504062

[Śniady 2004] P. Śniady, "Factoriality of Bożejko–Speicher von Neumann algebras", *Comm. Math. Phys.* **246**:3 (2004), 561–567. MR 2005d:46130 Zbl 1064.46047

[Ştefan 2005] M. B. Ştefan, "Indecomposability of free group factors over nonprime subfactors and abelian subalgebras", *Pacific J. Math.* **219**:2 (2005), 365–390. MR 2009e:46060 Zbl 1098.46050

[Voiculescu 1994] D. Voiculescu, "The analogues of entropy and of Fisher's information measure in free probability theory, II", *Invent. Math.* **118**:3 (1994), 411–440. MR 96a:46117 Zbl 0820.60001

[Voiculescu 1996] D. Voiculescu, "The analogues of entropy and of Fisher's information measure in free probability theory, III: The absence of Cartan subalgebras", *Geom. Funct. Anal.* **6**:1 (1996), 172–199. MR 96m:46119 Zbl 0856.60012

[Voiculescu 1998] D. Voiculescu, "The analogues of entropy and of Fisher's information measure in free probability theory. V. Noncommutative Hilbert transforms", *Invent. Math.* **132**:1 (1998), 189–227. MR 99d:46087 Zbl 0930.46053

[Voiculescu 1999] D. Voiculescu, "The analogues of entropy and of Fisher's information measure in free probability theory, VI: Liberation and mutual free information", *Adv. Math.* **146**:2 (1999), 101–166. MR 2001a:46064 Zbl 0956.46045

[Voiculescu 2002a] D. Voiculescu, "Cyclomorphy", *Int. Math. Res. Not.* **6** (2002), 299–332. MR 2003a:46092 Zbl 1027.46089

[Voiculescu 2002b] D. Voiculescu, "Free entropy", *Bull. London Math. Soc.* **34**:3 (2002), 257–278. MR 2003c:46077 Zbl 1036.46051

Received 10 Jan 2008. Revised 10 Nov 2008. Accepted 24 Mar 2009.

DIMITRI SHLYAKHTENKO: shlyakht@math.ucla.edu
Department of Mathematics, University of California, Los Angeles, Los Angeles, CA 90095, United States

ANALYSIS AND PDE
Vol. 2, No. 2, 2009

HEAT-FLOW MONOTONICITY OF STRICHARTZ NORMS

JONATHAN BENNETT, NEAL BEZ, ANTHONY CARBERY AND DIRK HUNDERTMARK

Our main result is that for $d=1,2$ the classical Strichartz norm $\|e^{is\Delta}f\|_{L^{2+4/d}_{s,x}(\mathbb{R}\times\mathbb{R}^d)}$ associated to the free Schrödinger equation is nondecreasing as the initial datum f evolves under a certain quadratic heat flow.

1. Introduction

For $d\in\mathbb{N}$ let the Fourier transform $\widehat{f}:\mathbb{R}^d\to\mathbb{C}$ of a Lebesgue integrable function f on $\mathbb{R}^d$ be given by

$$\widehat{f}(\xi)=\frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d}e^{-ix\cdot\xi}f(x)\,dx.$$

For each $s\in\mathbb{R}$ the Fourier multiplier operator $e^{is\Delta}$ is defined via the Fourier transform by

$$\widehat{e^{is\Delta}f}(\xi)=e^{-is|\xi|^2}\widehat{f}(\xi)$$

for all f belonging to the Schwartz class $\mathcal{S}(\mathbb{R}^d)$ and $\xi\in\mathbb{R}^d$. Thus for each $f\in\mathcal{S}(\mathbb{R}^d)$ and $x\in\mathbb{R}^d$,

$$e^{is\Delta}f(x)=\frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d}e^{i(x\cdot\xi-s|\xi|^2)}\widehat{f}(\xi)\,d\xi.$$

By an application of the Fourier transform in x it is easily seen that $e^{is\Delta}f(x)$ solves the Schrödinger equation

$$i\partial_s u=-\Delta u, \tag{1-1}$$

with initial datum $u(0,x)=f(x)$. It is well known that the solution operator $e^{is\Delta}$ extends to a bounded operator from $L^2(\mathbb{R}^d)$ to $L^p_sL^q_x(\mathbb{R}\times\mathbb{R}^d)$ if and only if (d,p,q) is Schrödinger-admissible; that is, there exists a finite constant $C_{p,q}$ such that

$$\|e^{is\Delta}f\|_{L^p_sL^q_x(\mathbb{R}\times\mathbb{R}^d)}\le C_{p,q}\|f\|_{L^2(\mathbb{R}^d)} \tag{1-2}$$

if and only if

$$p,q\ge 2,\quad (d,p,q)\ne(2,2,\infty),\ \text{and}\quad \frac{2}{p}+\frac{d}{q}=\frac{d}{2}. \tag{1-3}$$

For $p=q=2+4/d$, this classical inequality is due to Strichartz [1977], who followed arguments of Stein and Tomas (see [Tomas 1975]). For $p\ne q$ the reader is referred to [Keel and Tao 1998] for historical references and a full treatment of (1-2) for suboptimal constants $C_{p,q}$.

MSC2000: 35K05, 35Q40.
Keywords: heat flow, Strichartz estimates, Schrödinger equation.
Bennett and Bez were supported by EPSRC grant EP/E022340/1.

Foschi [2007] and independently Hundertmark and Zharnitsky [2006] showed that in the cases where one can "multiply out" the Strichartz norm

$$\|e^{is\Delta} f\|_{L^p_s L^q_x(\mathbb{R}\times\mathbb{R}^d)}, \tag{1-4}$$

that is, when q is an even integer dividing p, the sharp constants $C_{p,q}$ in the inequalities above are obtained by testing on isotropic centered Gaussians. (These authors considered $p = q$ only.) The main purpose of this paper is to highlight a startling monotonicity property of such Strichartz norms as the function f evolves under a certain quadratic heat flow.

Theorem 1.1. *Let $f \in L^2(\mathbb{R}^d)$. If (d, p, q) is Schrödinger-admissible and q is an even integer which divides p, the quantity*

$$Q_{p,q}(t) := \big\|e^{is\Delta}(e^{t\Delta}|f|^2)^{1/2}\big\|_{L^p_s L^q_x(\mathbb{R}\times\mathbb{R}^d)} \tag{1-5}$$

is nondecreasing for all $t > 0$; that is, $Q_{p,q}$ is nondecreasing in the cases $(1, 6, 6)$, $(1, 8, 4)$, *and* $(2, 4, 4)$.

The heat operator $e^{t\Delta}$ is of course defined to be the Fourier multiplier operator with multiplier $e^{-t|\xi|^2}$, and so

$$e^{t\Delta}|f|^2 = H_t * |f|^2,$$

where the heat kernel $H_t : \mathbb{R}^d \to \mathbb{R}$ is given by

$$H_t(x) = \frac{1}{(4\pi t)^{d/2}}\, e^{-|x|^2/4t}. \tag{1-6}$$

By making an appropriate rescaling one may rephrase the above result in terms of "sliding" Gaussians in the following way. For $f \in L^2(\mathbb{R}^d)$ let $u : (0,\infty)\times\mathbb{R}^d \to \mathbb{R}$ be given by $u(t,x) = H_t * |f|^2(x)$ and $\widetilde{u} : (0,\infty)\times\mathbb{R}^d \to \mathbb{R}$ be given by

$$\widetilde{u}(t,x) = t^{-d}u(t^{-2}, t^{-1}x) = \frac{1}{(4\pi)^{d/2}} \int_{\mathbb{R}^d} e^{-|x-tv|^2/4}|f(v)|^2\,dv.$$

We interpret $\widetilde{u}$ as a superposition of translates of a fixed Gaussian which simultaneously slide to the origin as t tends to zero. By a simple change of variables it follows that

$$Q_{p,q}(t^{-2}) = \big\|e^{is\Delta}(\widetilde{u}(t,\cdot)^{1/2})\big\|_{L^p_s L^q_x(\mathbb{R}\times\mathbb{R}^d)}. \tag{1-7}$$

The reader familiar with the standard wave-packet analysis in the context of Fourier extension estimates may find it more enlightening to interpret Theorem 1.1 via this rescaling.

The claimed monotonicity of $Q_{p,q}$ yields the sharp constant $C_{p,q}$ in (1-2) as a simple corollary. To see this, suppose that the function f is bounded and has compact support. Then, by rudimentary calculations,

$$\lim_{t\to 0} Q_{p,q}(t) = \big\|e^{is\Delta}|f|\big\|_{L^p_s L^q_x(\mathbb{R}\times\mathbb{R}^d)},$$

which, by virtue of the fact that q is an even integer which divides p, is greater than or equal to $\|e^{is\Delta} f\|_{L^p_s L^q_x(\mathbb{R}\times\mathbb{R}^d)}$. Furthermore, because of (1-7) it follows that

$$\lim_{t\to\infty} Q_{p,q}(t) = \big\|e^{is\Delta}(H_1^{1/2})\big\|_{L^p_s L^q_x(\mathbb{R}\times\mathbb{R}^d)}\|f\|_{L^2(\mathbb{R}^d)},$$

where H_1 is the heat kernel at time $t = 1$. Therefore Theorem 1.1 gives the sharp constant $C_{p,q}$ in (1-2) for the triples $(1, 6, 6)$, $(1, 8, 4)$, and $(2, 4, 4)$, and shows that Gaussians are maximisers. In particular, if

$$C_{p,q} := \sup\{\|e^{is\Delta} f\|_{L^p_s L^q_x(\mathbb{R}\times\mathbb{R}^d)} : \|f\|_{L^2(\mathbb{R}^d)} = 1\}$$

then $C_{6,6} = 12^{-1/12}$, $C_{8,4} = 2^{-1/4}$, and $C_{4,4} = 2^{-1/2}$. As we have already noted, $C_{6,6}$ and $C_{4,4}$ were found recently by Foschi [2007] and independently by Hundertmark and Zharnitsky [2006]. In the $(1, 8, 4)$ case, we shall see in the proof of Theorem 1.1 below that the monotonicity (and hence sharp constant) follows easily from the $(2, 4, 4)$ case.

Heat-flow methods have already proved effective in treating certain d-linear analogues of the Strichartz estimate (1-2) [Bennett et al. 2006]. Also intimately related (as we shall see) are the articles [Carlen et al. 2004; Bennett et al. 2008a] in the setting of the multilinear Brascamp–Lieb inequalities.

The proof of Theorem 1.1 is contained in Section 2. We discuss some further results in Section 3. In particular we show that the Strichartz norm is nondecreasing under a certain quadratic Mehler flow and observe that one may relax the quadratic nature of the heat flow in Theorem 1.1 by inserting a mitigating factor which is a power of t. We also consider extensions of Theorem 1.1 to higher dimensions.

2. Proof of Theorem 1.1

The idea behind the proof of Theorem 1.1 is simply to express the Strichartz norm

$$\|e^{is\Delta} f\|_{L^p_s L^q_x(\mathbb{R}\times\mathbb{R}^d)}$$

in terms of quantities which are already known to be monotone under the heat flow that we consider. As we shall see, this essentially amounts to bringing together the Strichartz-norm representation formulae of Hundertmark and Zharnitsky [2006] and the following heat-flow monotonicity property inherent in the Cauchy–Schwarz inequality.

Lemma 2.1. *For $n \in \mathbb{N}$ and nonnegative integrable functions f_1 and f_2 on $\mathbb{R}^n$, the quantity*

$$\Lambda(t) := \int_{\mathbb{R}^n} (e^{t\Delta} f_1)^{1/2} (e^{t\Delta} f_2)^{1/2}$$

is nondecreasing for all $t > 0$.

Proof. Let $0 < t_1 < t_2$. If H_t denotes the heat kernel on $\mathbb{R}^n$ given by (1-6) then,

$$\begin{aligned}\Lambda(t_1) &= \int_{\mathbb{R}^n} (H_{t_1} * f_1)^{1/2} (H_{t_1} * f_2)^{1/2} = \int_{\mathbb{R}^n} H_{t_2-t_1} * \big((H_{t_1} * f_1)^{1/2} (H_{t_1} * f_2)^{1/2}\big) \\ &= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \big(H_{t_2-t_1}(x-y) H_{t_1} * f_1(y)\big)^{1/2} \big(H_{t_2-t_1}(x-y) H_{t_1} * f_2(y)\big)^{1/2} dy\, dx \\ &\le \int_{\mathbb{R}^n} \big(H_{t_2-t_1} * (H_{t_1} * f_1)\big)^{1/2} \big(H_{t_2-t_1} * (H_{t_1} * f_2)\big)^{1/2} \\ &= \Lambda(t_2),\end{aligned}$$

where we have used the Cauchy–Schwarz inequality on $L^2(\mathbb{R}^n)$ and the semigroup property of the heat kernel. □

This proof of Lemma 2.1 originates in [Ball 1989] and was developed further in [Bennett et al. 2008a]. An alternative method of proof, used in [Carlen et al. 2004] and [Bennett et al. 2008a], is based on the divergence theorem and produces the explicit formula

$$\Lambda'(t) = \tfrac{1}{4}\int_{\mathbb{R}^n} \big|\nabla(\log e^{t\Delta} f_1) - \nabla(\log e^{t\Delta} f_2)\big|^2 (e^{t\Delta} f_1)^{1/2} (e^{t\Delta} f_2)^{1/2} \tag{2-1}$$

for each $t > 0$, provided f_1 and f_2 are sufficiently well behaved (for instance, bounded with compact support). We remark in passing that the Cauchy–Schwarz inequality on $L^2(\mathbb{R}^n)$ follows from Lemma 2.1 by comparing the limiting values of $\Lambda(t)$ for t at zero and infinity.

The next lemma is an observation of Hundertmark and Zharnitsky [2006], who showed that multiplied out expressions for the Strichartz norm in the $(1, 6, 6)$ and $(2, 4, 4)$ cases have a particularly simple geometric interpretation.

Lemma 2.2. (1) *For nonnegative* $f \in L^2(\mathbb{R})$,

$$\|e^{is\Delta} f\|^6_{L^6_s L^6_x(\mathbb{R}\times\mathbb{R})} = \frac{1}{2\sqrt{3}} \int_{\mathbb{R}^3} (f\otimes f\otimes f)(X) P_1(f\otimes f\otimes f)(X)\, dX,$$

where $P_1 : L^2(\mathbb{R}^3) \to L^2(\mathbb{R}^3)$ *is the projection operator onto the subspace of functions on* $\mathbb{R}^3$ *invariant under the isometries that fix the direction* $(1, 1, 1)$.

(2) *For nonnegative* $f \in L^2(\mathbb{R}^2)$,

$$\|e^{is\Delta} f\|^4_{L^4_s L^4_x(\mathbb{R}\times\mathbb{R}^2)} = \frac{1}{4} \int_{\mathbb{R}^4} (f\otimes f)(X) P_2(f\otimes f)(X)\, dX,$$

where $P_2 : L^2(\mathbb{R}^4) \to L^2(\mathbb{R}^4)$ *is the projection operator onto the subspace of functions on* $\mathbb{R}^4$ *invariant under the isometries that fix the directions* $(1, 0, 1, 0)$ *and* $(0, 1, 0, 1)$.

Proof of Theorem 1.1. We begin with the case where (p, q, d) is equal to $(1, 6, 6)$. For functions $G \in L^2(\mathbb{R}^3)$ we may write

$$P_1 G(X) = \int_O G(\rho X)\, d\mathcal{H}(\rho), \tag{2-2}$$

where O is the group of isometries on $\mathbb{R}^3$ that coincide with the identity on the span of $(1, 1, 1)$ and $d\mathcal{H}$ denotes the right-invariant Haar probability measure on O.

If, for $f \in L^2(\mathbb{R})$, we let $F := f\otimes f\otimes f$ then it is easy to see that

$$e^{t\Delta}|f|^2 \otimes e^{t\Delta}|f|^2 \otimes e^{t\Delta}|f|^2 = e^{t\Delta}|F|^2, \tag{2-3}$$

because, in general, the heat operator $e^{t\Delta}$ commutes with tensor products. It is also easy to check that for each isometry ρ on $\mathbb{R}^3$,

$$\big(e^{t\Delta}|f|^2 \otimes e^{t\Delta}|f|^2 \otimes e^{t\Delta}|f|^2\big)(\rho\,\cdot) = e^{t\Delta}|F_\rho|^2, \tag{2-4}$$

where $F_\rho := F(\rho\,\cdot)$. In (2-3) and (2-4) the Laplacian Δ acts in the number of variables dictated by context. Therefore, by Lemma 2.2(1),

$$Q_{6,6}(t)^6 = \frac{1}{2\sqrt{3}} \int_O \int_{\mathbb{R}^3} (e^{t\Delta}|F|^2)^{1/2}(X) (e^{t\Delta}|F_\rho|^2)^{1/2}(X)\, dX\, d\mathcal{H}(\rho)$$

and, by Lemma 2.1 and the nonnegativity of the measure $d\mathscr{H}$, it follows that $Q_{6,6}(t)$ is nondecreasing for each $t > 0$.

For the $(2,4,4)$ case, we use a representation of the form (2-2) for the projection operator P_2 where the averaging group O is replaced by the group of isometries on $\mathbb{R}^4$ which coincide with the identity on the span of $(1,0,1,0)$ and $(0,1,0,1)$. Of course, the analogous statements to (2-3) and (2-4) involving two-fold tensor products hold. Hence the nondecreasingness of $Q_{4,4}$ follows from Lemma 2.2(2) and Lemma 2.1.

Finally, for the $(1,8,4)$ case we observe that

$$\left\|e^{is\Delta}(e^{t\Delta}|f|^2)^{1/2}\right\|^2_{L^8_sL^4_x(\mathbb{R}\times\mathbb{R})} = \left\|e^{is\Delta}(e^{t\Delta}(|f|^2\otimes|f|^2))^{1/2}\right\|_{L^4_sL^4_x(\mathbb{R}\times\mathbb{R}^2)} \tag{2-5}$$

because both solution operators $e^{is\Delta}$ and $e^{t\Delta}$ commute with tensor products. Therefore, the claimed monotonicity in the $(1,8,4)$ case follows from the corresponding claim in the $(2,4,4)$ case. This completes the proof of Theorem 1.1. □

It is transparent from the proof of Theorem 1.1 and (2-1) how one may obtain an explicit formula for $Q'_{p,q}(t)$ provided q is an even integer which divides p and f is sufficiently well behaved (say, bounded with compact support). For example, using the notation used in the proof of Theorem 1.1,

$$\frac{d}{dt}(Q_{6,6}(t)^6) = \frac{1}{8\sqrt{3}}\int_O\int_{\mathbb{R}^3}\left|V(t,X)-\rho^tV(t,\rho X)\right|^2(e^{t\Delta}|F|^2)^{1/2}(e^{t\Delta}|F_\rho|^2)^{1/2}\,dX\,d\mathscr{H}(\rho),$$

where $V(t,\cdot)$ denotes the time-dependent vector field on $\mathbb{R}^3$ given by

$$V(t,X) = \nabla\big(\log e^{t\Delta}|F|^2\big)(X)$$

and ρ^t denotes the transpose of ρ.

Lemma 2.2, combined with a further argument from [Hundertmark and Zharnitsky 2006] (where explicit details can be found), shows that Gaussians are the only extremisers of the Strichartz inequality in the cases $(d,p,q) = (1,6,6), (2,4,4)$. The same conclusion for the case $(d,p,q) = (1,8,4)$ follows quickly from that for the case $(d,p,q) = (2,4,4)$ by (2-5).

3. Further results

Mehler flow. The operator $L := \Delta - \langle x, \nabla\rangle$ generates the Mehler semigroup e^{tL} (sometimes called the Ornstein–Uhlenbeck semigroup) given by

$$e^{tL}f(x) = \int_{\mathbb{R}^d} f\big(e^{-t}x + \sqrt{1-e^{-2t}}\,y\big)\,d\gamma_d(y)$$

for suitable functions f on $\mathbb{R}^d$, where $d\gamma_d$ is the Gaussian probability measure on $\mathbb{R}^d$ given by

$$d\gamma_d(y) = \frac{1}{(2\pi)^{d/2}}e^{-|y|^2/2}dy.$$

Naturally, $u(t,\cdot) := e^{tL}f$ satisfies the evolution equation

$$\partial_t u = Lu$$

with initial datum $u(0, x) = f(x)$. It will be convenient to restrict our attention to functions f which are bounded and compactly supported.

The purpose of this remark is to highlight that when (d, p, q) is one of $(1, 6, 6)$, $(1, 8, 4)$, or $(2, 4, 4)$ the Strichartz norm also exhibits a certain monotonicity subject to the input evolving according to a quadratic Mehler flow.

Theorem 3.1. *Suppose f is a bounded and compactly supported function on $\mathbb{R}^d$. If (d, p, q) is Schrödinger admissible and q is an even integer dividing p, then the quantity*

$$Q(t) := \left\| e^{is\Delta}(e^{-|\cdot|^2/2} e^{tL}|f|^2)^{1/2} \right\|_{L^p_s L^q_x(\mathbb{R}\times\mathbb{R}^d)}$$

is nondecreasing for all $t > 0$.

As a consequence of Theorem 3.1, we may again recover sharp forms of the Strichartz estimates in (1-2) for such exponents by considering the limiting values of $Q(t)$ as t approaches zero and infinity. In particular, since

$$e^{tL}|f|^2(x) = \int_{\mathbb{R}^d} |f|^2\big(e^{-t}x + \sqrt{1-e^{-2t}}\,y\big)\, d\gamma_d(y),$$

it follows that, for each $x \in \mathbb{R}^d$, $e^{tL}|f|^2(x)$ tends to $\int_{\mathbb{R}^d} |f|^2 d\gamma_d$ as t tends to infinity. Thus, the monotonicity of Q implies that

$$\left\| e^{is\Delta}(e^{-|\cdot|^2/4}|f|) \right\|_{L^p_s L^q_x(\mathbb{R}\times\mathbb{R}^d)} \le \left\| e^{is\Delta}(e^{-|\cdot|^2/4}) \right\|_{L^p_s L^q_x(\mathbb{R}\times\mathbb{R}^d)} \left(\int_{\mathbb{R}^d} |f|^2 d\gamma_d \right)^{1/2}$$

for each bounded and compactly supported function f on $\mathbb{R}^d$. Thus,

$$\|e^{is\Delta} g\|_{L^p_s L^q_x(\mathbb{R}\times\mathbb{R}^d)} \le \left\| e^{is\Delta}\Big(\tfrac{1}{(2\pi)^{d/2}} e^{-|\cdot|^2/2}\Big)^{1/2} \right\|_{L^p_s L^q_x(\mathbb{R}\times\mathbb{R}^d)} \|g\|_{L^2(\mathbb{R}^d)}$$

for each $g \in L^2(\mathbb{R}^d)$.

The first key ingredient in the proof of Theorem 3.1 is to observe that an analogue of Lemma 2.1 holds for Mehler flow.

Lemma 3.2. *Let $n \in \mathbb{N}$ and let f_1 and f_2 be nonnegative, bounded and compactly supported functions on $\mathbb{R}^n$. Then the quantity*

$$\Lambda(t) := \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \big(e^{-|\cdot|^2/2} e^{tL} f_1\big)^{1/2} \big(e^{-|\cdot|^2/2} e^{tL} f_2\big)^{1/2}$$

is nondecreasing for all $t > 0$.

Proof. Notice that

$$e^{\log \frac{1}{\sqrt{1-2T}} L} f_j\Big(\frac{x}{\sqrt{1-2T}}\Big) = e^{T\Delta} f_j(x) = H_T * f_j(x)$$

for each $0 < T < 1/2$. Thus, for $0 < T_1 < T_2 < 1/2$ we have

$$\begin{aligned}\Lambda\Big(\log\frac{1}{\sqrt{1-2T_1}}\Big) &= \int_{\mathbb{R}^n} (f_1 * H_{T_1})^{1/2}(f_2 * H_{T_1})^{1/2} H_{1/2-T_1}\\ &= \int_{\mathbb{R}^n} (f_1 * H_{T_1})^{1/2}(f_2 * H_{T_1})^{1/2}(H_{T_2-T_1} * H_{1/2-T_2})\\ &= \int_{\mathbb{R}^n} \big[H_{T_2-T_1} * ((f_1 * H_{T_1})^{1/2}(f_2 * H_{T_1})^{1/2})\big] H_{1/2-T_2}\end{aligned}$$

using the semigroup property and evenness of the heat kernel. As in the proof of Lemma 2.1 it follows from the Cauchy–Schwarz inequality and another application of the semigroup property of the heat kernel that

$$H_{T_2-T_1} * \big((f_1 * H_{T_1})^{1/2}(f_2 * H_{T_1})^{1/2}\big) \le (f_1 * H_{T_2})^{1/2}(f_2 * H_{T_2})^{1/2},$$

and thus

$$\Lambda\Big(\log\frac{1}{\sqrt{1-2T_1}}\Big) \le \Lambda\Big(\log\frac{1}{\sqrt{1-2T_2}}\Big).$$

Hence, $\Lambda(t_1) \le \Lambda(t_2)$ for $0 < t_1 < t_2$. □

As with Lemma 2.1, it is possible to prove Lemma 3.2 in a way that produces an explicit formula for $\Lambda'(t)$ for each $t > 0$, from which the monotonicity of Λ is manifest. To see this, let $\mathfrak{u}_j : (0,\infty)\times\mathbb{R}^n \to \mathbb{R}$ be given by

$$\mathfrak{u}_j(t,x) = e^{-|x|^2/2} e^{tL} f_j(x) = e^{-|x|^2/2} \int_{\mathbb{R}^n} f_j\big(e^{-t}x + \sqrt{1-e^{-2t}}\,y\big)\, d\gamma_n(y) \tag{3-1}$$

for $j = 1, 2$. It is straightforward to check that

$$\partial_t \mathfrak{u}_j = \Delta \mathfrak{u}_j + \langle x, \nabla \mathfrak{u}_j\rangle + n\mathfrak{u}_j$$

and furthermore

$$\partial_t(\log \mathfrak{u}_j) = \operatorname{div}(v_j) + |v_j|^2 + \langle x, v_j\rangle + n,$$

where $v_j := \nabla(\log \mathfrak{u}_j)$. Therefore,

$$\Lambda'(t) = \mathrm{I} + \mathrm{II},$$

where

$$\mathrm{I} := \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \big(\langle x, \tfrac12 v_1 + \tfrac12 v_2\rangle + n\big)(t,x)\, \mathfrak{u}_1(t,x)^{1/2}\mathfrak{u}_2(t,x)^{1/2} dx$$

and

$$\mathrm{II} := \frac{1}{2(2\pi)^{n/2}} \int_{\mathbb{R}^n} \big(\operatorname{div}(v_1) + \operatorname{div}(v_2) + |v_1|^2 + |v_2|^2\big)(t,x)\, \mathfrak{u}_1(t,x)^{1/2}\mathfrak{u}_2(t,x)^{1/2} dx.$$

Since

$$\mathrm{I} = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \operatorname{div}(\mathfrak{u}_1(t,x)^{1/2}\mathfrak{u}_2(t,x)^{1/2} x)\, dx,$$

it follows from the divergence theorem that I vanishes. Using the fact that each f_j is bounded with compact support it follows from the explicit formula for $\mathfrak{u}_j$ in (3-1) that $v_j(t,x)$ grows at most polynomially in x for each fixed $t > 0$, so $\int_{\mathbb{R}^n} \operatorname{div}(\mathrm{u}_1^{1/2}\mathrm{u}_2^{1/2}v_j)$ vanishes by the divergence theorem. Therefore, for each $t > 0$,

$$\Lambda'(t) = \frac{1}{4(2\pi)^{n/2}} \int_{\mathbb{R}^n} \big|v_1(t,x) - v_2(t,x)\big|^2 \mathrm{u}_1(t,x)^{1/2}\mathrm{u}_2(t,x)^{1/2} dx,$$

which is manifestly nonnegative.

The above argument which proves Lemma 3.2 based on the divergence theorem is very much in the spirit of the heat-flow monotonicity results in [Carlen et al. 2004] and [Bennett et al. 2008a] and naturally extends to the setting of the geometric Brascamp–Lieb inequality. In particular, for $j = 1, \dots, m$ suppose that $p_j \geq 1$ and $B_j : \mathbb{R}^n \to \mathbb{R}^{n_j}$ is a linear mapping such that $B_j^* B_j$ is a projection and $\sum_{j=1}^m \frac{1}{p_j} B_j^* B_j = I_{\mathbb{R}^n}$. Then the quantity

$$\frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \prod_{j=1}^m \big(e^{-|B_j x|^2/2}(e^{tL} f_j)(B_j x)\big)^{1/p_j} dx = \int_{\mathbb{R}^n} \prod_{j=1}^m (e^{tL} f_j)(B_j x)^{1/p_j} d\gamma_n(x)$$

is nondecreasing for each $t > 0$ provided each f_j is a nonnegative, bounded and compactly supported function on $\mathbb{R}^{n_j}$. This is due to Barthe and Cordero-Erausquin [2004] in the case where each B_j has rank one. A modification of the argument gives the general rank case (see [Carlen and Lieb 2008] for closely related results).

By following the same argument employed in our proof of Theorem 1.1, to conclude the proof of Theorem 3.1 it suffices to note that Mehler flow appropriately respects tensor products and isometries. In particular we need that if F is the m-fold tensor product of f then

$$\bigotimes_{j=1}^m e^{-|\cdot|^2/2} e^{tL} |f|^2 = e^{-|\cdot|^2/2} e^{tL} |F|^2 \tag{3-2}$$

and, for each isometry ρ on $(\mathbb{R}^d)^m$,

$$\bigotimes_{j=1}^m e^{-|\cdot|^2/2} e^{tL} |f|^2(\rho\,\cdot) = e^{-|\cdot|^2/2} e^{tL} |F_\rho|^2, \tag{3-3}$$

where $F_\rho := F(\rho\,\cdot)$. Here, the operators $|\cdot|$ and L are acting on the number of variables dictated by context. The verification of (3-2) and (3-3) is an easy exercise.

Mitigating powers of t. It is possible to relax the quadratic nature of the heat flow in the quantity $Q_{p,q}$ in Theorem 1.1 by inserting as a mitigating factor a well-chosen power of t.

Theorem 3.3. *Suppose that (p, q, d) is Schrödinger-admissible and q is an even integer which divides p. If f is a nonnegative integrable function on $\mathbb{R}^d$ and $\alpha \in [1/2, 1]$, the quantity*

$$t^{d(\alpha-1/2)/2} \| e^{is\Delta}(e^{t\Delta} f)^\alpha \|_{L^p_s L^q_x(\mathbb{R}\times\mathbb{R}^d)}$$

is nondecreasing for each $t > 0$.

By [Bennett et al. 2008a], Lemma 2.1 generalises to the statement that

$$t^{n(\alpha-1/2)} \int_{\mathbb{R}^n} (e^{t\Delta} f_1)^\alpha (e^{t\Delta} f_2)^\alpha \tag{3-4}$$

is nondecreasing for all $t > 0$ provided $n \in \mathbb{N}$, $\alpha \in [1/2, 1]$ and f_1, f_2 are nonnegative integrable functions on $\mathbb{R}^n$. Thus Theorem 3.3 follows by the same argument in our proof of Theorem 1.1.

Higher dimensions. Theorem 1.1 raises obvious questions about higher-dimensional analogues and consequently the potential of our approach to prove the sharp form of (1-2) in all dimensions (at least for nonnegative initial data f). Shao [2009] has shown that for nonendpoint Schrödinger-admissible triples (p, q, d),

$$\sup\bigl\{\|e^{is\Delta} f\|_{L^p_s L^q_x(\mathbb{R}\times\mathbb{R}^d)} : \|f\|_{L^2(\mathbb{R}^d)} = 1\bigr\}$$

is at least attained, although he does not determine the explicit form of an extremiser. There is some anecdotal evidence in [Bennett et al. 2008b] to suggest that Theorem 1.1 may not extend to all Schrödinger-admissible triples (d, p, q). Nevertheless, we end this section with a discussion of some results in this direction which we believe to be of some interest.

We shall consider the case $p = q = 2 + 4/d$ and it will be convenient to denote this number by $p(d)$. Since $p(d)$ is not an even integer for $d \geq 3$, one possible approach to the question of monotonicity of $Q_{p(d),p(d)}$, given by (1-5), is to attempt to embed the Strichartz norm

$$|||f|||_{p(d)} := \|e^{is\Delta} f\|_{L^{2+4/d}_{s,x}(\mathbb{R}\times\mathbb{R}^d)}$$

in a one-parameter family of norms $|||\cdot|||_p$ which are appropriately monotone under a quadratic flow for $p \in 2\mathbb{N}$, and for which the resulting monotonicity formula may be extrapolated, in a sign-preserving way, to $p = p(d)$. Such an approach has proved effective in the context of the general Brascamp–Lieb inequalities, and was central to the approach to the multilinear Kakeya and Strichartz inequalities in [Bennett et al. 2006].

Our analysis for $d = 1, 2$ suggests (albeit rather indirectly) a natural candidate for such a family of norms. For each $d \in \mathbb{N}$ and $p > p(d)$, we define a norm $|||\cdot|||_p$ on $\mathcal{S}(\mathbb{R}^d)$ by

$$|||f|||_p^p = \frac{(p(d)/\pi)^{d/2}}{(2\pi)^{d+2}} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_0^\infty\int_{\mathbb{R}} \left|\int_{\mathbb{R}^d} e^{-|z-\sqrt{\zeta}\xi|^2} e^{i(x\cdot\xi - s|\xi|^2)} \widehat{f}(\xi)\, d\xi\right|^p \frac{\zeta^{\nu-1}}{\Gamma(\nu)} ds\, d\zeta\, dz\, dx,$$

where $\nu = d(p - p(d))/4$.

Theorem 3.4. *As p tends to $p(d)$, the norm $|||f|||_p$ converges to the Strichartz norm $\|e^{is\Delta} f\|_{L^{p(d)}_{s,x}}$ for each f belonging to the Schwartz class on $\mathbb{R}^d$. Additionally, if $\alpha \in [1/2, 1]$ and f is a nonnegative integrable function on $\mathbb{R}^d$ then*

$$\widetilde{Q}_{\alpha,p}(t) := t^{d(\alpha-1/2)/2} |||(e^{t\Delta} f)^\alpha|||_p$$

is nondecreasing for all $t > 0$ whenever p is an even integer.

Remarks. (1) This modified Strichartz norm $|||f|||_p$ is related in spirit to the norm

$$\|I_\beta\, e^{is\Delta} f\|_{L^p_{s,x}(\mathbb{R}\times\mathbb{R}^d)},$$

where I_β denotes the fractional integral of order $\beta = d(p - p(d))/2p$. Although it is true that for all $p \geq p(d)$,

$$\|I_\beta\, e^{is\Delta} f\|_{L^p_{s,x}(\mathbb{R}\times\mathbb{R}^d)} \leq C\|f\|_{L^2(\mathbb{R}^d)}$$

for some finite constant C, the desired heat-flow monotonicity for $p \in 2\mathbb{N}$ is far from apparent for these norms.

(2) Both the Strichartz norm and the modified Strichartz norms $|||\cdot|||_p$ are invariant under the Fourier transform; that is

$$\|e^{is\Delta}\widehat{f}\|_{L^{p(d)}_{s,x}(\mathbb{R}\times\mathbb{R}^d)} = \|e^{is\Delta} f\|_{L^{p(d)}_{s,x}(\mathbb{R}\times\mathbb{R}^d)} \tag{3-5}$$

for all $d \in \mathbb{N}$ and

$$|||\widehat{f}|||_p = |||f|||_p \tag{3-6}$$

for all $p > p(d)$ and $d \in \mathbb{N}$. This observation follows by direct computation and simple changes of variables; for the Strichartz norm it was noted for $d = 1, 2$ in [Hundertmark and Zharnitsky 2006]. We note that in the proof of Theorem 3.4 below we use the invariance in (3-6) for even integers p which (as we will see) follows from Parseval's theorem.

(3) For every integer $m \geq 2$ and in all dimensions $d \geq 1$, a corollary to the case $\alpha = 1/2$ of Theorem 3.4 is the sharp inequality

$$|||f|||_{2m} \leq C_{d,m}\|f\|_{L^2(\mathbb{R}^d)},$$

where the constant $C_{d,m}$ is given by

$$C_{d,m}^{2m} = \frac{\pi^\nu}{2^{\nu+1}m^d\Gamma(\nu+1)}\left(\frac{p(d)}{2}\right)^{d/2}. \tag{3-7}$$

Here $\nu = d(2m - p(d))/4$ as before.

(4) It is known that for nonnegative integrable functions f on $\mathbb{R}^d$ the quantity

$$\|\widehat{(e^{t\Delta} f)^{1/p}}\|_{L^{p'}(\mathbb{R}^d)}$$

is nondecreasing for each $t > 0$ provided the conjugate exponent p' is an even integer; this follows from [Bennett et al. 2008a] and [Bennett and Bez 2009]. However, tying in with our earlier comment on the extension of Theorem 1.1 to all Schrödinger-admissible exponents, in [Bennett et al. 2008b] we show that whenever $p' > 2$ is not an even integer there exists a nonnegative integrable function f such that $Q(t)$ is *strictly decreasing* for all sufficiently small $t > 0$.

Proof of Theorem 3.4. To see the claimed limiting behaviour of $|||f|||_p$ as p tends to $p(d)$ observe that

$$\lim_{\nu\to 0}\frac{1}{\Gamma(\nu)}\int_0^\infty \phi(\nu,\zeta)\zeta^{\nu-1}\,d\zeta = \phi(0,0) \tag{3-8}$$

for any ϕ on $[0,\infty)\times[0,\infty)$ satisfying certain mild regularity conditions. For example, (3-8) holds if ϕ is continuous at the origin and there exist constants $C, \varepsilon > 0$ such that, locally uniformly in ν, one has $|\phi(\nu,\zeta) - \phi(\nu,0)| \leq C|\zeta|^\varepsilon$ for all ζ in a neighbourhood of zero and $|\phi(\nu,\zeta)| \leq C|\zeta|^{-\varepsilon}$ for all ζ bounded away from a neighbourhood of zero. One can check that standard estimates (for example, Strichartz estimates of the form (1-2) for compactly supported functions) imply that for f belonging to the Schwartz class on $\mathbb{R}^d$,

$$\phi(\nu,\zeta) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}}\left|\int_{\mathbb{R}^d} e^{-|z-\sqrt{\zeta}\xi|^2} e^{i(x\cdot\xi - s|\xi|^2)}\widehat{f}(\xi)\,d\xi\right|^p ds\,dx\,dz$$

satisfies such conditions.

We now turn to the monotonicity claim, beginning with some notation. Suppose that $p = 2m$ for some positive integer m. For a nonnegative $f \in \mathcal{S}(\mathbb{R}^d)$ let $F : \mathbb{R}^{md} \to \mathbb{R}$ be given by $F(X) = \otimes_{j=1}^{m} f(X)$, where $X = (\xi_1, \dots, \xi_m) \in (\mathbb{R}^d)^m \cong \mathbb{R}^{md}$. Next we define the subspace W of $\mathbb{R}^{md}$ to be the linear span of $\mathbf{1}_1, \dots, \mathbf{1}_d$, where for each $1 \le j \le d$, $\mathbf{1}_j := (e_j, \dots, e_j)/\sqrt{m}$ and e_j denotes the jth standard basis vector of $\mathbb{R}^d$. For a vector $X \in \mathbb{R}^{md}$ we denote by X_W and $X_{W^\perp}$ the orthogonal projections of X onto W and $W^\perp$ respectively. Now,

$$\|\!| f \|\!|_{2m}^{2m} = \tfrac{1}{2^{d+1}\pi} \big(\tfrac{p(d)}{m\pi}\big)^{d/2} \int \delta(X_W - Y_W)\delta(|X|^2 - |Y|^2) K(X,Y) F(X) F(Y)\, dX dY,$$

where we integrate over $\mathbb{R}^{md} \times \mathbb{R}^{md}$ and

$$\begin{aligned} K(X,Y) &= \int_0^\infty \frac{\zeta^{\nu-1}}{\Gamma(\nu)} e^{-\zeta(|X|^2+|Y|^2)} \int_{\mathbb{R}^d} e^{\sqrt{m\zeta} z\cdot(X_W+Y_W)} e^{-m|z|^2/2} dz\, d\zeta \\ &= \big(\tfrac{2\pi}{m}\big)^{d/2} \int_0^\infty \frac{\zeta^{\nu-1}}{\Gamma(\nu)} e^{-\zeta(|X|^2+|Y|^2)} e^{\zeta|X_W+Y_W|^2/2} d\zeta \end{aligned}$$

for $(X,Y) \in \mathbb{R}^{md} \times \mathbb{R}^{md}$. Thus, on the support of the delta distributions ($X_W = Y_W$ and $|X|^2 = |Y|^2$) we have

$$K(X,Y) = \big(\tfrac{2\pi}{m}\big)^{d/2} \int_0^\infty \frac{\zeta^{\nu-1}}{\Gamma(\nu)} e^{-2\zeta(|X|^2-|X_W|^2)} d\zeta = \tfrac{1}{2^\nu}\big(\tfrac{2\pi}{m}\big)^{d/2} \frac{1}{(|X|^2 - |X_W|^2)^\nu} = \tfrac{1}{2^\nu}\big(\tfrac{2\pi}{m}\big)^{d/2} \frac{1}{|X_{W^\perp}|^{2\nu}}.$$

Therefore

$$\|\!| f \|\!|_{2m}^{2m} = \tfrac{\pi^\nu}{2^{\nu+1} m^d \Gamma(\nu+1)} \big(\tfrac{p(d)}{2}\big)^{d/2} \int_{\mathbb{R}^{md}} F(X) PF(X)\, dX, \tag{3-9}$$

where P is given by

$$PF(X) = \tfrac{\Gamma(\nu+1)}{\pi^{\nu+1}} \frac{1}{|X_{W^\perp}|^{2\nu}} \int_{\mathbb{R}^{md}} \delta(X_W - Y_W)\delta(|X|^2 - |Y|^2) F(Y)\, dY.$$

Using polar coordinates in $W^\perp$ in the above integral and recalling that $\nu = d(2m - p(d))/4$ identifies P as the orthogonal projection onto functions on $\mathbb{R}^{md}$ which are invariant under the action of O, the group of isometries on $\mathbb{R}^{md}$ which coincide with the identity on W; that is,

$$PF(X) = \int_O F(\rho X)\, d\mathcal{H}(\rho),$$

where $d\mathcal{H}$ denotes the right-invariant Haar probability measure on O.

Finally, applying the representation of $\|\!| f \|\!|_{2m}^{2m}$ in (3-9) to the quantity $\widetilde{Q}_{\alpha,2m}$, and appealing to the nondecreasingness of the quantity in (3-4), we conclude that $\widetilde{Q}_{\alpha,2m}(t)$ is nondecreasing for all $t > 0$ and all $\alpha \in [1/2, 1]$. This completes the proof of Theorem 3.4. □

References

[Ball 1989] K. Ball, "Volumes of sections of cubes and related problems", pp. 251–260 in *Geometric aspects of functional analysis (1987–88)*, edited by J. Lindenstrauss and V. D. Milman, Lecture Notes in Math. **1376**, Springer, Berlin, 1989. MR 90i:52019 Zbl 0674.46008

[Barthe and Cordero-Erausquin 2004] F. Barthe and D. Cordero-Erausquin, "Inverse Brascamp–Lieb inequalities along the heat equation", pp. 65–71 in *Geometric aspects of functional analysis*, edited by V. D. Milman and G. Schechtman, Lecture Notes in Math. **1850**, Springer, Berlin, 2004. MR 2005i:52012 Zbl 1086.26010

[Bennett and Bez 2009] J. Bennett and N. Bez, "Closure properties of solutions to heat inequalities", *J. Geom. Anal.* **19**:3 (2009), 584–600. MR 2496567 Zbl 05589981

[Bennett et al. 2006] J. Bennett, A. Carbery, and T. Tao, "On the multilinear restriction and Kakeya conjectures", *Acta Math.* **196**:2 (2006), 261–302. MR 2007h:42019 Zbl 05114945

[Bennett et al. 2008a] J. Bennett, A. Carbery, M. Christ, and T. Tao, "The Brascamp–Lieb inequalities: finiteness, structure and extremals", *Geom. Funct. Anal.* **17**:5 (2008), 1343–1415. MR 2009c:42052 Zbl 1132.26006

[Bennett et al. 2008b] J. M. Bennett, N. Bez, and A. Carbery, "Heat-flow monotonicity related to the Hausdorff–Young inequality", preprint, 2008. To appear in *Bull. London Math. Soc.*, 2009. arXiv 0806.4329

[Carlen and Lieb 2008] E. A. Carlen and E. H. Lieb, "Brascamp–Lieb inequalities for non-commutative integration", *Doc. Math.* **13** (2008), 553–584. MR 2452876 Zbl 05367928

[Carlen et al. 2004] E. A. Carlen, E. H. Lieb, and M. Loss, "A sharp analog of Young's inequality on S^N and related entropy inequalities", *J. Geom. Anal.* **14**:3 (2004), 487–520. MR 2005k:82046 Zbl 1056.43002

[Foschi 2007] D. Foschi, "Maximizers for the Strichartz inequality", *J. Eur. Math. Soc.* (*JEMS*) **9**:4 (2007), 739–774. MR 2008k:35389 Zbl 05255314

[Hundertmark and Zharnitsky 2006] D. Hundertmark and V. Zharnitsky, "On sharp Strichartz inequalities in low dimensions", *Int. Math. Res. Not.* (2006), id 34080. MR 2007b:35277 Zbl 1131.35308

[Keel and Tao 1998] M. Keel and T. Tao, "Endpoint Strichartz estimates", *Amer. J. Math.* **120**:5 (1998), 955–980. MR 2000d:35018 Zbl 0922.35028

[Shao 2009] S. Shao, "Maximizers for the Strichartz inequalities and the Sobolev–Strichartz inequalities for the Schrödinger equation", preprint, 2009. Zbl 05561914

[Strichartz 1977] R. S. Strichartz, "Restrictions of Fourier transforms to quadratic surfaces and decay of solutions of wave equations", *Duke Math. J.* **44**:3 (1977), 705–714. MR 58 #23577 Zbl 0372.35001

[Tomas 1975] P. A. Tomas, "A restriction theorem for the Fourier transform", *Bull. Amer. Math. Soc.* **81** (1975), 477–478. MR 50 #10681 Zbl 0298.42011

Received 15 Oct 2008. Revised 26 Jan 2009. Accepted 24 Mar 2009.

JONATHAN BENNETT: J.Bennett@bham.ac.uk
School of Mathematics, The University of Birmingham, The Watson Building, Edgbaston, Birmingham, B15 2TT, United Kingdom

NEAL BEZ: n.bez@maths.gla.ac.uk
Department of Mathematics, University Gardens, University of Glasgow, G12 8QW, United Kingdom

ANTHONY CARBERY: A.Carbery@ed.ac.uk
School of Mathematics and Maxwell Institute for Mathematical Sciences, The University of Edinburgh, James Clerk Maxwell Building, The King's Buildings, Edinburgh, EH3 9JZ, United Kingdom

DIRK HUNDERTMARK: dirk@math.uiuc.edu
Department of Mathematics and Institute for Condensed Matter Theory, Altgeld Hall, University of Illinois at Urbana–Champaign, Urbana, IL 61801, United States

ANALYSIS AND PDE
Vol. 2, No. 2, 2009

DYNAMICS OF VORTICES FOR THE COMPLEX GINZBURG–LANDAU EQUATION

EVELYNE MIOT

We study a complex Ginzburg–Landau equation in the plane, which has the form of a Gross–Pitaevskii equation with some dissipation added. We focus on the regime corresponding to well-prepared unitary vortices and derive their asymptotic motion law.

1. Introduction

We study the dynamics of vortices for a complex Ginzburg–Landau equation on the plane, namely

$$\frac{\delta}{|\log \varepsilon|}\partial_t u_\varepsilon + \alpha i \partial_t u_\varepsilon = \Delta u_\varepsilon + \frac{1}{\varepsilon^2} u_\varepsilon (1 - |u_\varepsilon|^2), \tag{CGL$_\varepsilon$}$$

where $u_\varepsilon : \mathbb{R}_+ \times \mathbb{R}^2 \to \mathbb{R}^2$ is a complex-valued map. Here δ, α, and ε denote positive real parameters, and we will mainly focus on the asymptotics as ε tends to zero while δ and α are kept fixed. Up to a change of scale, we may further assume that $\alpha = 1$, and we set $k_\varepsilon = \delta/|\log \varepsilon|$. The complex Ginzburg–Landau equation $(\mathrm{CGL})_\varepsilon$ reduces to the Gross–Pitaevskii equation when $\delta = 0$ and to the parabolic Ginzburg–Landau equation when $\alpha = 0$. Both the Gross–Pitaevskii and the Ginzburg–Landau equations have been widely investigated in the regime we will consider (see, for example, [Colliander and Jerrard 1998; Lin and Xin 1999; Jerrard and Spirn 2008; Bethuel et al. 2008] for the Gross–Pitaevskii equation and [Jerrard and Soner 1998; Serfaty 2007; Bethuel et al. 2007] and references therein for the parabolic Ginzburg–Landau equation). Typical functions u_ε in this regime are given explicitly by

$$u_\varepsilon^*(a_i, d_i)(z) := \prod_{i=1}^{l} u_{\varepsilon,d_i}(z - a_i) = \prod_{i=1}^{l} f_{1,d_i}\left(\frac{|z-a_i|}{\varepsilon}\right)\left(\frac{z-a_i}{|z-a_i|}\right)^{d_i},$$

where the points $a_i \in \mathbb{R}^2$, $d_i = \pm 1$, and the functions $f_{1,d_i} : \mathbb{R}^+ \mapsto [0, 1]$, which satisfy $f_{1,d_i}(0) = 0$, $f_{1,d_i}(+\infty) = 1$, are in some sense optimal profiles. The points a_i are called the vortices of the fields u_ε^* and the d_i their degrees. This class of functions u_ε^* is, of course, not invariant by any of the flows corresponding to these equations, but it is not far from it (see the notion of well-preparedness in Definition 1.2). In particular, it is possible to define notions of point vortices for solutions of $(\mathrm{CGL})_\varepsilon$, at least in an asymptotic way as $\varepsilon \to 0$, and to study their dynamics. This dynamics is eventually governed by a system of ordinary differential equations, at least before collisions.

MSC2000: 35B20, 35B40, 35Q40, 82D55.
Keywords: complex Ginzburg–Landau equation, vortex dynamics.
This work was partly supported by the grant JC05-51279 of the Agence Nationale de la Recherche.

Two relevant quantities in the study of vortex dynamics are the Ginzburg–Landau energy

$$E_\varepsilon(u) = \int_{\mathbb{R}^2} e_\varepsilon(u)\,dx = \int_{\mathbb{R}^2} \frac{|\nabla u|^2}{2} + \frac{(1-|u|^2)^2}{4\varepsilon^2}\,dx$$

through its energy density $e_\varepsilon(u)$, and the Jacobian

$$Ju = \tfrac{1}{2}\,\mathrm{curl}(u \times \nabla u)$$

through its primitive $j(u) = u \times \nabla u$. In the regime we will consider, one has

$$\frac{e_\varepsilon(u_\varepsilon)}{|\log\varepsilon|}\,dx \rightharpoonup \pi \sum_{i=1}^{l} \delta_{a_i} \quad \text{and} \quad Ju_\varepsilon\,dx \rightharpoonup \pi \sum_{i=1}^{l} d_i \delta_{a_i}$$

as $\varepsilon \to 0$, which describes asymptotically the positions and degrees of the vortices. The quantity $e_\varepsilon(u_\varepsilon)$ has been especially used in the study of the parabolic Ginzburg–Landau equation, while $j(u_\varepsilon)$ has been used in the study of the Gross–Pitaevskii equation. Here, we will rely on both of them.

In the case of the domain being the entire plane $\mathbb{R}^2$, which we consider here, the reference fields $u^*_\varepsilon(a_i, d_i)$ have infinite Ginzburg–Landau energy E_ε whenever $d = \sum d_i \neq 0$. In [Bethuel and Smets 2007], a notion of renormalized energy for such data — not to be confused with the one in [Bethuel et al. 1994] — was introduced to solve the Cauchy problem for the Gross–Pitaevskii equation. This notion was later used in [Bethuel et al. 2008] to study the dynamics of vortices for the Gross–Pitaevskii equation in the plane. Our definition of well-prepared data below and part of the subsequent analysis is borrowed from this last reference.

The complex Ginzburg–Landau equation $(\mathrm{CGL})_\varepsilon$, either in the plane or on the real line, has been widely considered in the literature, especially as a model for amplitude oscillation in weakly nonlinear systems undergoing a Hopf bifurcation (see [Aranson and Kramer 2002] for a survey). The mathematical analysis of vortices for $(\mathrm{CGL})_\varepsilon$ was initiated in [Lin and Xin 1999], where it was presented as an alternative approach (a regularized version) for the study of the Gross–Pitaevskii equation. We believe, however, that the conclusion regarding the dynamics of vortices for $(\mathrm{CGL})_\varepsilon$ in [Lin and Xin 1999] is erroneous, and that Theorem 1.3 represents the correct version.

After the completion of this work we were informed that Kurzke, Melcher, Moser, and Spirn [Kurzke et al. 2008] independently obtained similar results concerning the dynamics of vortices for $(\mathrm{CGL})_\varepsilon$ in bounded and simply connected domains.

Renormalized energy and the Cauchy problem. As mentioned, for $d = \sum d_i \neq 0$ the Ginzburg–Landau energy of $u^*_\varepsilon(a_i, d_i)$ is infinite. It can actually be computed that

$$\int_{\mathbb{R}^2} \frac{|\nabla |u^*_\varepsilon(a_i, d_i)||^2}{2} + \frac{(1-|u^*_\varepsilon(a_i, d_i)|^2)^2}{4\varepsilon^2}\,dz < +\infty,$$

whereas as $|z| \to +\infty$,

$$|\nabla u^*_\varepsilon(a_i, d_i)|^2(z) \sim \frac{d^2}{|z|^2},$$

so that

$$\int_{\mathbb{R}^2} \frac{|\nabla u^*_\varepsilon(a_i, d_i)|^2}{2} = +\infty.$$

The renormalized energy introduced in [Bethuel and Smets 2007] is obtained by subtracting the divergent part of the gradient at infinity. More precisely, given a smooth map U_d such that

$$U_d = \left(\frac{z}{|z|}\right)^d \quad \text{on } \mathbb{R}^2 \setminus B(0,1),$$

we have as $|z| \to +\infty$

$$|\nabla u_\varepsilon^*(a_i, d_i)|^2 \sim |\nabla U_d|^2$$

and one may define

$$\mathscr{E}_{\varepsilon,U_d}(u_\varepsilon^*(a_i,d_i)) := \lim_{R\to+\infty} \int_{B(R)} \left(e_\varepsilon(u_\varepsilon^*(a_i,d_i)) - \tfrac{1}{2}|\nabla U_d|^2\right) < +\infty. \tag{1-1}$$

This definition extends to a larger class of functions, and is a useful ingredient in solving the Cauchy problem. Following Bethuel and Smets, we define

$$\mathscr{V} = \left\{U \in L^\infty(\mathbb{R}^2, \mathbb{C}) : \nabla^k U \in L^2 \text{ for all } k \geq 2,\ (1-|U|^2) \in L^2,\ \nabla|U| \in L^2\right\}.$$

In particular, the space $\mathscr{V}$ contains all the maps u_ε^* as well as the reference maps U_d. Our first result, which we prove in the Appendix, establishes global well-posedness in the class $\mathscr{V} + H^1(\mathbb{R}^2)$. (In passing, we mention that Ginibre and Velo [1997] investigated the Cauchy problem in local spaces for a more general class of complex Ginzburg–Landau equations.)

Theorem 1.1. *Let $u_0 = U + w_0$ be in $\mathscr{V} + H^1(\mathbb{R}^2)$. There exists a unique global solution u to $(\mathrm{CGL})_\varepsilon$ such that $u \in C^0(\{U\} + H^1(\mathbb{R}^2))$. If we write $u(t) = U + w(t)$, then w is the unique solution in $C^0(\mathbb{R}_+, H^1(\mathbb{R}^2))$ to*

$$\begin{cases} (k_\varepsilon + i)\partial_t w = \Delta w + f_U(w), \\ w(0) = w_0, \end{cases} \tag{1-2}$$

where

$$f_U(w) = \Delta U + \varepsilon^{-2}(U+w)(1-|U+w|^2).$$

In addition, w satisfies

$$w \in L^1_{\mathrm{loc}}\big(\mathbb{R}_+, H^2(\mathbb{R}^2)\big) \cap L^\infty_{\mathrm{loc}}\big(\mathbb{R}_+^*, L^\infty(\mathbb{R}^2)\big), \quad \partial_t w \in L^1_{\mathrm{loc}}\big(\mathbb{R}_+, L^2(\mathbb{R}^2)\big), \quad w \in C^\infty\big(\mathbb{R}_+^*, C^\infty(\mathbb{R}^2)\big).$$

Finally, the functional $E_{\varepsilon,U}(u) := E_{\varepsilon,U}(w)$ defined by

$$E_{\varepsilon,U}(u) = \int_{\mathbb{R}^2} \frac{|\nabla w|^2}{2} - \int_{\mathbb{R}^2} \Delta U \cdot w + \int_{\mathbb{R}^2} \frac{(1-|U+w|^2)^2}{4\varepsilon^2}$$

satisfies

$$\frac{d}{dt} E_{\varepsilon,U}(u) = -k_\varepsilon \int_{\mathbb{R}^2} |\partial_t w|^2\, dx \quad \textit{for all } t \geq 0.$$

As a matter of fact, it follows from integration by parts that if $u \in \{U\} + H^1(\mathbb{R}^2)$ is as in Theorem 1.1 and if U satisfies in addition $|\nabla U(x)| \leq C/\sqrt{|x|}$, then

$$E_{\varepsilon,U}(u(t)) \equiv \mathscr{E}_{\varepsilon,U}(u(t)) = \lim_{R\to+\infty} \int_{B(R)} \left(e_\varepsilon(u(t)) - \tfrac{1}{2}|\nabla U|^2\right) dx.$$

The functions $u_\varepsilon^*(a_i, d_i)$ are not H^1 perturbations of one another, even for fixed $d = \sum d_i$, unless algebraic relations connect the a_i and d_i. To handle a class of functions containing all the u_ε^*, it is useful to introduce on the set $\mathscr{V}$ the equivalence relation defined, for $U, U' \in \mathscr{V}$, by

$$U \sim U' \quad \Longleftrightarrow \quad \deg_\infty(U) = \deg_\infty(U') \text{ and } |\nabla U|^2 - |\nabla U'|^2 \in L^1(\mathbb{R}^2).$$

Denoting by $[U]$ the corresponding equivalence class of U, we observe that $u_\varepsilon^*(a_i, d_i) \in [U_d]$ for any configuration (a_i, d_i) such that $\sum d_i = d$. Therefore the space $[U_d] + H^1(\mathbb{R}^2)$ contains all H^1 perturbations of reference maps u_ε^* of degree d at infinity.

For a map u in $[U_d] + H^1(\mathbb{R}^2)$, we may now define

$$\mathscr{E}_{\varepsilon,[U_d]}(u) := \lim_{R\to+\infty} \int_{B(R)} \bigl(e_\varepsilon(u) - \tfrac{1}{2}|\nabla U_d|^2\bigr),$$

which is a finite quantity. Moreover, for any solution $u \in C^0(\{U\} + H^1(\mathbb{R}^2))$ with $U \in [U_d]$, we infer from Theorem 1.1 that

$$\frac{d}{dt}\mathscr{E}_{\varepsilon,[U_d]}(u) = \frac{d}{dt}\mathscr{E}_{\varepsilon,U}(u) = -k_\varepsilon \int_{\mathbb{R}^2} |\partial_t u|^2,$$

which means that the renormalized energy has the same dissipation rate as the Ginzburg–Landau energy in the finite energy case $d = 0$.

Statement of the main result. In the sequel, A_n denotes the annulus $B(2^{n+1}) \setminus B(2^n)$ for $n \in \mathbb{N}$, so that $\mathbb{R}^2 = B(2^{n_0}) \cup \bigl(\bigcup_{n\ge n_0} A_n\bigr)$.

Definition 1.2. Let $a_1, \dots, a_l$ be l distinct points in $\mathbb{R}^2$, $d_i \in \{-1, +1\}$ for $i = 1, \dots, l$ and set $d = \sum d_i$. Let $(u_\varepsilon)_{0<\varepsilon<1}$ be a family of maps in $[U_d] + H^1(\mathbb{R}^2)$. We say that $(u_\varepsilon)_{0<\varepsilon<1}$ is *well-prepared* with respect to the configuration (a_i, d_i) if there exist $R = 2^{n_0} > \max|a_i|$ and a constant $K_0 > 0$ such that, with $E_\varepsilon(u, B) \equiv \int_B e_\varepsilon(u)$, the following conditions are satisfied:

$$\lim_{\varepsilon\to 0} \Bigl\| Ju_\varepsilon - \pi \sum_{i=1}^{l} d_i \delta_{a_i} \Bigr\|_{W_0^{1,\infty}(B(R))^*} = 0, \tag{WP$_1$}$$

$$\sup_{0<\varepsilon<1} E_\varepsilon(u_\varepsilon, A_n) \le K_0 \quad \text{for all } n \ge n_0, \tag{WP$_2$}$$

$$\lim_{\varepsilon\to 0} \bigl(\mathscr{E}_{\varepsilon,[U_d]}(u_\varepsilon) - \mathscr{E}_{\varepsilon,[U_d]}(u_\varepsilon^*(a_i, d_i))\bigr) = 0. \tag{WP$_3$}$$

Theorem 1.3. *Let $(u_\varepsilon^0)_{0<\varepsilon<1}$ in $[U_d]+H^1(\mathbb{R}^2)$ be a family of well-prepared initial data with respect to the configuration (a_i^0, d_i) with $d_i = \pm 1$, and let $(u_\varepsilon(t))_{0<\varepsilon<1}$ in $C(\mathbb{R}_+, [U_d]+H^1(\mathbb{R}^2))$ be the corresponding solution of (CGL)$_\varepsilon$. Let $\{a_i(t)\}_{\{i=1,\dots,l\}}$ denote the solution of the ordinary differential equation*

$$\begin{cases} \pi \dot{a}_i(t) = C_i(\delta d_i \mathbb{I}_2 - \mathbb{J}_2)\nabla_{a_i} W, \\ a_i(0) = a_i, \end{cases} \qquad i = 1, \dots, l, \tag{1-5}$$

where $C_i = -d_i/(1+\delta^2)$, $\mathbb{I}_2 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, $\mathbb{J}_2 = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, and W is the Kirchhoff–Onsager functional defined by $W(a_i, d_i) = -\pi \sum_{i\ne j} d_i d_j \log|a_i - a_j|$. Denote by $[0, T^*)$ its maximal interval of existence. Then, for every $t \in [0, T^*)$, the family $(u_\varepsilon(t))_{0<\varepsilon<1}$ is well-prepared with respect to the configuration $(a_i(t), d_i)$.

2. Evolution formula for u_ε

We now recall or derive a number of evolution formulae involving quantities related to u_ε which we introduce now.

Notation. For $x=(x_1,x_2)\in\mathbb{R}^2$, we set $x^\perp=\mathbb{J}_2x=(-x_2,x_1)$, or $x^\perp=ix$ under the standard identification of $\mathbb{R}^2$ with $\mathbb{C}$. For z and $z'\in\mathbb{C}$, we denote by $z\cdot z'=\mathrm{Re}(z\bar{z}')$ the scalar product and $z\times z'=z^\perp\cdot z'=-\mathrm{Im}(z\bar{z}')$ the exterior product of z and z' in $\mathbb{R}^2$. For $\vec{a}:\mathbb{R}^2\to\mathbb{R}^2$, we define $\mathrm{curl}(\vec{a})=\partial_1a_2-\partial_2a_1$. If $u:\mathbb{R}^2\to\mathbb{C}$, we denote by

$$j(u)=u\times\nabla u=iu\cdot\nabla u=u^\perp\cdot\nabla u$$

the linear momentum and by

$$J(u)=\partial_1u\times\partial_2u=\det(\nabla u)$$

the Jacobian of u. For $u\in H^1_{\mathrm{loc}}(\mathbb{R}^2)$, it can be checked that $J(u)=\frac{1}{2}\,\mathrm{curl}\,j(u)$ in the distribution sense. On the set where u does not vanish, we have for $k=1,2$

$$\partial_ku=\partial_ku\cdot\frac{u}{|u|}\frac{u}{|u|}+\partial_ku\cdot\frac{iu}{|u|}\frac{iu}{|u|}.$$

This yields

$$\partial_ku=\partial_k|u|\frac{u}{|u|}+\frac{j_k(u)}{|u|}\frac{u^\perp}{|u|};\tag{2-1}$$

hence

$$\partial_ku\cdot\partial_lu=\partial_k|u|\,\partial_l|u|+\frac{j_k(u)j_l(u)}{|u|^2},\tag{2-2}$$

and it follows that

$$|\nabla u|^2=\big|\nabla|u|\big|^2+\frac{|j(u)|^2}{|u|^2}.\tag{2-3}$$

The Hopf differential of u is defined as

$$\omega(u)=|\partial_1u|^2-|\partial_2u|^2-2i\partial_1u\cdot\partial_2u=4\partial_zu\overline{\partial_{\bar{z}}u}.$$

It follows from (2-2) that $\omega(u)$ may be rewritten in terms of the components of $\nabla|u|$ and $j(u)$ as

$$\omega(u)=\partial_1|u|^2-\partial_2|u|^2-2i\partial_1|u|\,\partial_2|u|+\frac{1}{|u|^2}\big(j_1^2(u)-j_2^2(u)-2ij_1(u)j_2(u)\big).\tag{2-4}$$

We recall that the Ginzburg–Landau energy density is defined by

$$e_\varepsilon(u)=\frac{|\nabla u|^2}{2}+\frac{(1-|u|^2)^2}{4\varepsilon^2}=\frac{|\nabla u|^2}{2}+V(u),$$

and we set

$$\mu_\varepsilon(u)=\frac{e_\varepsilon(u)}{|\log\varepsilon|}.$$

In view of (2-3), we then have

$$e_\varepsilon(u)=e_\varepsilon(|u|)+\frac{|j(u)|^2}{|u|^2}.\tag{2-5}$$

Finally, we write the right-hand side in (CGL)$_\varepsilon$ as

$$\nabla E(u) = \nabla E_\varepsilon(u) = \Delta u + \frac{1}{\varepsilon^2} u(1-|u|^2).$$

Evolution formulae involving the Jacobian and the energy density. For a smooth map u in space-time, direct computations by integration by parts yield for the energy

$$\frac{d}{dt}\int_{\mathbb{R}^2} e_\varepsilon(u)\varphi\,dx = -\int_{\mathbb{R}^2} \partial_t u\cdot\nabla E(u)\varphi\,dx - \int_{\mathbb{R}^2} \nabla\varphi\cdot(\partial_t u\cdot\nabla u)\,dx \tag{2-6}$$

and for the Jacobian

$$\frac{d}{dt}\int_{\mathbb{R}^2} J(u)\chi\,dx = -\int_{\mathbb{R}^2} \nabla^\perp\chi\cdot(\partial_t u^\perp\cdot\nabla u)\,dx, \tag{2-7}$$

where $\chi,\varphi\in\mathfrak{D}(\mathbb{R}^2)$.

At the same time, the Pohozaev identity (see [Bethuel et al. 2005], for example) yields, for any vector field $\vec{X}\in\mathfrak{D}(\mathbb{R}^2,\mathbb{C})$,

$$\int_{\mathbb{R}^2} \vec{X}\cdot(\nabla E(u)\cdot\nabla u)\,dx = -\int_{\mathbb{R}^2} \mathrm{Re}\Big(\omega(u)\frac{\partial\vec{X}}{\partial\bar{z}}\Big)\,dz + \int_{\mathbb{R}^2} V(u)\nabla\cdot\vec{X}\,dx.$$

In particular, the choice of $\vec{X}=\nabla\varphi$, for which $\partial_{\bar z}X = 2\dfrac{\partial^2\varphi}{\partial\bar{z}^2}$, or $\vec{X}=\nabla^\perp\chi$, for which $\partial_{\bar z}X = 2i\dfrac{\partial^2\chi}{\partial\bar{z}^2}$, leads to

$$\int_{\mathbb{R}^2} \nabla\varphi\cdot(\nabla E(u)\cdot\nabla u)\,dx = -2\int_{\mathbb{R}^2} \mathrm{Re}\Big(\omega(u)\frac{\partial^2\varphi}{\partial\bar{z}^2}\Big)\,dz + \int_{\mathbb{R}^2} V(u)\Delta\varphi\,dx$$

and

$$\int_{\mathbb{R}^2} \nabla^\perp\chi\cdot(\nabla E(u)\cdot\nabla u)\,dx = 2\int_{\mathbb{R}^2} \mathrm{Im}\Big(\omega(u)\frac{\partial^2\chi}{\partial\bar{z}^2}\Big)\,dz. \tag{2-8}$$

We next consider a solution u of (CGL)$_\varepsilon$, which is smooth in view of Theorem 1.1. In this case, $\nabla E(u)$ and $\partial_t u$ are related by

$$\partial_t u = \frac{1}{\alpha_\varepsilon}\nabla E(u) = \beta_\varepsilon\nabla E(u), \tag{2-9}$$

where $\alpha_\varepsilon = \dfrac{\delta}{|\log\varepsilon|} + i = k_\varepsilon + i$. Using (2-9) in (2-6) and (2-7), we obtain

$$\frac{d}{dt}\int_{\mathbb{R}^2} e_\varepsilon(u)\varphi\,dx = -k_\varepsilon\int_{\mathbb{R}^2} |\partial_t u|^2\varphi\,dx - \int_{\mathbb{R}^2} \nabla\varphi\cdot(\beta_\varepsilon\nabla E(u)\cdot\nabla u)\,dx,$$

$$\frac{d}{dt}\int_{\mathbb{R}^2} J(u)\chi\,dx = -\int_{\mathbb{R}^2} \nabla^\perp\chi\cdot(i\beta_\varepsilon\nabla E(u)\cdot\nabla u)\,dx.$$

To get rid of the terms of the form $\displaystyle\int_{\mathbb{R}^2}\vec{X}\cdot(i\nabla E(u)\cdot\nabla u)$, we compute

$$\frac{d}{dt}\int_{\mathbb{R}^2} [bJ(u)\chi - ae_\varepsilon(u)\varphi],$$

where $\beta_\varepsilon = a + ib$. This yields

$$\frac{d}{dt}\int_{\mathbb{R}^2}[bJ(u)\chi - ae_\varepsilon(u)]\varphi$$
$$= (b^2+a^2)\int_{\mathbb{R}^2}\nabla^\perp\chi\cdot(\nabla E\cdot\nabla u) + ak_\varepsilon\int_{\mathbb{R}^2}|\partial_t u|^2\,dx + \int_{\mathbb{R}^2}(\nabla\varphi - \nabla^\perp\chi)\cdot(a(a+ib)\nabla E\cdot\nabla u). \tag{2-10}$$

Since $a = k_\varepsilon/(k_\varepsilon^2+1)$ and $b = -1/(k_\varepsilon^2+1)$, we finally infer from this relation and (2-8) the following:

Proposition 2.1. *Let u solve $(\mathrm{CGL})_\varepsilon$. Then for all φ, $\chi \in \mathfrak{D}(\mathbb{R}^2)$,*

$$\frac{d}{dt}\int_{\mathbb{R}^2}[J(u)\chi + k_\varepsilon e_\varepsilon(u)\varphi] = -k_\varepsilon^2\int_{\mathbb{R}^2}|\partial_t u|^2\varphi - 2\int_{\mathbb{R}^2}\mathrm{Im}\left(\omega(u)\frac{\partial^2\chi}{\partial\bar{z}^2}\right) + R_\varepsilon(t,\varphi,\chi,u),$$

where the remainder R_ε is defined by either of the equivalent relations

$$R_\varepsilon(t,\varphi,\chi,u) = -k_\varepsilon\int_{\mathbb{R}^2}(\nabla\varphi - \nabla^\perp\chi)\cdot(\beta_\varepsilon\nabla E(u)\cdot\nabla u),$$
$$R_\varepsilon(t,\varphi,\chi,u) = -k_\varepsilon\int_{\mathbb{R}^2}(\nabla\varphi - \nabla^\perp\chi)\cdot(\partial_t u\cdot\nabla u).$$

Proposition 2.1 allows us to derive formally the motion law for the vortices. Indeed, assume that we have

$$Ju_\varepsilon(t) \to \pi\sum_{i=1}^{l} d_i\delta_{a_i(t)}, \qquad \mu_\varepsilon(u_\varepsilon)(t) \to \pi\sum_{i=1}^{l}\delta_{a_i(t)},$$

and $u_\varepsilon(t)$ is close in some sense to $u_\varepsilon^*(a_i(t), d_i)$ and therefore to $u^*(a_i(t), d_i)$, where

$$u^*(a_i, d_i) = \prod_{i=1}^{l}\left(\frac{z-a_i}{|z-a_i|}\right)^{d_i}.$$

We use Proposition 2.1 with u formally replaced by $u^*(a_i(t), d_i)$ and with choices of test functions φ and χ which are localized and affine near each point $a_i(t)$ and satisfy $\nabla\varphi = \nabla^\perp\chi$ there, so that both terms $k_\varepsilon^2\int_{\mathbb{R}^2}|\partial_t u|^2\varphi$ and $R_\varepsilon(t,\varphi,\chi,u_\varepsilon)$ vanish in the limit $\varepsilon \to 0$. Using the formula

$$2\int_{\mathbb{R}^2}\mathrm{Im}\left(\omega\big(u^*(a_i,d_i)\big)\frac{\partial^2\chi}{\partial\bar{z}^2}\right) = 2\pi\sum_{j\neq i} d_i d_j\frac{(a_i-a_j)^\perp}{|a_i-a_j|^2}\cdot\nabla\chi(a_i)$$

from [Bethuel et al. 2005, (7.2)], we then obtain that for each i

$$\pi d_i\dot{a}_i(t)\cdot\nabla\chi(a_i) + \delta\pi\dot{a}_i(t)\cdot\nabla\varphi(a_i) = -2\pi\sum_{j:j\neq i} d_i d_j\frac{(a_i-a_j)^\perp}{|a_i-a_j|^2}\cdot\nabla\chi(a_i).$$

Taking into account that $\nabla\varphi(a_i) = \nabla^\perp\chi(a_i)$, we infer that

$$\pi\left(d_i\dot{a}_i(t) - \delta\dot{a}_i^\perp(t)\right)\cdot\nabla\chi(a_i) = -2\pi\sum_{j\neq i} d_i d_j\frac{(a_i-a_j)^\perp}{|a_i-a_j|^2}\cdot\nabla\chi(a_i),$$

which yields the ODE (1-5).

In Sections 4 and 5, in order to give a rigorous meaning to the previous computations, we will prove the convergence of the Jacobians and of the energy densities to the weighted sums of Dirac masses mentioned above, and then show that both the energy dissipation $k_\varepsilon^2 \int_{\mathbb{R}^2} |\partial_t u_\varepsilon|^2$ and the remainder $R_\varepsilon(t, \varphi, \chi, u_\varepsilon)$ vanish when ε tends to zero. In Section 6, we will establish some asymptotic control of $\omega(u_\varepsilon) - \omega(u^*(a_i), d_i)$ away from the vortices in terms of the excess energy $\mathscr{E}_{\varepsilon,[U_d]}(u_\varepsilon) - \mathscr{E}_{\varepsilon,[U_d]}(u_\varepsilon^*(a_i))$, and finally prove that this excess energy converges to zero by mean of a Grönwall inequality.

3. Some results on the renormalized energy

Degree and energy at infinity. In this paragraph, we collect some results from [Bethuel et al. 2008] related to the energy at infinity, which require the notion of degree at infinity.

Let A be the annulus $B(2) \setminus B(1)$. We define

$$T_d = \left\{u \in H^1(A) : \text{some } B \subset B(u) \text{ satisfies } |B| \geq \tfrac{3}{4} \text{ and } \deg(u, \partial B(r)) = d \text{ for all } r \in B\right\},$$

where $B(u)$ is the set of radii $r \in [1, 2]$ such that the restriction $u_{|\partial B(r)}$ is continuous and does not vanish, and we define the sublevel sets

$$E_\varepsilon^\Lambda = \{u \in H^1(A) : E_\varepsilon(u, A) < \Lambda\}.$$

The topological sector of degree d is then defined as

$$S_{d,\varepsilon}^\Lambda = E_\varepsilon^\Lambda \cap T_d.$$

Theorem 3.1 [Almeida 1999]. *For all $\Lambda > 0$, there exists $\varepsilon_\Lambda > 0$ such that for every $0 < \varepsilon < \varepsilon_\Lambda$, we have*

$$E_\varepsilon^\Lambda = \bigcup_{d \in \mathbb{Z}} S_{d,\varepsilon}^\Lambda.$$

The map $\deg : E_\varepsilon^\Lambda \to \mathbb{Z}$, $u \in S_{d,\varepsilon}^\Lambda \mapsto d$ is continuous.

For the rest of this section, we fix $\Lambda > \Lambda_d = 2\pi d^2 \log 2$ and we set

$$S_d \equiv S_{d,\varepsilon_\Lambda}^\Lambda,$$

so in particular the map U_d belongs to S_d, since $|U_d| \equiv 1$ on A and $\int_A \frac{1}{2}|\nabla U_d|^2 = \pi d^2 \log 2$.

One easily infers from Theorem 3.1 that if $u \in [U_d] + H^1(\mathbb{R}^2)$, then for any sufficiently large k the map $u(2^k \cdot)$ belongs to some $S_{d(k)}$. In fact, one can find a radius from which $d(k) \equiv d$, that is, u has well defined and constant degree d at infinity.

Proposition 3.2 [Bethuel et al. 2008]. *Let $d \in \mathbb{Z}$, $\Lambda > \Lambda_d$ and $u \in [U_d] + H^1(\mathbb{R}^2)$. There exists an integer $n \in \mathbb{N}^*$ such that for any $k \geq n$, the map $u_k : z \in A \mapsto u(2^k z)$ belongs to the topological sector S_d. We denote by $n(u)$ the smallest integer with this property.*

For maps $u \in [U_d] + H^1(\mathbb{R}^2)$ satisfying in addition a uniform energy bound on large annuli one can characterize $n(u)$ as follows (see, for example, the proof of Lemma 7.1 in [Bethuel et al. 2008]).

Lemma 3.3. *Let $\Lambda > \Lambda_d$ be given and $0 < \varepsilon < \varepsilon_\Lambda$. Let $u \in [U_d] + H^1(\mathbb{R}^2)$ and assume that there exists some $n_0 \in \mathbb{N}^*$ such that $E_\varepsilon(u, A_n) < \Lambda$ for all $n \geq n_0$. Then $n(u) \leq n_0$.*

The next lemma provides a lower bound for the energy on large annuli.

Lemma 3.4 [Bethuel et al. 2008]. *Let $d \in \mathbb{Z}$ and $u \in [U_d] + H^1(\mathbb{R}^2)$. Then, for any $k \geq n(u)$, we have for $0 < \varepsilon < \varepsilon_\Lambda$*

$$\int_{A_k} \left(e_\varepsilon(u) - \tfrac{1}{2}|\nabla U_d|^2\right) \geq -C2^{-2k}\varepsilon^2.$$

One can then derive from Lemma 3.4 an upper bound for $E_\varepsilon(u, B) - E_\varepsilon(u_\varepsilon^*, B)$ on large balls B in terms of the excess energy $\mathscr{E}_{\varepsilon,[U_d]}(u) - \mathscr{E}_{\varepsilon,[U_d]}(u_\varepsilon^*)$. We will therefore be able to rely on properties of the Ginzburg–Landau energy on bounded domains in the course of the proof of Theorem 1.3.

Lemma 3.5 [Bethuel et al. 2008]. *Let $d \in \mathbb{Z}$, $u \in [U_d] + H^1(\mathbb{R}^2)$, $a_1, \dots, a_l \in \mathbb{R}^2$ and $d_1, \dots, d_l \in \mathbb{Z}^*$ such that $d = \sum d_i$. Let $k \geq 1 + \max\{\log_2 |a_1|, \dots, \log_2 |a_l|, n(u)\}$ and $R = 2^k$. Then, we have*

$$\int_{B(R)} \left[e_\varepsilon(u) - e_\varepsilon(u_\varepsilon^*(a_i, d_i))\right] \leq \mathscr{E}_{\varepsilon,[U_d]}(u) - \mathscr{E}_{\varepsilon,[U_d]}(u_\varepsilon^*(a_i, d_i)) + \frac{C}{R},$$

where C depends only on l and d.

Explicit identities for the reference map u_ε^*. We present here an account of some classical identities for the energy of u_ε^*, borrowed from [Bethuel et al. 2008].

We consider a configuration (a_i, d_i) with $d_i \in \mathbb{Z}^*$ and we set $d = \sum d_i$. We begin with an explicit expansion near each vortex a_j.

Lemma 3.6. *For $j \in \{1, \dots, l\}$ and $0 < \varepsilon < 1$,*

$$\int_{B(a_j,r)} e_\varepsilon(u_\varepsilon^*(a_i, d_i)) = \pi d_j^2 \log\left(\frac{r}{\varepsilon}\right) + \gamma(|d_j|) + O\left(\frac{r}{r_a}\right)^2 + O\left(\frac{\varepsilon}{r}\right)^2,$$

where $\gamma(|d_j|)$ is some universal constant.

On the other hand, $u_\varepsilon^*(a_i, d_i)$ behaves as $u^*(a_i, d_i)$ away from the vortices, so its energy on $\Omega_{R,r} = B(R) \setminus \bigcup_{j=1}^l B(a_j, r)$ is close to the energy of $u^*(a_i, d_i)$ on $\Omega_{R,r}$ which we can compute explicitly [Bethuel et al. 1994]. Combining the previous expansions, we obtain:

Proposition 3.7. *Let*

$$r_a = \tfrac{1}{8} \min_{i \neq j}\{|a_i - a_j|\}, \qquad R_a = \max\{|a_i|\}.$$

Then for $R > R_a + 1$, we have as $\varepsilon \to 0$

$$\int_{B(R)} e_\varepsilon(u_\varepsilon^*(a_i, d_i)) = \pi \sum_{i=1}^l d_i^2 |\log \varepsilon| + W(a_i, d_i) + \sum_{i=1}^l \gamma(|d_i|) + \pi d^2 \log R + O\left(\frac{R_a}{R}\right) + o_\varepsilon(1).$$

Observe that $\pi d^2 \log R = \int_{B(R)\setminus B(1)} \frac{1}{2}|\nabla U_d|^2$. This yields an expansion for the renormalized energy:

Corollary 3.8. *When $\varepsilon \to 0$,*

$$\mathscr{E}_{\varepsilon,[U_d]}(u_\varepsilon^*(a_i, d_i)) = \pi \sum_{i=1}^l d_i^2 |\log \varepsilon| + W(a_i, d_i) + \sum_{i=1}^l \gamma(|d_i|) - \int_{B(1)} \frac{|\nabla U_d|^2}{2} + o_\varepsilon(1).$$

Concerning the energy on annuli, we finally quote the following:

Lemma 3.9. *For $R > R_a$, we have*

$$\int_{B(2R)\setminus B(R)} e_\varepsilon(u^*_\varepsilon(a_i,d_i)) = \pi d^2 \log 2 + O\Big(\frac{R_a}{R}\Big)$$

or, in view of the properties of U_d at infinity,

$$\int_{B(2R)\setminus B(R)} e_\varepsilon(u^*_\varepsilon(a_i,d_i)) = \int_{B(2R)\setminus B(R)} \frac{|\nabla U_d|^2}{2} + O\Big(\frac{R_a}{R}\Big).$$

4. Coercivity for the renormalized energy

In this section, we supplement some results from [Bethuel et al. 2008] and [Jerrard and Spirn 2007] with estimates to be used later. These results establish precise estimates in various norms for maps u being close to $u^*_\varepsilon(a_i,d_i)$ in terms of the excess energy with respect to the configuration (a_i,d_i). For a map $u \in [U_d]+H^1(\mathbb{R}^2)$ and a given configuration (a_i,d_i) with $d_i = \pm 1$, we define this excess energy Σ_ε as

$$\Sigma_\varepsilon = \Sigma_\varepsilon(a_i,d_i) = \mathscr{E}_{\varepsilon,[U_d]}(u) - \mathscr{E}_{\varepsilon,[U_d]}(u^*_\varepsilon(a_i,d_i)).$$

We also set

$$r_a = \tfrac{1}{8}\min_{i\neq j}\{|a_i - a_j|\}, \qquad R_a = \max_{i=1,\dots,l}\{|a_i|\}.$$

Theorem 4.1. *Let $r \le r_a$ and let $2^{n_0} = R_0 > R_a$ be such that $\bigcup_{i=1}^l B(a_i,r) \subset B(R_0)$. There exist ε_0 and η_0, depending only on l, r, r_a, R_a, and R_0, such that for all $u \in [U_d]+H^1(\mathbb{R}^2)$ satisfying*

$$\eta = \|Ju - \pi\sum_{i=1}^l d_i\delta_{a_i}\|_{W_0^{1,\infty}(B(R_0))^*} \le \eta_0 \quad \textit{and} \quad 2^{n(u)} \le R_0, \tag{4-1}$$

we have

$$\int_{B(R_0)\setminus\cup B(a_i,r)} e_\varepsilon(|u|) + \frac{1}{8}\left|\frac{j(u)}{|u|} - j(u^*(a_i,d_i))\right|^2 \le \Sigma_\varepsilon + C\Big(\eta,\varepsilon,\frac{1}{R_0}\Big) \quad \textit{for } \varepsilon \le \varepsilon_0, \tag{4-2}$$

where C is a continuous function on $\mathbb{R}^3$ that vanishes at the origin. Furthermore, there exist points $b_i \in B(a_i,r/2)$ such that, for some continuous functions f on $\mathbb{R}^2$ and g on $\mathbb{R}^4$, we have

$$\Big\|Ju - \pi\sum_{i=1}^l d_i\delta_{b_i}\Big\|_{W_0^{1,\infty}(B(R_0))^*} \le f(R_0,\Sigma_\varepsilon)\varepsilon|\log\varepsilon|, \tag{4-3}$$

$$\Big\|\mu_\varepsilon(u) - \pi\sum_{i=1}^l \delta_{b_i}\Big\|_{W_0^{1,\infty}(B(R_0))^*} \le \frac{g(R_0,r,r_a,\Sigma_\varepsilon)}{|\log\varepsilon|}. \tag{4-4}$$

Proof. Except for the energy concentration (4-4), each of the statements is proved in [Bethuel et al. 2008, Theorem 6.1]. We first infer from (4-1) that

$$\|Ju - \pi d_i\delta_{a_i}\|_{W_0^{1,\infty}(B(a_i,r))^*} \le \eta_0 \quad \text{for all } i.$$

If η_0 is small enough with respect to r this gives in view of [Jerrard and Spirn 2007, Theorem 3] that $K_0^i \ge C(r)$, where K_0^i is the local excess energy near the vortex i defined by $K_0^i = \int_{B(a_i,r)} e_\varepsilon(u) - \pi\log\big(r/\varepsilon\big)$.

It follows that

$$\int_{B(a_i,r)} e_\varepsilon(u) \le \int_{B(R_0)} e_\varepsilon(u) - \pi(l-1)|\log\varepsilon| - C(r).$$

At the same time, since $n(u) \le n_0$, we have according to Lemma 3.5 and Proposition 3.7

$$\int_{B(R_0)} e_\varepsilon(u) \le \int_{B(R_0)} e_\varepsilon(u_\varepsilon^*(a_i,d_i)) + \Sigma_\varepsilon + \frac{C}{R_0} \le \pi l|\log\varepsilon| + \Sigma_\varepsilon + C.$$

This first implies that $K_0^i \le C + \Sigma_\varepsilon$. Also, replacing r by $3r/4$ we see that

$$\int_{B(R_0)\setminus\cup B(a_i,3r/4)} \mu_\varepsilon(u) \le \frac{C+\Sigma_\varepsilon}{|\log\varepsilon|},$$

where C only depends on R_0, r, r_a, R_a.

Now, according to [Jerrard and Spirn 2007, Theorem 2′], the energy density $\mu_\varepsilon(u)$ on $B(a_i,r)$ is concentrated at the point $b_i \in B(a_i,r/2)$ where $J(u)$ concentrates. From [Colliander and Jerrard 1999, Theorem 3.2.1] and the estimate for K_0^i it follows that

$$\|\mu_\varepsilon(u) - \pi\delta_{b_i}\|_{W_0^{1,\infty}(B(a_i,r))^*} \le \frac{f(\Sigma_\varepsilon, C)}{|\log\varepsilon|}.$$

Combining this and the upper bound for the energy density outside the vortex balls yields (4-4). □

5. Convergence to Lipschitz vortex paths

In this section, we establish compactness for the Jacobians and the energy densities in a more general situation, replacing assumption (WP$_3$) in Theorem 1.3 by a uniform bound on the initial excess energy.

Theorem 5.1. *Let (a_i^0, d_i) with $d_i = \pm 1$ be a configuration of vortices. Let $R = 2^{n_0}$ and $(u_\varepsilon^0)_{0<\varepsilon<1}$ in $[U_d] + H^1(\mathbb{R}^2)$ such that*

$$\lim_{\varepsilon\to 0}\Big\| Ju_\varepsilon^0 - \pi\sum_{i=1}^{l} d_i\delta_{a_i^0}\Big\|_{W_0^{1,\infty}(B(R))^*} = 0, \tag{WP$_1$}$$

$$\sup_{0<\varepsilon<1} E_\varepsilon(u_\varepsilon^0, A_n) \le K_0 \quad \text{for all } n \ge n_0, \tag{WP$_2$}$$

$$\sup_{0<\varepsilon<1}\big(\mathscr{E}_{\varepsilon,[U_d]}(u_\varepsilon^0) - \mathscr{E}_{\varepsilon,[U_d]}(u_\varepsilon^*(a_i^0,d_i))\big) \le K_1. \tag{WP$_{3'}$}$$

Then there exist $R' = 2^{n_1}$ and $T > 0$ depending only on K_1, R, r_{a^0} and R_{a^0}, a sequence $\varepsilon_k \to 0$ and l Lipschitz paths $b_i : [0,T] \to \mathbb{R}^2$ starting from a_i^0 such that

$$\sup_{t\in[0,T]}\Big\| Ju_{\varepsilon_k}(t) - \pi\sum_{i=1}^{l} d_i\delta_{b_i(t)}\Big\|_{W_0^{1,\infty}(B(R'))^*} \to 0, \qquad k\to+\infty, \tag{5-1}$$

$$\sup_{t\in[0,T]}\Big\| \mu_{\varepsilon_k}(u_{\varepsilon_k})(t) - \pi\sum_{i=1}^{l}\delta_{b_i(t)}\Big\|_{W^{1,\infty}(B(R'))^*} \to 0, \qquad k\to+\infty. \tag{5-2}$$

Moreover, there exist a constant $C_0 > 0$ depending only on r_{a^0}, R, K_1 and K_0 and a constant $C_1 > 0$ depending on r_{a^0}, R and K_1 such that for all $t \in [0, T]$ and $k \in \mathbb{N}$, we have

$$E_{\varepsilon_k}(u_{\varepsilon_k}(t), A_n) \le C_0 \quad \textit{for all } n \ge n_1, \tag{5-3}$$

$$\mathscr{E}_{\varepsilon_k,[U_d]}(u_{\varepsilon_k}(t)) - \mathscr{E}_{\varepsilon_k,[U_d]}(u^*_{\varepsilon_k}(b_i(t), d_i)) \le C_1. \tag{5-4}$$

Proof. The proof is very similar to that of [Bethuel et al. 2008, Theorem 4]. In the sequel, C will be a constant depending only on r_{a^0}, R, R_{a^0}, and K_1. To simplify the notations further we will set $r_a = r_{a^0}$ and $R_a = R_{a^0}$.

We first consider $\Lambda > \max(K_0, \Lambda_d)$. Thanks to Lemma 3.3 and (WP$_2$), there exists $\varepsilon_\Lambda > 0$ such that $n(u^0_\varepsilon) \le n_0$ for all $0 < \varepsilon < \varepsilon_\Lambda$. We fix such a Λ and from now on only consider $0 < \varepsilon < \varepsilon_\Lambda$.

We next introduce the smallest integer $n_1 \ge n_0$ such that $2^{n_1} \ge \max(R, R_a + r_a)$ and define $R' = 2^{n_1}$. In the remainder of the proof, we write $\|\cdot\|$ instead of $\|\cdot\|_{W^{1,\infty}_0(B(R'))^*}$. Our aim is to apply Theorem 4.1 to each $u_\varepsilon(t)$ for the choice $r = r_a$ and $R_0 = R'$. Let η_0 and ε_0 be the constants provided by Theorem 4.1 for this choice. First, thanks to (WP$_2$) and (WP$_{3'}$), the convergence in (WP$_1$) still holds on the larger ball $B(R')$ (see the proof of Lemma 7.3 in [Bethuel et al. 2008]). Therefore, since $t \mapsto Ju_\varepsilon(t) \in L^1(B(R'))$ is continuous for each ε, there exists a time $T_\varepsilon > 0$ such that

$$\Big\| Ju_\varepsilon(s) - \pi \sum_{i=1}^{l} d_i \delta_{a^0_i} \Big\| < \eta_0, \qquad \forall s \in [0, T_\varepsilon).$$

We take T_ε as the maximum time smaller than T^* having this property, where T^* is as in Theorem 1.3.

Meanwhile, since $t \mapsto E_\varepsilon(u_\varepsilon(t), A_n)$ is uniformly continuous with respect to n and $\Lambda > K_0$, we infer from (WP$_2$) that there exists $T'_\varepsilon > 0$ such that for $s \in [0, T'_\varepsilon]$

$$E_\varepsilon(u_\varepsilon(s), A_n) < \Lambda \quad \text{for all } n \ge n_1,$$

so according to Lemma 3.3 we have $n(u_\varepsilon(s)) \le n_1$ for $s \in [0, T'_\varepsilon]$. We take $T'_\varepsilon \le T^*$ maximal with this property.

We claim that there exists a constant D depending on K_1, r_a, R, and K_0 such that for all $s \in [0, \min(T_\varepsilon, T'_\varepsilon))$,

$$E_\varepsilon(u_\varepsilon(s), A_n) \le D \quad \text{for all } n \ge n_1. \tag{5-5}$$

Consequently, if we assume from the beginning that $\Lambda > \max(K_0, \Lambda_d, D)$, then $T'_\varepsilon \ge T_\varepsilon$, and it follows from Lemma 3.3 that $n(u_\varepsilon(s)) \le n_1$ on $[0, T_\varepsilon]$.

Proof of (5-5). As in [Bethuel et al. 2008], we decompose each $E_\varepsilon(u_\varepsilon(s), A_n) - E_\varepsilon(u^*_\varepsilon(a^0_i, d_i), A_n)$, for $n \ge n_1$, as

$$\sum_{\substack{k=n_1\\k\neq n}}^{+\infty} \big(E_\varepsilon(u^*_\varepsilon(a^0_i, d_i), A_k) - E_\varepsilon(u_\varepsilon(t), A_k)\big) \\ + E_\varepsilon(u^*_\varepsilon(a^0_i, d_i), B(R')) - E_\varepsilon(u_\varepsilon(s), B(R')) + \mathscr{E}_{\varepsilon,[U_d]}(u_\varepsilon(s)) - \mathscr{E}_{\varepsilon,[U_d]}(u^*_\varepsilon(a^0_i, d_i)).$$

We first handle each term of the sum in the right. In view of Lemmas 3.4 and 3.9, we have for $k \ge n_1$

$$E_\varepsilon(u_\varepsilon(s), A_k) \ge -C\varepsilon^2 2^{-2k} + \int_{A_k} \frac{|\nabla U_d|^2}{2} \ge E_\varepsilon(u^*_\varepsilon(a^0_i, d_i), A_k) - C(R_a)2^{-k} - C\varepsilon^2 2^{-2k},$$

so we deduce that

$$\sum_{\substack{k=n_1\\ k\neq n}}^{+\infty}\big(E_\varepsilon(u^*_\varepsilon(a_i^0,d_i),A_k)-E_\varepsilon(u_\varepsilon(s),A_k)\big)\le C.$$

Next, we infer from the definition of T_ε and [Jerrard and Spirn 2007, Theorem 3] that

$$\int_{B(a_i^0,r_a)} e_\varepsilon(u_\varepsilon(s))\ge \pi\,|\log\varepsilon|-C.$$

Observe that R' is chosen so that $\cup B(a_i^0,r_a)\subset B(R')$, so this leads to

$$E_\varepsilon(u_\varepsilon(s),B(R'))\ge \pi\, l|\log\varepsilon|-C.$$

Using Proposition 3.7, we thus find

$$E_\varepsilon(u^*_\varepsilon(a_i^0,d_i),B(R'))-E_\varepsilon(u_\varepsilon(s),B(R'))\le C. \tag{5-6}$$

Finally, we define $\Sigma_\varepsilon^0(s):=\mathscr{E}_{\varepsilon,[U_d]}(u_\varepsilon(s))-\mathscr{E}_{\varepsilon,[U_d]}(u^*_\varepsilon(a_i^0,d_i))$. Since $\mathscr{E}_{\varepsilon,[U_d]}(u_\varepsilon(t))$ is nonincreasing, ($\text{WP}_{3'}$) yields $\Sigma_\varepsilon^0(s)\le\mathscr{E}_{\varepsilon,[U_d]}(u_\varepsilon^0)-\mathscr{E}_{\varepsilon,[U_d]}(u^*_\varepsilon(a_i^0,d_i))\le K_1$, and (5-5) follows. □

We can now apply Theorem 4.1 to each $u_\varepsilon(t)$ on $[0,T_\varepsilon]$. This provides points $b_i^\varepsilon(s)\in B\big(a_i^0,r_a/2\big)$ for $0\le s\le T_\varepsilon$. Since $\Sigma_\varepsilon^0(s)\le K_1$, the estimate (4-2) turns into

$$\int_{\Omega_{R',r_a}} e_\varepsilon(|u_\varepsilon(s)|)+\frac18\left|\frac{j(u_\varepsilon(s))}{|u_\varepsilon(s)|}-j(u^*(a_i^0,d_i))\right|^2\le C,$$

where $\Omega_{R',r_a}=B(R')\setminus\bigcup B(a_i^0,r_a)$. Also, we have by (2-4) and (2-5)

$$\int_{\Omega_{R',r_a}} e_\varepsilon(u_\varepsilon(s))\le C \tag{5-7}$$

and

$$\|\omega(u_\varepsilon(s))\|_{L^1(\Omega_{R',r_a})}\le C, \tag{5-8}$$

where $C=C(R,r_a,K_1)$. For convenience, we will now write μ_ε instead of $\mu_\varepsilon(u_\varepsilon)$.

Given any configuration (a_i,d_i), we denote by $\mathscr{K}(a_i)$ the set of functions $\chi,\varphi\in\mathfrak{D}(\mathbb{R}^2)$ such that

$$\chi=\sum_{i=1}^{l}\chi_i,\qquad \varphi=\sum_{i=1}^{l}\varphi_i,$$

where for all i

$$\chi_i,\varphi_i\in\mathfrak{D}\Big(B\Big(a_i,\frac{3r_a}{2}\Big)\Big),\qquad \nabla\varphi_i=\nabla^\perp\chi_i \text{ on } B(a_i,r_a),$$

and χ_i (hence φ_i) is affine on $B(a_i,r_a)$ with $|\nabla\chi_i(a_i)|=|\nabla\varphi_i(a_i)|\le 1$.

By definition of r_a such functions χ and φ always exist, and we can moreover estimate their L^∞ norms by

$$\|D\varphi\|_\infty,\|D\chi\|_\infty\le\frac{C}{r_a},\qquad \|D^2\varphi\|_\infty,\|D^2\chi\|_\infty\le\frac{C}{r_a^2}.$$

We next control the remainder terms appearing in Proposition 2.1.

Lemma 5.2. *There exists a constant $C = C(r_a, R, K_1, T^*)$ such that*

$$\int_0^{T_\varepsilon}\int_{\mathbb{R}^2} \frac{|\partial_t u_\varepsilon|^2}{|\log\varepsilon|^2}\,ds \le \frac{C}{|\log\varepsilon|}$$

and for all $\chi, \varphi \in \mathcal{H}(a_i^0)$

$$\left|\int_0^{T_\varepsilon}\int_{\mathbb{R}^2} (\nabla^\perp\chi - \nabla\varphi)\cdot\frac{\partial_t u_\varepsilon\cdot\nabla u_\varepsilon}{|\log\varepsilon|}\,ds\right| \le \frac{C}{|\log\varepsilon|^{1/2}}.$$

Proof. To prove the first inequality, we use Theorem 1.1 and obtain

$$\frac{\delta}{|\log\varepsilon|}\int_0^{T_\varepsilon}\int_{\mathbb{R}^2}|\partial_t u_\varepsilon|^2 = \mathscr{E}_{\varepsilon,[U_d]}(u_\varepsilon^0) - \mathscr{E}_{\varepsilon,[U_d]}(u_\varepsilon(T_\varepsilon)) \le K_1 + \mathscr{E}_{\varepsilon,[U_d]}(u_\varepsilon^*(a_i^0,d_i)) - \mathscr{E}_{\varepsilon,[U_d]}(u_\varepsilon(T_\varepsilon)).$$

Since $n(u_\varepsilon(T_\varepsilon)) \le n_1$ we have by Lemma 3.5

$$\mathscr{E}_{\varepsilon,[U_d]}(u_\varepsilon^*(a_i^0,d_i)) - \mathscr{E}_{\varepsilon,[U_d]}(u_\varepsilon(T_\varepsilon)) \le \int_{B(R')} e_\varepsilon(u_\varepsilon^*(a_i^0,d_i)) - \int_{B(R')} e_\varepsilon(u_\varepsilon(T_\varepsilon)) + \frac{C}{R'},$$

which is bounded in view of (5-6). It then suffices to divide all terms by $|\log\varepsilon|$.

For the second assertion, we set $\xi = \nabla^\perp\chi - \nabla\varphi$, which has compact support in $A = \bigcup A_i$, where $A_i = B(a_i^0, 3r_a/2)\setminus B(a_i^0, r_a)$, and we apply the Cauchy–Schwarz inequality. We obtain

$$\left(\int_0^{T_\varepsilon}\int_{\mathbb{R}^2}(\nabla^\perp\chi - \nabla\varphi)\cdot\frac{\partial_t u_\varepsilon\cdot\nabla u_\varepsilon}{|\log\varepsilon|}\right)^2 \le \left(\int_0^{T_\varepsilon}\int_{\mathbb{R}^2}\frac{|\partial_t u_\varepsilon|^2}{|\log\varepsilon|^2}\right)\cdot\left(\int_0^{T_\varepsilon}\int_A |\nabla u_\varepsilon|^2|\xi|^2\right).$$

Since $A \subset \Omega_{R',r_a}$ and $\sup_{0<\varepsilon<1} T_\varepsilon \le T^*$, we infer from (5-7) that

$$\int_0^{T_\varepsilon}\int_A |\nabla u_\varepsilon|^2|\xi|^2 \le \|\xi\|_\infty^2 \int_0^{T_\varepsilon}\int_A |\nabla u_\varepsilon|^2 \le CT^*\|\xi\|_\infty^2,$$

and the conclusion finally follows from the first part of the proof. □

Lemma 5.3. *There exists $T = T(r_a, R_a, R, K_1) > 0$ such that*

$$\liminf_{\varepsilon\to 0} T_\varepsilon \ge T.$$

Proof. We first show that for $(\chi,\varphi) \in \mathcal{H}(a_i^0)$, for $s, t \in [0, T_\varepsilon]$ and $i = 1, \dots, l$ we have

$$\big|\langle\chi_i, Ju_\varepsilon(t) - Ju_\varepsilon(s)\rangle + \delta\langle\varphi_i, \mu_\varepsilon(t) - \mu_\varepsilon(s)\rangle\big| \le C|t-s| + \frac{C}{|\log\varepsilon|^{1/2}}. \tag{5-9}$$

Indeed, we fix i and we invoke Proposition 2.1 for $u \equiv u_\varepsilon$ and the choice of test functions (χ_i, φ_i). Integrating the formula in that proposition over $[s, t]$ yields

$$\begin{aligned}&\big|\langle\chi_i, Ju_\varepsilon(t) - Ju_\varepsilon(s)\rangle + \delta\langle\varphi_i, \mu_\varepsilon(t) - \mu_\varepsilon(s)\rangle\big| \\ &\qquad \le 2\int_s^t\int\left|\operatorname{Im}\Big(\omega(u_\varepsilon)\frac{\partial^2\chi_i}{\partial\bar z^2}\Big)\right| + \int_s^t\int\left|\frac{|\partial_t u_\varepsilon|^2}{|\log\varepsilon|^2}\varphi_i + (\nabla^\perp\chi_i - \nabla\varphi_i)\cdot\frac{\partial_t u_\varepsilon\cdot\nabla u_\varepsilon}{|\log\varepsilon|}\right|,\end{aligned}$$

where $\partial^2\chi_i/\partial\bar z^2$ has support in $C_i \subset \Omega_{R',r_a}$, and it finally suffices to use (5-8) and Lemma 5.2.

In a second step, we take advantage of the equality $\left\| J u_\varepsilon(T_\varepsilon) - \pi \sum_{i=1}^{l} d_i \delta_{a_i^0} \right\| \equiv \eta_0$. We set

$$\nu_{i,\varepsilon} = d_i \frac{b_i^\varepsilon(T_\varepsilon) - a_i^0}{|b_i^\varepsilon(T_\varepsilon) - a_i^0|}, \qquad i = 1, \ldots, l$$

and we define $\chi_{i,\varepsilon}$, $\varphi_{i,\varepsilon}$ so that for $x \in B(a_i^0, r_a)$,

$$\chi_{i,\varepsilon}(x) = \nu_{i,\varepsilon} \cdot x, \qquad \varphi_{i,\varepsilon}(x) = \nu_{i,\varepsilon}^\perp \cdot x.$$

We require additionally that $\chi = \sum \chi_{i,\varepsilon}$ and $\varphi = \sum \varphi_{i,\varepsilon}$ belong to $\mathcal{K}(a_i^0)$; we can moreover choose $\varphi_{i,\varepsilon}$ and $\chi_{i,\varepsilon}$ so that their norms in $C^2(B(R))$ remain bounded uniformly in ε. Since $b_i^\varepsilon(T_\varepsilon) \in B(a_i^0, r_a/2)$, we have $|d_i| \left| b_i^\varepsilon(T_\varepsilon) - a_i^0 \right| = d_i \chi\left(b_i^\varepsilon(T_\varepsilon) - a_i^0\right) + \delta \varphi\left(b_i^\varepsilon(T_\varepsilon) - a_i^0\right)$, so

$$\left\| \pi \sum_{i=1}^{l} d_i (\delta_{b_i^\varepsilon(T_\varepsilon)} - \delta_{a_i^0}) \right\| = \left\langle \pi \sum_{i=1}^{l} d_i (\delta_{b_i^\varepsilon(T_\varepsilon)} - \delta_{a_i^0}), \chi \right\rangle + \delta \left\langle \pi \sum_{i=1}^{l} (\delta_{b_i^\varepsilon(T_\varepsilon)} - \delta_{a_i^0}), \varphi \right\rangle.$$

On the other hand, we have

$$\left\| J u_\varepsilon(T_\varepsilon) - \pi \sum_{i=1}^{l} d_i \delta_{a_i^0} \right\| \leq \left\| J u_\varepsilon(T_\varepsilon) - \pi \sum_{i=1}^{l} d_i \delta_{b_i^\varepsilon(T_\varepsilon)} \right\| + \left\| \pi \sum_{i=1}^{l} d_i (\delta_{b_i^\varepsilon(T_\varepsilon)} - \delta_{a_i^0}) \right\|.$$

The second term in the right-hand side may be rewritten as

$$\left\langle \pi \sum_{i=1}^{l} d_i (\delta_{b_i^\varepsilon(T_\varepsilon)} - \delta_{a_i^0}), \chi \right\rangle + \delta \left\langle \pi \sum_{i=1}^{l} (\delta_{b_i^\varepsilon(T_\varepsilon)} - \delta_{a_i^0}), \varphi \right\rangle = A + B + C,$$

where

$$\begin{aligned}
A &= \left\langle \pi \sum_{i=1}^{l} d_i \delta_{b_i^\varepsilon(T_\varepsilon)} - J u_\varepsilon(T_\varepsilon), \chi \right\rangle + \delta \left\langle \pi \sum_{i=1}^{l} \delta_{b_i^\varepsilon(T_\varepsilon)} - \mu_\varepsilon(T_\varepsilon), \varphi \right\rangle \\
&\leq C\left(\left\| J u_\varepsilon(T_\varepsilon) - \sum_{i=1}^{l} d_i \delta_{b_i^\varepsilon(T_\varepsilon)} \right\| + \delta \left\| \mu_\varepsilon(T_\varepsilon) - \sum_{i=1}^{l} \delta_{b_i^\varepsilon(T_\varepsilon)} \right\| \right), \\
B &= \langle J u_\varepsilon(T_\varepsilon) - J u_\varepsilon(0), \chi \rangle + \delta \langle \mu_\varepsilon(T_\varepsilon) - \mu_\varepsilon(0), \varphi \rangle, \\
C &= \left\langle J u_\varepsilon^0 - \pi \sum_{i=1}^{l} d_i \delta_{a_i^0}, \chi \right\rangle + \delta \left\langle \mu_\varepsilon(u_\varepsilon^0) - \pi \sum_{i=1}^{l} \delta_{a_i^0}, \varphi \right\rangle \\
&\leq C\left(\left\| J u_\varepsilon^0 - \sum_{i=1}^{l} d_i \delta_{a_i^0} \right\| + \delta \left\| \mu_\varepsilon(u_\varepsilon^0) - \sum_{i=1}^{l} \delta_{a_i^0} \right\| \right).
\end{aligned}$$

In view of the bound provided by (5-9) for B, estimates (4-3)–(4-4) and the fact that $\Sigma_\varepsilon^0(s) \leq K_1$ for $0 \leq s \leq T_\varepsilon$, this implies

$$\eta_0 = \left\| J u_\varepsilon(T_\varepsilon) - \pi \sum_{i=1}^{l} d_i \delta_{a_i^0} \right\| \leq C\left(\varepsilon |\log \varepsilon| + |\log \varepsilon|^{-1} + |\log \varepsilon|^{-\frac{1}{2}}\right) + C T_\varepsilon,$$

and letting $\varepsilon \to 0$ yields the conclusion. Lemma 5.3 is proved. □

Conclusion of the proof of Theorem 5.1. We consider $t, s \in [0, T]$. Arguing as in the proof of Lemma 5.3 (with T_ε and 0 replaced by t and s), we find that for all χ, φ belonging to $\mathcal{K}(a_i^0)$,

$$\left|\sum_{i=1}^{l} d_i\big[\chi(b_i^\varepsilon(t)) - \chi(b_i^\varepsilon(s))\big] + \delta\big[\varphi(b_i^\varepsilon(t)) - \varphi(b_i^\varepsilon(s))\big]\right|$$
$$\le C \sup_{\tau\in[0,T]}\left(\Big\|Ju_\varepsilon(\tau) - \sum_{i=1}^{l} d_i\delta_{b_i^\varepsilon(\tau)}\Big\| + \delta\Big\|\mu_\varepsilon(\tau) - \sum_{i=1}^{l}\delta_{b_i^\varepsilon(\tau)}\Big\|\right)$$
$$+\left|\langle Ju_\varepsilon(t) - Ju_\varepsilon(s), \chi\rangle + \delta\langle \mu_\varepsilon(t) - \mu_\varepsilon(s), \varphi\rangle\right|,$$

which is bounded by $o_\varepsilon(1) + c|t-s|$ because of (4-3), (4-4) and (5-9). Considering successively $\chi(x) = e_1 \cdot x$ and $\chi(x) = e_2 \cdot x$ on each $B(a_i^0, r_a)$, we obtain

$$|b_i^\varepsilon(t) - b_i^\varepsilon(s)| \le c|t-s| + o_\varepsilon(1). \tag{5-10}$$

Next, using that $b_i^\varepsilon \in B(a_i^0, r_a)$ and a standard diagonal argument, we may construct a sequence $(\varepsilon_k) \to 0$ and paths $b_i(t)$ such that $b_i^{\varepsilon_k}(t)$ converges to $b_i(t)$ for all $t \in \mathbb{Q} \cap [0, T]$. We infer then from (4-3)–(4-4) that the convergence statements (5-1)–(5-2) in Theorem 5.1 hold for these times. Moreover, in view of (5-10) these paths are Lipschitz on $[0, T] \cap \mathbb{Q}$, so that they can be extended in a unique way to Lipschitz paths (still denoted by $b_i(t)$) on the whole of $[0, T]$. We can finally establish that the convergence (5-1)–(5-2) holds uniformly with respect to $t \in [0, T]$ by again using (5-10) and (4-3)–(4-4).

Finally, we already know from (5-5) that the estimate (5-3) holds for the full family $(u_\varepsilon)_{\varepsilon<\varepsilon_\Lambda}$. To show (5-4), we now recall the uniform bound $\mathscr{E}_{\varepsilon,[U_d]}(u_\varepsilon(t)) - \mathscr{E}_{\varepsilon,[U_d]}(u_\varepsilon^*(a_i^0, d_i)) \le K_1$, and observe also that Corollary 3.8 gives

$$\mathscr{E}_{\varepsilon,[U_d]}(u_\varepsilon^*(a_i^0, d_i)) - \mathscr{E}_{\varepsilon,[U_d]}(u_\varepsilon^*(b_i(t), d_i)) = W(a_i^0, d_i) - W(b_i(t), d_i) \le C,$$

since the b_i are continuous and remain separated on $[0, T]$. This yields the bound (5-4) and concludes the proof of Theorem 5.1. □

As mentioned early in the proof of Theorem 5.1, the convergence of the initial data in (WP$_1$) actually holds on every large ball $B(L)$, $L = 2^n \ge R$, so we find the same conclusions replacing R by L.

Lemma 5.4 [Bethuel et al. 2008, Lemma 7.3]. *There exists a subsequence, still denoted by ε_k, such that for all $L \ge 2^{n_1}$,*

$$\eta_k := \sup_{[0,T]}\Big\|Ju_{\varepsilon_k}(t) - \pi\sum_{i=1}^{l} d_i\delta_{b_i(t)}\Big\|_{W_0^{1,\infty}(B(L))^*} \to 0, \qquad k \to +\infty.$$

For $t \in [0, T]$ and sufficiently large $k \in \mathbb{N}$, we may therefore apply Theorem 4.1 to $u_{\varepsilon_k}(t)$ with respect to the configuration $(b_i(t), d_i)$ and with the choice $R_0 = L = 2^n$ for each $n \ge n_1$. We are led to introduce the excess energy at time t with respect to the configuration $(b_i(t), d_i)$ by

$$\Sigma_{\varepsilon_k}(t) = \mathscr{E}_{\varepsilon_k,U_d}(u_{\varepsilon_k}(t)) - \mathscr{E}_{\varepsilon_k,U_d}(u_{\varepsilon_k}^*(b_i(t), d_i)),$$

which is uniformly bounded on $[0, T]$ in view of (5-4). Letting first k, then n tend to $+\infty$, we can get rid of the dependence on R in (4-2).

Lemma 5.5. *For all $r \le r_a/2$ and $K \ge 2^{n_1}$, we have for sufficiently large k and $t, t_1, t_2 \in [0, T]$*

$$\int_{B(K)\setminus\cup B(b_i(t),r)} e_{\varepsilon_k}\big(|u_{\varepsilon_k}(t)|\big) + \frac{1}{8}\left|\frac{j(u_{\varepsilon_k}(t))}{|u_{\varepsilon_k}(t)|} - j(u^*(b_i(t), d_i))\right|^2 \le \Sigma_{\varepsilon_k}(t) + C\Big(\varepsilon_k, \eta_k, \frac{1}{K}\Big).$$

Therefore, we have as $k \to +\infty$

$$\limsup_{k\to+\infty} \int_{t_1}^{t_2}\int_{B(K)\setminus\cup B(b_i(t),r)} e_{\varepsilon_k}\big(|u_{\varepsilon_k}(t)|\big) + \frac{1}{8}\left|\frac{j(u_{\varepsilon_k})(t)}{|u_{\varepsilon_k}(t)|} - j(u^*(b_i(t), d_i))\right|^2 \le \limsup_{k\to+\infty}\int_{t_1}^{t_2} \Sigma_{\varepsilon_k}(t).$$

Thus, the distance between $u_{\varepsilon_k}(t)$ and $u^*(b_i(t), d_i)$ can be asymptotically controlled by $\limsup \Sigma_{\varepsilon_k}(t)$.

We now define the trajectory set

$$\mathcal{T} = \{(t, b_i(t)),\ t \in [0, T],\ i = 1, \dots, l\}$$

and its complement

$$\mathcal{G} = [0, T] \times \mathbb{R}^2 \setminus \mathcal{T}.$$

Thanks to the uniform bounds in $L^2_{\text{loc}}(\mathcal{G})$ provided by Lemma 5.5, we establish:

Proposition 5.6. *There exists a subsequence, still denoted ε_k, such that*

$$\frac{j(u_{\varepsilon_k})}{|u_{\varepsilon_k}|} \rightharpoonup j(u^*(b_i(\cdot), d_i))$$

weakly in $L^2_{\text{loc}}(\mathcal{G})$ as $k \to +\infty$.

Proof. Let B be any bounded subset of $\mathbb{R}^2$. According to Lemma 5.4,

$$\operatorname{curl}(j(u_{\varepsilon_k})) = 2Ju_{\varepsilon_k} \to 2\pi\sum_{i=1}^{l} d_i\delta_{b_i(\cdot)} = \operatorname{curl}\big(j(u^*(b_i(\cdot), d_i))\big) \quad \text{in } \mathcal{D}'([0, T] \times B). \tag{5-11}$$

At the same time, we have

$$\operatorname{div}(j(u_{\varepsilon_k})) \to 0 = \operatorname{div}\big(j(u^*(b_i(\cdot), d_i))\big) \quad \text{in } \mathcal{D}'([0, T] \times B). \tag{5-12}$$

Indeed, since u_{ε_k} solves (CGL)$_\varepsilon$, we obtain by considering the exterior product

$$k_{\varepsilon_k}u_{\varepsilon_k} \times \partial_t u_{\varepsilon_k} + u_{\varepsilon_k}\cdot\partial_t u_{\varepsilon_k} = u_{\varepsilon_k} \times \Delta u_{\varepsilon_k} = \operatorname{div}(j(u_{\varepsilon_k})),$$

so we are led to

$$\operatorname{div}\big(j(u_{\varepsilon_k})\big) = k_{\varepsilon_k}u_{\varepsilon_k} \times \partial_t u_{\varepsilon_k} + \tfrac{1}{2}\varepsilon_k\frac{d}{dt}\left(\frac{|u_{\varepsilon_k}|^2 - 1}{\varepsilon_k}\right). \tag{5-13}$$

Now, applying Lemma 3.5 to u_{ε_k}, we find

$$\sup_{[0,T]} E_{\varepsilon_k}(u_{\varepsilon_k}(t), B) \le \pi l|\log\varepsilon| + \Sigma_{\varepsilon_k}(t) + C \le \pi l|\log\varepsilon_k| + C, \tag{5-14}$$

where the second inequality is itself a consequence of (5-4). This implies that $|u_{\varepsilon_k}| \to 1$ in $L^2([0, T] \times B)$. Moreover, we infer that the second term on the right-hand side of (5-13) converges to zero in the distribution sense on $[0, T] \times B$. For the first one, it suffices to use the Cauchy–Schwarz inequality combined with the L^2 bound provided by Lemma 5.2 and the already mentioned uniform bounds of $|u_{\varepsilon_k}|$ in L^2_{loc}.

We then infer from Lemma 5.4 and (5-14) that $j(u_{\varepsilon_k})$ is uniformly bounded in $L^p_{\mathrm{loc}}([0,T]\times\mathbb{R}^2)$ for all $p<2$. This is a consequence of, for example, [Colliander and Jerrard 1999, Theorem 3.2.1 and subsequent remarks]. We deduce from (5-11) and (5-12) that up to a subsequence, we have

$$j(u_{\varepsilon_k}) \rightharpoonup j_1 = j(u^*(b_i(\cdot),d_i)) + H \tag{5-15}$$

weakly in $L^p_{\mathrm{loc}}([0,T]\times\mathbb{R}^2)$, where H is harmonic in x on $[0,T]\times\mathbb{R}^2$.

On the other hand, it follows from the first part of Lemma 5.5 that there exists j_2 such that, taking subsequences if necessary, $j(u_{\varepsilon_k})/|u_{\varepsilon_k}| \rightharpoonup j_2$ weakly in $L^2_{\mathrm{loc}}(\mathcal{G})$.

Taking into account the strong convergence $|u_{\varepsilon_k}|\to 1$ in $L^2_{\mathrm{loc}}([0,T]\times\mathbb{R}^2)$, we obtain $j_1=j_2\in L^2_{\mathrm{loc}}(\mathcal{G})$. The second part of Lemma 5.5 combined with (5-15) then yields

$$\|H\|_{L^2_{\mathrm{loc}}(\mathcal{G})} \le \liminf_{k\to+\infty} \left\| \frac{j(u_{\varepsilon_k})}{|u_{\varepsilon_k}|} - j(u^*(b_i,d_i)) \right\|_{L^2_{\mathrm{loc}}(\mathcal{G})} \le CT,$$

where C depends only on K_1, R, and r_a, so finally $\|H\|_{L^2([0,T]\times\mathbb{R}^2)}\le CT$. Since H is harmonic in x, we find that $H(t,\cdot)$ is bounded on $\mathbb{R}^2$ for almost every t and therefore is identically zero. We end up with $j_1=j_2=j(u^*(b_i(\cdot),d_i))$ in $\mathcal{G}$, and the conclusion follows. □

6. Proof of Theorem 1.3

Let $\{b_i(t)\}$ be the l Lipschitz paths on $[0,T]$ provided by Theorem 5.1, and $\{a_i(t)\}$ the unique maximal solution defined on $I=[0,T^*)$ to (1-5) with initial conditions a_i^0. Our aim is to show that $a_i(t)\equiv b_i(t)$ on I. We prove this first on $[0,T]$. By Rademacher's Theorem, the time derivatives $\dot b_i(t)$ exist and are bounded almost everywhere on $[0,T]$. Without loss of generality, we may assume $T<T^*$, so

$$|\dot a_i(t)|\le C,\ |\dot b_i(t)|\le C \qquad \text{a.e. on } [0,T]. \tag{6-1}$$

Moreover, we may assume, possibly after decreasing T, that $|a_i(t)-b_i(t)|\le r_a/2$ for all i. Hence, the trajectories $a_i(t)$ remain in $B(a_i^0,r_a)$ on $[0,T]$. We introduce

$$h(t)=\sum_{i=1}^{l}\int_0^t \big|\dot a_i(s)-\dot b_i(s)\big|\,ds, \qquad \sigma(t)=\sum_{i=1}^{l}\big|a_i(t)-b_i(t)\big|.$$

Then h is Lipschitz on $[0,T]$ and for almost every $t\in[0,T]$ we have $h'(t)=\sum_{i=1}^{l}\big|\dot a_i(t)-\dot b_i(t)\big|$. Note that since σ is absolutely continuous and $\sigma(0)=0$, we have for all $t\in[0,T]$

$$\sigma(t)=\int_0^t \sigma'(s)\,ds\le h(t).$$

Therefore it suffices to show that h is identically zero on $[0,T]$. This will be done by means of Grönwall's Lemma.

Lemma 6.1. *For all $t_1,t_2,t\in[0,T]$, we have*

$$\limsup_{k\to+\infty}\Sigma_{\varepsilon_k}(t)\le Ch(t) \quad \textit{and} \quad \limsup_{k\to+\infty}\int_{t_1}^{t_2}\Sigma_{\varepsilon_k}(s)\,ds\le C\int_{t_1}^{t_2}h(s)\,ds,$$

where C only depends on r_a, K_0, and R_a.

Proof. For $t \in [0, T]$, we decompose $\Sigma_{\varepsilon_k}(t)$ as

$$\Sigma_{\varepsilon_k}(t) = \mathscr{E}_{\varepsilon_k,[U_d]}(u_{\varepsilon_k}(t)) - \mathscr{E}_{\varepsilon_k,[U_d]}(u^0_{\varepsilon_k}) + \Sigma_{\varepsilon_k}(0) + \mathscr{E}_{\varepsilon_k,[U_d]}(u^*_{\varepsilon_k}(a^0_i, d_i)) - \mathscr{E}_{\varepsilon_k,[U_d]}(u^*_{\varepsilon_k}(b_i(t), d_i)).$$

Appealing to Corollary 3.8 and Theorem 1.1, we obtain

$$\Sigma_{\varepsilon_k}(t) = -\delta \int_0^t \int_{\mathbb{R}^2} \frac{|\partial_t u_{\varepsilon_k}|^2}{|\log \varepsilon_k|} + \Sigma_{\varepsilon_k}(0) + W(a^0_i, d_i) - W(b_i(t), d_i) + o_{\varepsilon_k}(1).$$

Using that W is Lipschitz away from zero, we estimate the difference on the right by

$$\begin{aligned} W(a^0_i, d_i) - W(b_i(t), d_i) &= W(a^0_i, d_i) - W(a_i(t), d_i) + W(a_i(t), d_i) - W(b_i(t), d_i) \\ &\le -\int_0^t \sum_{i=1}^{l} \dot{a}_i(s) \cdot \nabla_{a_i} W(s)\, ds + C\sigma(t). \end{aligned}$$

Since the a_i solve the Cauchy problem (1-5), an explicit computation gives

$$\dot{a}_i(s) \cdot \nabla_{a_i} W(s) = \frac{\delta}{\pi} C_i d_i |\nabla_{a_i} W|^2 = -\delta\pi\, |\dot{a}_i(s)|^2,$$

so

$$\Sigma_{\varepsilon_k}(t) \le \Sigma_{\varepsilon_k}(0) + \delta\pi \int_0^t \sum_{i=1}^{l} |\dot{a}_i(s)|^2\, ds - \delta \int_0^t \int_{\mathbb{R}^2} \frac{|\partial_t u_{\varepsilon_k}|^2}{|\log \varepsilon_k|} + C\sigma(t) + o_{\varepsilon_k}(1).$$

For the energy dissipation on the right-hand side, we need a lower bound as ε_k tends to zero. In view of the convergence of the Jacobians (5-1) and the upper bound for the energy

$$\sup_{t\in[0,T]} E_{\varepsilon_k}(u_{\varepsilon_k}(t), B(R')) \le \pi l |\log \varepsilon_k| + C$$

stated in (5-14), Proposition 3 in [Jerrard 1999] (see also Corollary 7 in [Sandier and Serfaty 2004]) provides the lower mobility bound

$$\liminf_{k\to+\infty} \int_0^t \int_{\mathbb{R}^2} \frac{|\partial_t u_{\varepsilon_k}|^2}{|\log \varepsilon_k|} \ge \pi \sum_{i=1}^{l} \int_0^t \left|\dot{b}_i(t)\right|^2 ds. \tag{6-2}$$

Now, thanks to (6-1), we have

$$\sum_{i=1}^{l} \int_0^t \left(|\dot{a}_i(s)|^2 - |\dot{b}_i(s)|^2\right) \le C \sum_{i=1}^{l} \int_0^t \left|\dot{a}_i(s) - \dot{b}_i(s)\right| ds = Ch(t),$$

whereas $\Sigma_{\varepsilon_k}(0) \to 0$ by assumption; hence we get

$$\limsup_{k\to+\infty} \Sigma_{\varepsilon_k}(t) \le C\,(\sigma(t) + h(t)).$$

Applying Fatou's Lemma in (6-2) yields the corresponding integral version as well. We conclude by using that $\sigma \le h$. □

As suggested in the introduction, the map $u^*(a_i(t), d_i)$ solves the evolution formula provided by Proposition 2.1 in the asymptotic limit where $\varepsilon \to 0$.

Lemma 6.2. *For $t \in [0,T]$ and $\chi, \varphi \in \mathscr{H}(a_i^0)$,*

$$\pi \frac{d}{dt} \sum_{i=1}^{l} \big[d_i \chi(a_i(t)) + \delta\varphi(a_i(t))\big] = -2 \int_{\mathbb{R}^2} \operatorname{Im}\left(\omega(u^*(a_i(t), d_i)) \frac{\partial^2 \chi}{\partial \bar{z}^2}\right).$$

Proof. We use the following formula, proved in [Bethuel et al. 2005] and valid for any configuration (a_i, d_i) and any test function χ that is affine near the point vortices:

$$-2 \int_{\mathbb{R}^2} \operatorname{Im}\left(\omega(u^*(a_i(t), d_i)) \frac{\partial^2 \chi}{\partial \bar{z}^2}\right) = -2\pi \sum_{i \neq j} d_i d_j \frac{(a_i(t) - a_j(t))^{\perp}}{|a_i(t) - a_j(t)|^2} \cdot \nabla \chi(a_i(t)).$$

We also compute

$$\frac{d}{dt} \sum_{i=1}^{l} \big[d_i \chi(a_i) + \delta\varphi(a_i)\big] = \sum_{i=1}^{l} \big[d_i \nabla\chi(a_i^0) \cdot \dot{a}_i(t) + \delta \nabla\varphi(a_i^0) \cdot \dot{a}_i(t)\big] = \sum_{i=1}^{l} d_i \nabla\chi(a_i^0) \cdot \big(\dot{a}_i(t) - \delta d_i \dot{a}_i^{\perp}(t)\big),$$

where the second equality follows from the relation $\nabla\varphi(a_i^0) = \nabla^{\perp}\chi(a_i^0)$. Next, we deduce from (1-5)

$$\pi\big(\dot{a}_i(t) - \delta d_i \dot{a}_i^{\perp}(t)\big) = -C_i(1 + \delta^2 d_i^2) \nabla_{a_i}^{\perp} W = d_i \nabla_{a_i}^{\perp} W,$$

and we obtain

$$\pi \frac{d}{dt} \sum_{i=1}^{l} \big[d_i \chi(a_i) + \delta\varphi(a_i)\big] = \sum_{i=1}^{l} \nabla\chi(a_i) \cdot \nabla_{a_i}^{\perp} W = -2\pi \sum_{i \neq j} d_i d_j \frac{(a_i - a_j)^{\perp}}{|a_i - a_j|^2} \cdot \nabla\chi(a_i),$$

which yields the conclusion. □

Lemma 6.3. *Set $A = \bigcup_{i=1}^{l} B(a_i^0, 2r_a) \setminus B(a_i^0, r_a)$ and let $t_1, t_2 \in [0,T]$. For all $\varphi \in \mathfrak{D}(A)$, we have*

$$\limsup_{k \to +\infty} \left| \int_{t_1}^{t_2} \int_A \big[\omega(u_{\varepsilon_k}(s)) - \omega(u^*(b_i(s), d_i))\big] \varphi \right| \le C \|\varphi\|_{\infty} \int_{t_1}^{t_2} h(s)\, ds.$$

Proof. We apply the pointwise equality (2-4) to $u \equiv u_{\varepsilon_k}(t)$ and $u^* \equiv u^*(b_i(t), d_i)$ for all t. Since $|u^*(b_i(t), d_i)| = 1$, this gives

$$\omega(u) - \omega(u^*) = \sum_{k,l=1}^{2} \left(a_{k,l} \partial_l |u| \partial_k |u| + b_{k,l} \left[\frac{j_k(u)}{|u|} \frac{j_l(u)}{|u|} - j_k(u^*) j_l(u^*) \right] \right),$$

where $a_{k,l}, b_{k,l} \in \mathbb{C}$. We rewrite the terms involving the components of j as

$$\begin{aligned} \frac{j_k(u)}{|u|} \frac{j_l(u)}{|u|} - j_k(u^*) j_l(u^*) = &\left(\frac{j_k(u)}{|u|} - j_k(u^*) \right) \left(\frac{j_l(u)}{|u|} - j_l(u^*) \right) \\ &+ j_k(u^*) \left(\frac{j_l(u)}{|u|} - j_l(u^*) \right) + j_l(u^*) \left(\frac{j_k(u)}{|u|} - j_k(u^*) \right). \end{aligned}$$

We multiply the previous equality by φ, integrate on $[t_1,t_2]\times A$ and let k go to $+\infty$. Using the weak convergence in L^2 of $j(u_{\varepsilon_k})/|u_{\varepsilon_k}|$ to $j(u^*(b_i(.),d_i))$ on $[0,T]\times A\subset\mathcal{G}$, we deduce that

$$\limsup_{k\to+\infty}\left|\int_{t_1}^{t_2}\int_A\big[\omega(u_{\varepsilon_k}(s))-\omega(u^*(b_i(s),d_i))\big]\varphi\right| \\ \leq C\|\varphi\|_\infty \limsup_{k\to+\infty}\int_{t_1}^{t_2}\int_A\left(\big|\nabla|u_{\varepsilon_k}|\big|^2+\left|\frac{j(u_{\varepsilon_k})}{|u_{\varepsilon_k}|}-j(u^*(b_i,d_i))\right|^2\right).$$

The conclusion finally follows from Lemmas 5.5 and 6.1. □

We are now in a position to complete the proof of Theorem 1.3. We consider arbitrary χ,φ belonging to $\mathcal{K}(a_i^0)$, we fix $0\leq s\leq t\leq T$ and we integrate the evolution formula given by Proposition 2.1 on $[s,t]$. We obtain

$$\int_s^t\frac{d}{d\tau}\int_{\mathbb{R}^2}Ju_{\varepsilon_k}(\tau)\chi+\delta\int_{\mathbb{R}^2}\mu_{\varepsilon_k}(\tau)\varphi=\int_s^t g_k^1(\tau)+\int_s^t g_k^2(\tau),$$

where

$$g_k^1(\tau)=-\delta\int_{\mathbb{R}^2}\frac{|\partial_t u_{\varepsilon_k}|^2}{|\log\varepsilon_k|^2}+R_{\varepsilon_k}(\tau,\chi,\varphi,u_{\varepsilon_k}),\qquad g_k^2(\tau)=-2\int_{\mathbb{R}^2}\mathrm{Im}\left(\omega(u_{\varepsilon_k}(\tau))\frac{\partial^2\chi}{\partial\bar z^2}\right).$$

We decompose the latter as

$$g_k^2=-2\int_{\mathbb{R}^2}\mathrm{Im}\left(\big[\omega(u_{\varepsilon_k})-\omega(u^*(b_i,d_i))\big]\frac{\partial^2\chi}{\partial\bar z^2}\right)-2\int_{\mathbb{R}^2}\mathrm{Im}\left(\big[\omega(u^*(b_i,d_i))-\omega(u^*(a_i,d_i))\big]\frac{\partial^2\chi}{\partial\bar z^2}\right) \\ -2\int_{\mathbb{R}^2}\mathrm{Im}\left(\omega(u^*(a_i,d_i))\frac{\partial^2\chi}{\partial\bar z^2}\right)=A_k(\tau)+B_k(\tau)+C_k(\tau).$$

We next substitute the expression given in Lemma 6.2 for C_k in the previous equalities. Setting

$$f_{k,\chi,\varphi}(\tau)=\int_{\mathbb{R}^2}Ju_{\varepsilon_k}(\tau)\chi+\delta\int_{\mathbb{R}^2}\mu_{\varepsilon_k}(\tau)\varphi-\pi\sum_{i=1}^{l}\big[d_i\chi(a_i(\tau))+\delta\varphi(a_i(\tau))\big],$$

we obtain

$$f_{k,\chi,\varphi}(t)-f_{k,\chi,\varphi}(s)=\int_s^t g_k^1+\int_s^t A_k+\int_s^t B_k.$$

Lemma 5.2 with $T_\varepsilon=T$ first gives $\left|\int_s^t g_k^1(\tau)\,d\tau\right|\leq C|\log\varepsilon_k|^{-1/2}$ for all k. Moreover, it follows from Lemma 6.3 and inclusion $\operatorname{supp}\partial^2\chi/\partial\bar z^2\subset A$ that

$$\limsup_{k\to+\infty}\left|\int_s^t A_k(\tau)\,d\tau\right|\leq C\int_s^t h(\tau)\,d\tau.$$

Finally, the regularity of $\omega(u^*)$ away from the vortices gives

$$\int_s^t|B_k(\tau)|\,d\tau\leq C\int_s^t\sigma(\tau)\,d\tau\leq C\int_s^t h(\tau)\,d\tau.$$

Letting k go to $+\infty$, we deduce from the convergence statements in Theorem 5.1 that for $0 \le s \le t \le T$,

$$|f_{\chi,\varphi}(t) - f_{\chi,\varphi}(s)| \le C \int_s^t h(\tau)\, d\tau, \tag{6-3}$$

where the constant C depends only on χ, φ and the initial conditions, and $f_{\chi,\varphi}$ is defined by

$$f_{\chi,\varphi} = \pi \sum_{i=1}^{l} \big[d_i(\chi(b_i) - \chi(a_i)) + \delta(\varphi(b_i) - \varphi(a_i)) \big].$$

We now fix a time $t \in [0, T]$ at which all the vortices b_i have a time derivative. Since the a_i are C^1, it follows that $f_{\chi,\varphi}$ is differentiable at t with time derivative given by

$$f'_{\chi,\varphi}(t) = \pi \sum_{i=1}^{l} \big(d_i \nabla\chi(a_i^0) + \delta\nabla^\perp\chi(a_i^0) \big) \cdot \big(\dot{b}_i(t) - \dot{a}_i(t) \big).$$

Dividing by $t - s$ in (6-3) and letting $s \to t$ then gives

$$\left| \pi \sum_{i=1}^{l} \big(d_i \nabla\chi(a_i^0) + \delta\nabla^\perp\chi(a_i^0) \big) \cdot \big(\dot{b}_i(t) - \dot{a}_i(t) \big) \right| \le C\, h(t).$$

So, considering in particular $\chi, \varphi \in \mathscr{H}(a_i^0)$ such that χ and φ vanish near each point a_i^0 except for one, we obtain for all $i = 1, \dots, l$

$$\left| \pi \big(d_i \nabla\chi(a_i^0) + \delta\nabla^\perp\chi(a_i^0) \big) \cdot \big(\dot{b}_i(t) - \dot{a}_i(t) \big) \right| \le C\, h(t).$$

Choosing then successively $\chi(x) = x_1$ and $\chi(x) = x_2$ near a_i^0 we end up with $|\dot{b}_i(t) - \dot{a}_i(t)| \le Ch(t)$, and it follows by summation that $h'(t) \le Ch(t)$ for a.e. $t \in [0, T]$. Since $h(0) = 0$, this implies that $h = 0$ on $[0, T]$, and hence $\sigma = 0$ on $[0, T]$. Applying Lemma 6.1, we infer that $\limsup_{k\to+\infty} \Sigma_{\varepsilon_k}(t) \le 0$. Besides, Lemma 3.5 yields for all $L \ge 2^{n_1}$

$$\liminf_{k\to+\infty} \Sigma_{\varepsilon_k}(t) \ge \liminf_{k\to+\infty} \int_{B(L)} \big[e_{\varepsilon_k}(u_{\varepsilon_k}(t)) - e_{\varepsilon_k}(u^*_{\varepsilon_k}(a_i(t), d_i)) \big] - \frac{C}{L} \ge -\frac{C}{L},$$

where the second inequality follows from the convergence of Jacobians on $B(L)$ stated in Lemma 5.4; see [Jerrard and Spirn 2007; Lin and Xin 1999]. Letting L tend to $+\infty$, we obtain $\liminf_{k\to+\infty} \Sigma_{\varepsilon_k}(t) \ge 0$, so we deduce from (5-3) that $(u_{\varepsilon_k}(t))_{k\in\mathbb{N}}$ is well-prepared with respect to the configuration $(a_i(t), d_i)$. By the uniqueness of the limit, this holds for the full family $(u_\varepsilon(t))_{0<\varepsilon<1}$ on $[0, T]$.

In conclusion, we observe that in our definition T only depends on K_1, r_a and $\max(R, R_a + r_a)$, so we can extend our results to the whole of $[0, T^*)$ by repeating the previous arguments.

Appendix: The Cauchy problem for $(\mathrm{CGL})_\epsilon$

We present here the proof of Theorem 1.1. We omit the dependence on ε and rewrite (1-2) in the form

$$\begin{cases} \partial_t w = (a + ib)(\Delta w + f_{U_0}(w)), \\ w(0) = w_0 \in H^1(\mathbb{R}^2), \end{cases} \tag{CGL}$$

where a is positive, b is real and

$$f_{U_0}(w) = \Delta U_0 + (U_0 + w)(1 - |U_0 + w|^2).$$

We denote by $S = S(t, x)$ the semigroup operator associated to the corresponding homogeneous linear equation. Every solution $w \in C^0\big([0, T], H^1(\mathbb{R}^2)\big)$ to (CGL) satisfies the Duhamel formula

$$w(t, \cdot) = S(t, \cdot) * w_0 + \int_0^t S(t-s, \cdot) * g_{U_0}(w(s), \cdot)\, ds, \qquad s \in [0, T],$$

where $g_{U_0} = (a + ib) f_{U_0}$. The kernel S is explicitly given by

$$S(t, x) = \frac{1}{4\pi(a+ib)t} \exp\Big(\frac{-|x|^2}{4(a+ib)t}\Big).$$

Since a is positive, S decays at infinity like the standard heat kernel; therefore (CGL) enjoys the same smoothing properties as the parabolic Ginzburg–Landau equation. In particular, we have for all $t > 0$ and for all $1 \le r \le +\infty$

$$\big\| S(t, \cdot)\big\|_{L^r(\mathbb{R}^2)} \le \frac{C_{a,b}}{t^{1-1/r}} \tag{A-1}$$

and concerning the space derivatives of $S(t)$,

$$\big\| D^k S(t, \cdot)\big\|_{L^r(\mathbb{R}^2)} \le \frac{C_{a,b}}{t^{(|k|/2)+1-1/r}}. \tag{A-2}$$

We will often use Young's inequality, which states that, if $1 + \dfrac{1}{r} = \dfrac{1}{p} + \dfrac{1}{q}$ and $f \in L^p(\mathbb{R}^2)$, $g \in L^q(\mathbb{R}^2)$, then

$$\|f * g\|_{L^r(\mathbb{R}^2)} \le \|f\|_{L^p(\mathbb{R}^2)} \|g\|_{L^q(\mathbb{R}^2)}.$$

We first state a local well-posedness result for (CGL).

Proposition A.4. *Let $w_0 \in H^1(\mathbb{R}^2)$. There exists a positive time T^* depending on $\|w_0\|_{H^1}$ and a unique solution $w \in C^0\big([0, T^*), H^1(\mathbb{R}^2)\big)$ to (CGL).*

Proof. We intend to apply the fixed point theorem to the map $\psi : w \in H^1(\mathbb{R}^2) \mapsto \psi(w)$, where

$$\psi(w)(t) = S(t) * w_0 + \int_0^t S(t-s) * g_{U_0}(w(s))\, ds.$$

To this aim, we introduce $R = \|w_0\|_{H^1(\mathbb{R}^2)}$ and for $T > 0$

$$B(T, R) = \big\{ w \in L^\infty\big([0, T], H^1(\mathbb{R}^2)\big) : \|w\|_{L^\infty(H^1)} \le (2C_{a,b} + 1)R \big\},$$

where $C_{a,b}$ is the constant appearing in (A-1)–(A-2). We next show that we can choose $T = T(R)$ so that ψ maps $B(T(R), R)$ into itself and is a contraction on this ball.

For $T > 0$, we let $w \in B(T, R)$ and expand $f_{U_0}(w)$. Using that $H^1(\mathbb{R}^2)$ is continuously embedded in $L^p(\mathbb{R}^2)$ for all $2 \le p < +\infty$ and that U_0 belongs to $\mathcal{V}$, it can be shown that

$$\big\| f_{U_0}\big\|_{L^\infty([0,T], L^2)} \le C(U_0, R) \tag{A-3}$$

(see [Bethuel and Smets 2007, Lemma 1]), and that for $w_1, w_2 \in B(T, R)$

$$\| f_{U_0}(w_1) - f_{U_0}(w_2)\|_{L^\infty([0,T], L^2)} \le C(U_0, R) \|w_1 - w_2\|_{L^\infty([0,T], H^1)}. \tag{A-4}$$

We next apply Young's inequality to obtain

$$\begin{aligned}\|\psi(w)(t)\|_{H^1} &\le \|\psi(w)(t)\|_{L^2}+\|\nabla\psi(w)(t)\|_{L^2}\\ &\le 2\|S(t)\|_{L^1}\|w_0\|_{H^1}+\int_0^t \|S(t-s)+\nabla S(t-s)\|_{L^1}\|g_{U_0}(s)\|_{L^2}\,ds\\ &\le 2C_{a,b}\|w_0\|_{H^1}+C\int_0^t \big(1+(t-s)^{-1/2}\big)\|g_{U_0}(w(s))\|_{L^2}\,ds,\end{aligned}$$

where the last inequality is a consequence of (A-1) and (A-2) with the choice $r=1$. This yields, by (A-3) and (A-4),

$$\sup_{t\in[0,T]}\|\psi(w)(t)\|_{H^1}\le 2C_{a,b}\|w_0\|_{H^1}+C(U_0,R)(T+\sqrt{T})$$

and similarly,

$$\sup_{t\in[0,T]}\|\psi(w_1)(t)-\psi(w_2)(t)\|_{H^1}\le C'(U_0,R)(T+\sqrt{T})\sup_{t\in[0,T]}\|w_1(t)-w_2(t)\|_{H^1}.$$

The conclusion follows by choosing $T=T(R)$ sufficiently small so that $C(U_0,R)(T+\sqrt{T})\le R$ and $C'(U_0,R)(T+\sqrt{T})<1$. □

We next show additional regularity for a solution to (CGL).

Lemma A.5. *Let $w\in C^0\big([0,T],H^1(\mathbb{R}^2)\big)$ be a solution to (CGL). Then w belongs to*

$$L^1_{\text{loc}}\big([0,T],H^2(\mathbb{R}^2)\big)\cap C^0\big((0,T],H^2(\mathbb{R}^2)\big),$$

and therefore to $L^1_{\text{loc}}\big([0,T],L^\infty(\mathbb{R}^2)\big)$.

Proof. We first differentiate $f_{U_0}(w)$ and use [Bethuel and Smets 2007, Lemma 2] which states by means of various Sobolev embeddings, Hölder and Gagliardo–Nirenberg inequalities that

$$\partial_i f_{U_0}(w)=g_1(w)+g_2(w)\in L^\infty\big([0,T],L^2(\mathbb{R}^2)\big)+L^\infty\big([0,T],L^r(\mathbb{R}^2)\big)$$

for all $1<r<2$. Moreover, we have $\sup_{s\in[0,T]}\big(\|g_1(w)(s)\|_{L^2(\mathbb{R}^2)}+\|g_2(w)(s)\|_{L^r(\mathbb{R}^2)}\big)\le C(U_0,A(T),r)$, where $A(T)=\sup_{s\in[0,T]}\|w(s)\|_{H^1(\mathbb{R}^2)}$. Next, differentiating twice, Duhamel's formula gives

$$\partial_{ij}w(t)=\partial_j S(t)*\partial_i w_0+\int_0^t \partial_j S(t-s)*\partial_i f_{U_0}(s)\,ds,$$

so taking into account the decomposition $\partial_i f_{U_0}=g_1+g_2$ we get

$$\|\partial_{ij}w(t)\|_{L^2}\le\|\nabla S(t)\|_{L^1}\|\nabla w_0\|_{L^2}+\int_0^t\|\nabla S(t-s)\|_{L^1}\|g_1(s)\|_{L^2}\,ds+\int_0^t\|\nabla S(t-s)\|_{L^\alpha}\|g_2(s)\|_{L^r}\,ds,$$

where α is chosen so that $1+\dfrac{1}{2}=\dfrac{1}{\alpha}+\dfrac{1}{r}$. This finally yields, in view of (A-2),

$$\|\partial_{ij}w(t)\|_{L^2}\le Ct^{-1/2}\|w_0\|_{H^1}+C(U_0,A(T),r)\int_0^t\Big((t-s)^{-\frac12}+(t-s)^{-\frac12-1+\frac1\alpha}\Big)\,ds.$$

Since $\dfrac12+1-\dfrac1\alpha=\dfrac1r<1$, the right-hand side is finite, so $\partial_{ij}w\in L^1_{\text{loc}}\big([0,T],L^2(\mathbb{R}^2)\big)$. □

Lemma A.5 enables us to show that the renormalized energy is nonincreasing and hence to control $\|w(t)\|_{H^1(\mathbb{R}^2)}$. For (CGL), this energy is given by

$$E_{U_0}(w)(t) = \int_{\mathbb{R}^2} \frac{|\nabla w|^2}{2} - \int_{\mathbb{R}^2} \Delta U_0 \cdot w + \int_{\mathbb{R}^2} \frac{(1-|U_0+w|^2)^2}{4}.$$

It is well-defined and continuous in time for $w \in C^0(H^1(\mathbb{R}^2))$.

Lemma A.6. *Let $w \in C^0\big([0,T), H^1(\mathbb{R}^2)\big)$ be a solution to (CGL). Then for all $t \in (0,T)$ we have*

$$\frac{d}{dt} E_{U_0}(w)(t) \leq 0.$$

Moreover, there exists C_{U_0,w_0} depending only on U_0 and $\|w_0\|_{H^1}$ such that

$$\|w(t)\|_{H^1} \leq C_{U_0,w_0} \exp(C_{U_0,w_0} t) \quad \textit{for all } t \in [0,T). \tag{A-5}$$

Proof. We infer from (CGL) and Lemma A.5 that $\partial_t w$ belongs to $L^\infty_{\mathrm{loc}}\big((0,T], L^2(\mathbb{R}^2)\big)$, so we can compute

$$\begin{aligned}\frac{d}{dt} E_{U_0}(w(t)) &= \int_{\mathbb{R}^2} \nabla w \cdot \nabla \partial_t w - \Delta U_0 \cdot \partial_t w - \partial_t w \cdot (U_0+w)(1-|U_0+w|^2) \\ &= -\int_{\mathbb{R}^2} \partial_t w \cdot (\Delta w + f_{U_0}(w)) = -\int_{\mathbb{R}^2} \partial_t w \cdot \Big(\frac{1}{a+ib}\partial_t w\Big) = \frac{-a}{a^2+b^2}\int_{\mathbb{R}^2} |\partial_t w|^2 \leq 0.\end{aligned}$$

We now turn to (A-5). We compute, for $t \in (0,T)$,

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\|w(t)\|^2_{L^2(\mathbb{R}^2)} &= \int_{\mathbb{R}^2} w \cdot \partial_t w = \int_{\mathbb{R}^2} w \cdot [(a+ib)\Delta w] + \int_{\mathbb{R}^2} w \cdot [(a+ib) f_{U_0}(w)] \\ &= -a\int_{\mathbb{R}^2} |\nabla w|^2 + \int_{\mathbb{R}^2} w \cdot (a+ib)\Delta U_0 + \int_{\mathbb{R}^2} w \cdot \big[(a+ib)(U_0+w)(1-|U_0+w|^2)\big].\end{aligned}$$

We then split the last term in the previous equality as

$$\int_{\mathbb{R}^2} w \cdot \big[(a+ib)(U_0+w)(1-|U_0+w|^2)\big] = \int_{\mathbb{R}^2} w \cdot \big[(a+ib)U_0(1-|U_0+w|^2)\big] + a\int_{\mathbb{R}^2} |w|^2(1-|U_0+w|^2).$$

The last term on the right is clearly bounded by $a\|w(t)\|^2_{L^2(\mathbb{R}^2)}$. Using the Cauchy–Schwarz inequality for the first term, we obtain

$$\int_{\mathbb{R}^2} w \cdot \big[(a+ib)(U_0+w)(1-|U_0+w|^2)\big] \leq C(U_0)\|w(t)\|_{L^2} V(t)^{1/2} + a\|w(t)\|^2_{L^2},$$

where $V(t) = \int_{\mathbb{R}^2} (1-|U_0+w(t)|^2)^2$. We are led to

$$\frac{d}{dt}\|w(t)\|^2_{L^2(\mathbb{R}^2)} \leq C(U_0)\big(\|w(t)\|^2_{L^2} + 1 + V(t)\big). \tag{A-6}$$

On the other hand, Cauchy–Schwarz inequality gives

$$E_{U_0}(w)(t) \geq \int_{\mathbb{R}^2} \frac{|\nabla w|^2}{2}\, dx - C(U_0)\|w(t)\|_{L^2} + \frac{V(t)}{4},$$

which yields, since $E_{U_0}(w)$ is nonincreasing,

$$\frac{V(t)}{4}+\int_{\mathbb{R}^2}\frac{|\nabla w|^2}{2}\le E_{U_0}(w_0)+C(U_0)\|w(t)\|_{L^2}. \tag{A-7}$$

We infer from (A-6) and (A-7)

$$\|w(t)\|_{L^2}\le(1+\|w_0\|_{H^1})\exp(Ct)$$

and finally deduce (A-5) by using (A-7) once more. □

Lemma A.6 provides global well-posedness for (CGL).

Proposition A.7. *Let $w_0\in H^1(\mathbb{R}^2)$. Then there exists a unique and global solution $w\in C^0\big(\mathbb{R}_+, H^1(\mathbb{R}^2)\big)$ to* (CGL).

Proof. Let $w\in C^0\big([0,T^*), H^1(\mathbb{R}^2)\big)$ be the unique maximal solution with initial condition w_0. If T^* is finite, we have according to (A-5)

$$\limsup_{t\to T^*}\|w(t)\|_{H^1(\mathbb{R}^2)}\le C(U_0,T^*,w_0)<+\infty,$$

so that we can extend w to a solution $\overline{w}$ on $[0,T^*+\delta]$ for some positive δ. This yields a contradiction. □

We conclude this section with the following

Proposition A.8. *Let $w\in C^0\big(\mathbb{R}_+, H^1(\mathbb{R}^2)\big)$ be the solution to* (CGL). *Then $w\in C^\infty\big(\mathbb{R}_+^*, C^\infty(\mathbb{R}^2)\big)$.*

Proof. Step 1. Let $p\ge 2$ and $v\in H^p(\mathbb{R}^2)$. We show that $D^k f_{U_0}(v)\in L^2(\mathbb{R}^2)+L^{4/3}(\mathbb{R}^2)$ for all $|k|\le p$.

We may assume in view of the proof of Lemma A.5 that $|k|\ge 2$. We decompose $f_{U_0}(v)$ as $f_{U_0}(v)=\Delta U_0+h_{U_0}(v)$, where

$$h_{U_0}(v)=(U_0+v)(1-|U_0+v|^2).$$

Since $U_0\in\mathscr{V}$, it suffices to show that $D^k h_{U_0}(v)\in L^2(\mathbb{R}^2)+L^{4/3}(\mathbb{R}^2)$. Applying Leibniz's formula to $h_{U_0}(v)$, we obtain

$$\begin{aligned}D^k h_{U_0}(v)&=\sum_{m\le k}\binom{k}{m}D^{k-m}(U_0+v)D^m(1-|U_0+v|^2)\\&=D^k(U_0+v)-\sum_{\substack{m\le k\\ n\le m}}\binom{k}{m}\binom{m}{n}D^{k-m}(U_0+v)D^n(U_0+v)\cdot D^{m-n}(U_0+v).\end{aligned}$$

Since $2\le|k|\le p$, $v\in H^p(\mathbb{R}^2)$ and $U_0\in\mathscr{V}$, we clearly have $D^k(U_0+v)\in L^2(\mathbb{R}^2)$.

For the second term in the right-hand side, we write each product inside the sum as

$$D^a(U_0+v)D^b(U_0+v)\cdot D^c(U_0+v)$$

with $|a|+|b|+|c|=|k|\ge 2$, and we examine all cases. We observe that $D^a(v+U_0)$ belongs to $H^1(\mathbb{R}^2)$ whenever $1\le|a|\le p-1$ and hence to $L^4(\mathbb{R}^2)$, whereas $D^a(v+U_0)$ belongs to $L^2(\mathbb{R}^2)$ for $2\le|a|\le p$. Since $U_0+v\in L^\infty$, we finally obtain

$$D^a(U_0+v)D^b(U_0+v)\cdot D^c(U_0+v)\in L^2(\mathbb{R}^2)+L^{4/3}(\mathbb{R}^2),$$

which yields the conclusion.

Step 2: regularity in space for a solution to (CGL). Let $w \in C^0\big(\mathbb{R}_+, H^1(\mathbb{R}^2)\big)$ be the solution to (CGL). We show that $w \in C^0(\mathbb{R}_+^*, H^p(\mathbb{R}^2))$ for all $p \geq 1$.

We proceed by induction on p. The case $p = 2$ has already been treated in Lemma A.5, so we assume $w \in C^0\big(\mathbb{R}_+^*, H^p(\mathbb{R}^2)\big)$ for some $p \geq 2$. For $|k| \leq p+1$, we differentiate $w(t)$ and we find

$$D^k w(t) = D^k(S(t) * w_0) + D^k \int_0^t S(t-s) * g_{U_0}(s)\, ds$$

which we rewrite as

$$D^k w(t) = D^k S(t) * w_0 + \int_0^{t/2} (D^k S(t-s)) * g_{U_0}(s)\, ds + \int_{t/2}^t D^m S(t-s) * D^{k-m} g_{U_0}(s)\, ds,$$

where m is a multiindex such that $|m| = 1$.

It follows from (A-2) that $t \mapsto D^k S(t) * w_0 \in C^0(\mathbb{R}_+^*, L^2(\mathbb{R}^2))$. Next, arguing that $g_{U_0} \in C^0\big(\mathbb{R}_+, L^2(\mathbb{R}^2)\big)$ and using (A-2) with $r = 1$, we find

$$\left\| \int_0^{t/2} (D^k S(t-s)) * g_{U_0}(s)\, ds \right\|_{L^2} \leq C \int_0^{t/2} \frac{ds}{(t-s)^{|k|/2}} \leq \frac{C}{t^{(|k|/2)-1}}.$$

Also, since $|k-m| = |k|-1 \leq p$ and $w(s) \in H^p(\mathbb{R}^2)$ by assumption, Step 1 provides the decomposition

$$D^{k-m} g_{U_0}(s) = d^1(s) + d^2(s),$$

where d^1 belongs to $C^0(\mathbb{R}_+^*, L^2(\mathbb{R}^2))$ and d^2 to $C^0\big(\mathbb{R}_+^*, L^{4/3}(\mathbb{R}^2)\big)$. It follows from (A-2) that

$$\begin{aligned}\left\| \int_{t/2}^t D^m S(t-s) * D^{k-m} g_{U_0}(s)\, ds \right\|_{L^2} &\leq \int_{t/2}^t \big(\|\nabla S(t-s)\|_{L^1} \|d^1(s)\|_{L^2} + \|\nabla S(t-s)\|_{L^r} \|d^2(s)\|_{L^{4/3}} \big)\, ds \\ &\leq C(t) \int_{t/2}^t \big((t-s)^{-\frac{1}{2}} + (t-s)^{-\frac{1}{2}-1+\frac{1}{r}} \big)\, ds,\end{aligned}$$

where r satisfies $1 + \frac{1}{2} = \frac{1}{r} + \frac{3}{4}$. The last term is finite since $\frac{1}{2} + 1 - \frac{1}{r} = \frac{3}{4} < 1$, so we infer that $w \in C^0\big(\mathbb{R}_+^*, H^{p+1}(\mathbb{R}^2)\big)$, as we wanted.

Step 3. Let $w \in C^0(\mathbb{R}_+, H^1(\mathbb{R}^2))$ be the solution to (CGL). We show that $w \in C^k\big(\mathbb{R}_+^*, C^l(\mathbb{R}^2)\big)$ for all $k, l \in \mathbb{N}$.

Fix $k, l \in \mathbb{N}$. we show by induction on $0 \leq j \leq k$ that $w \in C^j\big(\mathbb{R}_+^*, C^{l+2k-2j}(\mathbb{R}^2)\big)$. This holds for $j = 0$ according to Step 2 and since H^p is embedded in C^{l+2k} for large enough p. We next assume that $w \in C^j\big(\mathbb{R}_+^*, C^{l+2k-2j}(\mathbb{R}^2)\big)$ for some $0 \leq j \leq k-1$, and it follows that

$$\Delta w,\ f_{U_0}(w) \in C^j\big(\mathbb{R}_+^*, C^{l+2k-2j-2}(\mathbb{R}^2)\big).$$

Going back to Equation (CGL), we obtain

$$w \in C^{j+1}\big(\mathbb{R}_+^*, C^{l+2k-2j-2}(\mathbb{R}^2)\big).$$

This concludes the proof of Proposition A.8. □

Acknowledgments

I warmly thank Didier Smets for his constant support during the preparation of this work. I am also indebted to Thierry Gallay and to Sylvia Serfaty for very helpful discussions.

References

[Almeida 1999] L. Almeida, "Threshold transition energies for Ginzburg–Landau functionals", *Nonlinearity* **12**:5 (1999), 1389–1414. MR 2000g:35190 Zbl 0953.35039

[Aranson and Kramer 2002] I. S. Aranson and L. Kramer, "The world of the complex Ginzburg–Landau equation", *Rev. Modern Phys.* **74**:1 (2002), 99–143. MR 2003h:35254

[Bethuel and Smets 2007] F. Bethuel and D. Smets, "A remark on the Cauchy problem for the 2D Gross–Pitaevskii equation with nonzero degree at infinity", *Differential Integral Equations* **20**:3 (2007), 325–338. MR 2008g:35177

[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, MA, 1994. MR 95c:58044 Zbl 0802.35142

[Bethuel et al. 2005] F. Bethuel, G. Orlandi, and D. Smets, "Collisions and phase-vortex interactions in dissipative Ginzburg–Landau dynamics", *Duke Math. J.* **130**:3 (2005), 523–614. MR 2007d:35140 Zbl 1087.35008

[Bethuel et al. 2007] F. Bethuel, G. Orlandi, and D. Smets, "Dynamics of multiple degree Ginzburg–Landau vortices", *Comm. Math. Phys.* **272**:1 (2007), 229–261. MR 2009f:35145 Zbl 1135.35014

[Bethuel et al. 2008] F. Bethuel, R. L. Jerrard, and D. Smets, "On the NLS dynamics for infinite energy vortex configurations on the plane", *Rev. Mat. Iberoam.* **24**:2 (2008), 671–702. MR MR2459209 Zbl pre05361879

[Colliander and Jerrard 1998] J. E. Colliander and R. L. Jerrard, "Vortex dynamics for the Ginzburg–Landau–Schrödinger equation", *Internat. Math. Res. Notices* **7** (1998), 333–358. MR 99c:35218 Zbl 0914.35128

[Colliander and Jerrard 1999] J. E. Colliander and R. L. Jerrard, "Ginzburg–Landau vortices: weak stability and Schrödinger equation dynamics", *J. Anal. Math.* **77** (1999), 129–205. MR 2001h:35170 Zbl 0933.35155

[Ginibre and Velo 1997] J. Ginibre and G. Velo, "The Cauchy problem in local spaces for the complex Ginzburg–Landau equation, II: Contraction methods", *Comm. Math. Phys.* **187**:1 (1997), 45–79. MR 98i:35174 Zbl 0889.35046

[Jerrard 1999] R. L. Jerrard, "Vortex dynamics for the Ginzburg–Landau wave equation", *Calc. Var. Partial Differential Equations* **9**:1 (1999), 1–30. MR 2000i:35189 Zbl 0941.35099

[Jerrard and Soner 1998] R. L. Jerrard and H. M. Soner, "Dynamics of Ginzburg–Landau vortices", *Arch. Rational Mech. Anal.* **142**:2 (1998), 99–125. MR 99m:58073 Zbl 0923.35167

[Jerrard and Spirn 2007] R. Jerrard and D. Spirn, "Refined Jacobian estimates for Ginzburg–Landau functionals", *Indiana Univ. Math. J.* **56**:1 (2007), 135–186. MR 2008h:35043 Zbl 1121.26008

[Jerrard and Spirn 2008] R. L. Jerrard and D. Spirn, "Refined Jacobian estimates and Gross–Pitaevsky vortex dynamics", *Arch. Ration. Mech. Anal.* **190**:3 (2008), 425–475. MR 2009i:35302 Zbl 1155.76010

[Kurzke et al. 2008] M. Kurzke, C. Melcher, R. Moser, and D. Spirn, "Dynamics for Ginzburg–Landau vortices under a mixed flow", preprint 428, Universität Bonn, SFB 611, 2008, Available at http://sfb611.iam.uni-bonn.de/uploads/428-komplett.pdf. to appear in *Indiana Univ. Math. J.*

[Lin and Xin 1999] F.-H. Lin and J. X. Xin, "On the incompressible fluid limit and the vortex motion law of the nonlinear Schrödinger equation", *Comm. Math. Phys.* **200**:2 (1999), 249–274. MR 2000g:35196 Zbl 0920.35145

[Sandier and Serfaty 2004] E. Sandier and S. Serfaty, "A product-estimate for Ginzburg–Landau and corollaries", *J. Funct. Anal.* **211**:1 (2004), 219–244. MR 2005e:35047 Zbl 1063.35144

[Serfaty 2007] S. Serfaty, "Vortex collisions and energy-dissipation rates in the Ginzburg–Landau heat flow, II: The dynamics", *J. Eur. Math. Soc.* **9**:3 (2007), 383–426. MR 2008g:35075

Received 21 Oct 2008. Accepted 27 Mar 2009.

EVELYNE MIOT: miot@ann.jussieu.fr
Laboratoire J.-L. Lions UMR 7598, Université Pierre et Marie Curie, 175 rue du Chevaleret, 75013 Paris, France

ANALYSIS AND PDE
Vol. 2, No. 2, 2009

ON THE GLOBAL WELL-POSEDNESS OF THE ONE-DIMENSIONAL SCHRÖDINGER MAP FLOW

IGOR RODNIANSKI, YANIR A. RUBINSTEIN AND GIGLIOLA STAFFILANI

We establish the global well-posedness of the initial value problem for the Schrödinger map flow for maps from the real line into Kähler manifolds and for maps from the circle into Riemann surfaces. This partially resolves a conjecture of W.-Y. Ding.

1. Introduction

In this article we study the Schrödinger map flow from a one-dimensional domain into a complete Kähler manifold. First, we show that when the domain is the real line the flow exists for all time. Second, we show that when the domain is the circle and the target is a Riemann surface the flow also exists for all time. The main contribution of this article is to bring Bourgain's work on the periodic cubic nonlinear Schrödinger equation (NLS) to bear on the geometric situation at hand.

Let (M, g) be a complete Riemannian manifold of dimension m, and let (N, ω, J, h) be a complete symplectic manifold of dimension $2n$ with a compatible almost complex structure J, that is, such that $\omega(J\,\cdot, J\,\cdot) = \omega(\cdot\,, \cdot)$ and such that $h(\cdot\,, \cdot) = \omega(\cdot\,, J\,\cdot)$ defines a complete Riemannian metric on N. Associated to this data is the space of all smooth maps from M to N, the Fréchet manifold $X := C^\infty(M, N)$, endowed with a symplectic structure,

$$\Omega(V, W)\big|_u = \int_M u^\star\omega(V, W)\, dV_{M,g} \quad \text{for all } V, W \in T_u X = \Gamma(M, u^\star TN),$$

where the tangent space to X at a map $u : M \to N$ is the space of smooth sections of $u^\star TN \to M$ and where $dV_{M,g}$ denotes the volume form on M induced by g. The form Ω is nondegenerate, i.e., it endows X with an injective map $TX \to T^\star X$.

Define the energy function on X by

$$E(u) = \tfrac{1}{2}\int_M |du|^2_{g^\sharp\otimes u^\star h}\, dV_{M,g},$$

where we denote by $g^\sharp$ the metric induced by g on $T^\star M$ and where we view du as a section of $T^\star M \otimes u^\star TN \to M$ and equip this bundle with the metric $g^\sharp \otimes u^\star h$.

The almost-complex structure on N induces one on X and a corresponding compatible Riemannian metric defined by

$$G(V, W)\big|_u = \int_M u^\star h(V, W)\, dV_{M,g} \quad \text{for all } V, W \in T_u X = \Gamma(M, u^\star TN).$$

MSC2000: primary 35Q55; secondary 53C44, 35B10, 32Q15, 42B35, 15A23.
Keywords: Schrödinger flow, periodic NLS, cubic NLS, Strichartz estimates, Kähler manifolds.

In the infinite-dimensional setting not every function will necessarily have a gradient. However, if we let $\{u_t\}_{t\in[-1,1]}$ be a smooth family of maps with $u_0 = u$ and denote by $W = (\partial u_t/\partial t)|_0$ a variation, then

$$dE(W)|_{u_0} = \frac{\partial E(u_t)}{\partial t}\bigg|_0 = \int_M g^\sharp \otimes u^\star h(du, dW)\, dV_{M,g} = -\int_M u^\star h(\mathrm{tr}_{g^\sharp} \nabla du, W)\, dV_{M,g}, \tag{1}$$

and hence the gradient of E exists and is given by

$$\nabla^G E\big|_u = -\tau(u), \tag{2}$$

where $\tau(u) := \mathrm{tr}_{g^\sharp} \nabla du$ is called the tension field of u and ∇ is the connection on $T^\star M \otimes u^\star TN \to M$ induced from the Levi-Civita connection on (M, g) and the pulled-back Levi-Civita connection from (N, h). The corresponding gradient flow

$$\frac{\partial u}{\partial t} = \tau(u), \quad u|_{\{0\}\times M} = u_0, \tag{3}$$

is the classical harmonic map flow introduced by Eells and Sampson [1964], which has been extensively studied.

Now the symplectic gradient of E also exists and is given by

$$\nabla^\Omega E\big|_u = -J\tau(u).$$

The corresponding Hamiltonian flow

$$\frac{\partial u}{\partial t} = -J\tau(u), \quad u|_{\{0\}\times M} = u_0, \tag{4}$$

on (X, Ω), introduced in [Ding and Wang 1998; Terng and Uhlenbeck 2006], is called the Schrödinger map flow.

While the energy decreases along (3), for (4) the flow is contained in an energy level set, since for maps of finite energy we have by (1)

$$\frac{dE(u(t))}{dt} = -\int_M u^\star h\Big(\tau(u), \frac{\partial u}{\partial t}\Big)\, dV_{M,g} = 0. \tag{5}$$

For (3) one typically expects to converge to a harmonic representative of the homotopy class of u_0 under some geometric assumptions (for example, negatively curved target [Eells and Sampson 1964]) while (4) seems to be describing some rather very different behavior. Analytically this may be described by the transition from the parabolic (3) to the borderline case (4) whose symbol has purely imaginary eigenvalues. Note also that for the Schrödinger flow there is no preferred time direction.

One problem common to both flows is the question of existence and uniqueness. Indeed since the flows are defined on infinite-dimensional spaces one cannot expect global existence[1] or well-posedness[2] in general. Restricting to the Kähler case, there is a similarity between the two flows as far as local existence is concerned: Results of Ding, Wang and McGahagan show that at least locally (4) can be approximated by equations of either parabolic (in the sense of Petrovskiĭ: see, for example, [Eidelman and Zhitarashu 1998]) or hyperbolic character. As a consequence the following result holds for maps of finite energy.

[1] For the Schrödinger flow the question of global existence is equivalent to the existence of a "symmetry" of (X, Ω), that is, a one-parameter subgroup of Hamiltonian diffeomorphisms of (X, Ω) for the energy function E integrating $\nabla^\Omega E$.

[2] Until recently no results were known for general symplectic targets; see [Chihara 2008] for recent work on local well-posedness in this setting.

Theorem 1.1 [Ding and Wang 2001; McGahagan 2007]. *Let (M^m, g) be a complete Riemannian manifold and let (N, J, h) be a complete Kähler manifold with bounded geometry. For integers $k > m/2+1$ the flow equation (4) with $u_0 \in W^{k,2}(M, N)$ admits a unique solution $u \in C^0\big([0, T], W^{k,2}(M, N)\big)$ where $T < T_0$ and T_0 depends on $\|\nabla u_0\|_{W^{[m/2]+1,2}}$ and the geometry of N alone. Moreover, there exist positive constants C_1, C_2 depending only on these quantities such that*

$$\|\nabla u(t)\|_{W^{[m/2]+1,2}} \le C_1/(T_0-t)^{C_2} \quad \text{for all } t \in [0, T_0).$$

In particular, if $u_0 \in W^{k,2}(M, N)$ for all $k \ge 2$ then $u \in C^\infty\big([0, T] \times M, N\big)$.

Here by bounded geometry we mean uniform bounds on the injectivity radius and the curvature tensor and its derivatives. This is automatically true for compact targets.

The main difficulty lies, therefore, in understanding the global behavior.

Previous results of a global nature are mostly concerned with the one-dimensional domain case and are all restricted to the case of a special target Kähler manifold. We recall the following nonexhaustive list of works. The flow on $(S^1, \text{can}) \to (S^2, \text{can})$, where "can" denotes the canonical metric, corresponding to the classical model for an isotropic ferromagnet was studied from the mathematical point of view by Sulem et al. [1986], who obtained local well-posedness for the initial value problem as well as partial global results. Zhou et al. [1991] studied the global well-posedness problem using a parabolic approximation which was later put to use in [Ding and Wang 1998; Pang et al. 2001] to prove global existence and uniqueness of smooth solutions of maps from (S^1, can) into a constant sectional curvature Kähler target (that is, a Riemann surface equipped with a constant curvature metric or a flat complex torus) as well as to Hermitian locally symmetric spaces [Pang et al. 2002] using a conservation law. The latter also treats the inhomogeneous flow which can be essentially viewed as the Schrödinger flow with domain S^1 equipped with a different metric. Terng and Uhlenbeck [2006] studied in detail the flow from the Euclidean line into Grassmannians. Chang et al. [2000] proved existence and uniqueness of global smooth solutions for maps of the Euclidean line into a compact Riemann surface. In addition, they treated maps of the Euclidean plane into a compact Riemann surface under the assumption of small initial energy and certain symmetries. Finally, see [Bejenaru et al. 2007; 2008] for recent work on global well-posedness in the case of maps from Euclidean space into (S^2, can) under a certain smallness assumption.

Note that in all of these results one restricts the target to a rather small class of Kähler manifolds.

We recall the following conjecture:

Conjecture 1.2 [Ding 2002]. *The Schrödinger map flow is globally well-posed for maps from one-dimensional domains into compact Kähler manifolds.*

The main results of this article are a partial answer to this conjecture. Namely, we establish the global well-posedness of the one-dimensional Schrödinger flow into general Kähler manifolds when the domain is the real line, and into Riemann surfaces when the domain is the circle.

Theorem 1.3. *Let $(M, g) = (\mathbb{R}, dx \otimes dx)$, let (N, J, h) be a complete Kähler manifold with bounded geometry, and let $k \ge 2$ be an integer. The flow equation (4) with $u_0 \in W^{k,2}(\mathbb{R}, N)$ admits a unique solution $u \in C^0(\mathbb{R}, W^{k,2}(\mathbb{R}, N))$. In particular u is smooth if u_0 is in $W^{k,2}(\mathbb{R}, N)$ for all $k \ge 2$.*

Theorem 1.4. *Let $(M, g) = (S^1, dx \otimes dx)$, let (N, J, h) be a complete Riemann surface with bounded geometry, and let $k \geq 2$ be an integer. The flow equation (4) with $u_0 \in W^{k,2}(S^1, N)$ admits a unique solution $u \in C^0(\mathbb{R}, W^{k,2}(S^1, N))$. In particular u is smooth if u_0 is in $W^{k,2}(S^1, N)$ for all $k \geq 2$.*

Remark 1.5. From the physical point of view, the Schrödinger map flow may also be introduced as a generalization of the Heisenberg model for a ferromagnetic spin system. The classical model for this physical system precisely corresponds to maps from the standard circle into $N = S^2$ with the standard metric and complex structure [Landau and Lifschitz 1935] (for some background see, for example, [Ding 2002; Ding and Wang 1998; McGahagan 2004; Sulem et al. 1986]). Perhaps the most physically natural generalization of the classical model would be to vary the metric on the target S^2, however it seems that even for small perturbations of the round metric on S^2 global well-posedness was not known before. Theorem 1.4 establishes the global well-posedness of the Cauchy problem describing this physical model when the metric on S^2 is arbitrary.

The global well-posedness thus established in these cases, several natural questions arise related to more precise information regarding the long-time behavior of the Schrödinger map flow. For example, for the case of maps from the real line it would be interesting to determine whether certain scattering occurs in some cases. In addition, we pose the following conjecture regarding the length of the image along the flow.

Conjecture 1.6. *In the setting of Theorem 1.3 one has* $\lim_{t\to\infty} \|u(t)\|_{W^{1,1}(\mathbb{R},N)} = \infty$. *In addition, for every* $\epsilon > 0$, *there exists a time* t_0 *and a geodesic ball* $B \subset N$ *of radius* ϵ *such that the image of* $u(t_0)$ *is contained in* B.

Is not hard to show this conjecture holds for the case $N = \mathbb{C}^n$, equipped with the Euclidean metric, with an estimate $\|u(t)\|_{W^{1,1}(\mathbb{R},N)} \geq Ct^{1/2}$.

Outline of proofs and organization of the paper. According to Theorem 1.1 we have existence of a time-local solution. The strategy of the proof is this: First, using the Kähler condition, we translate the flow equation into a system of nonlinear Schrödinger (NLS) equations. Then, for this system of equations we obtain an a priori estimate in a weaker norm than that in Theorem 1.1, namely in an appropriate Strichartz norm for $M = \mathbb{R}$ and in L^4 for $M = S^1$. These estimates are crucial since they only depend on the initial energy (which is a conserved quantity) and that in a manner that can be readily converted into a global a priori estimate in the same space. Taking derivatives of the flow equation and after additional work we then obtain global a priori estimates in stronger norms and these in turn may be converted back to imply global well-posedness for our original Cauchy problem in $W^{k,2}$ for all $k \geq 2$.

While the proofs of both Theorem 1.3 and Theorem 1.4 follow the same general scheme, nevertheless there are substantial differences between the two, as we now explain.

We start in Section 2 with the case of the real line, which is simpler due to simple connectivity and dispersiveness. Here we follow Chang, Shatah and Uhlenbeck [2000] and write the flow equation in terms of a parallel frame, with the added observation that the Kähler condition allows one to readily generalize their computations from the Riemann surface case to higher dimensions. The flow equation then reduces to a system of NLS equations. The same Strichartz-type calculations as in their study of the Riemann surface case then apply.

We then treat in Section 3 the case of maps from $M = S^1$ into a Riemann surface, which is considerably more difficult and is the main contribution of this article. There are two main difficulties. First, using a parallel frame introduces holonomy, so the resulting NLS equation lives on $\mathbb{R}$, the universal cover of S^1, instead of on S^1 itself. To overcome this we use a certain space-time transformation in order to obtain an NLS equation on S^1 in terms of the holonomy representation of N. In addition we need to estimate the variation of the holonomy along the flow. Second, since S^1 is compact the equations are no longer dispersive. To overcome that we adapt Bourgain's results on the cubic NLS to our setting in order to prove a time-local a priori estimate in L^4 that depends in such a way on the initial data that it may be used to obtain a global a priori estimate in the same space. Finally, we take derivatives of the NLS equation and after some more work obtain higher derivative a priori estimates.

In Section 4 we discuss some of the difficulties that arise when trying to apply our approach to treat maps from the circle into higher-dimensional Kähler manifolds. It is conceivable that some of these ideas might be related to showing finite time blow-up for higher-dimensional domains.

2. Maps from the real line into a Kähler manifold

In the case of maps from the Euclidean real line into a complete Kähler manifold (N, J, h) of complex dimension n, the Schrödinger equation (4) becomes

$$J\nabla_t u - \nabla_x \nabla_x u = 0, \qquad u(0) = u_0, \tag{6}$$

and we define the energy as

$$E(u_0) = \int_{\mathbb{R}} |du_0|^2_{\frac{\partial}{\partial x}\otimes\frac{\partial}{\partial x}\otimes u^\star h}\, dx < \infty \tag{7}$$

(departing from the convention in the Introduction by a factor of 2). Here we have used the abbreviated notation $\nabla_t = \nabla_{u_\star \partial/\partial t}$, $\nabla_x = \nabla_{u_\star\partial/\partial x}$, $\nabla_t u = u_\star \partial/\partial t = \partial u/\partial t$, $\nabla_x u = u_\star \partial/\partial x = \partial u/\partial x$, and we denote the derivatives of a function f by $f_{,x}$ and $f_{,t}$. The key idea in this section, going back to [Chang et al. 2000], is to rewrite (6) in an appropriate frame along the image, in such a way that (6) reduces to a system of nonlinear Schrödinger (NLS) equations. In fact our proof closely follows their approach for the Riemann surface case observing that it readily generalizes to Kähler targets of arbitrary dimension.

Assume that $u : I \times \mathbb{R} \to N$ is a solution of (6), where I is a neighborhood of 0 in $\mathbb{R}$ (given, for example, by Theorem 1.1). Choose an orthonormal frame $\{e_1, \dots, e_{2n}\}$ for $u^\star TN$ with respect to h. We further reduce the structure group to $U(n) \subseteq O(2n)$ by assuming $e_{n+1} = Je_1, \dots, e_{2n} = Je_n$. We identify $U(n)$ with its image in $O(2n)$ under the map $\imath : \mathrm{GL}(n, \mathbb{C}) \to \mathrm{GL}(2n, \mathbb{R})$ given by

$$\imath(A + \sqrt{-1}\, B) = \begin{pmatrix} A & -B \\ B & A \end{pmatrix}.$$

Note that if $v = x + \sqrt{-1}\, y \in \mathbb{C}^n$ and $\imath(v) = \begin{pmatrix} x \\ y \end{pmatrix}$ then

$$\imath(Av) = \imath(A)\imath(v).$$

We will use this identification frequently, sometimes omitting the reference to the map $\imath$. In the following we let Latin indices take values in $\{1, \dots, 2n\}$ and Greek indices in $\{1, \dots, n\}$. For both alphabets we use the notation

$$\bar{\cdot} = \cdot + n - 1 \pmod{2n} + 1.$$

Therefore barred Greek indices take values in $\{n+1,\dots,2n\}$. As mentioned $e_{\bar{j}} = Je_j$ so $e_{\bar{\alpha}} = Je_\alpha$ and $e_{\overline{\alpha+n}} = -e_\alpha$. We abbreviate the spaces $L^p(\mathbb{R},dx)$ and $L^p(\mathbb{R},dt)$ as $L^p(\mathbb{R}_x)$ and $L^p(\mathbb{R}_t)$, and so on. Finally, given a map $u:(M,g)\to(N,h)$, by $u\in W^{k,p}(M,N)$ we will mean that $\sum_{j=0}^{k-1}\big\||\nabla^j du|\big\|_{L^p}<\infty$. For example, in this notation we have $E(u)=\|u\|^2_{W^{1,2}(M,N)}$.

Now we may view the flow equation (6) in this frame. Write, for each $(t,x)\in I\times\mathbb{R}$,

$$\nabla_x u = \sum_{j=1}^{2n} h(\nabla_x u, e_j)e_j =: a^j e_j, \quad \nabla_t u = \sum_{j=1}^{2n} h(\nabla_t u, e_j)e_j =: b^j e_j, \tag{8}$$

where we use the Einstein summation convention, namely the appearance of an index both as a subscript and a superscript indicates summation. Then (6) can be rewritten as

$$b^j e_{\bar{j}} - a^j_{,x} e_j - a^j \nabla_x e_j = 0. \tag{9}$$

The conservation of energy (see (5)) is expressed as

$$E(u(t)) = E(u_0) = \int_{\mathbb{R}} \sum_{l=1}^{2n} (a_l)^2 dx = \|a(t)\|^2_{L^2(\mathbb{R}_x)} \quad \text{for all } t\in\mathbb{R}. \tag{10}$$

Note that

$$0 = u_\star\Big[\frac{\partial}{\partial t}, \frac{\partial}{\partial x}\Big] = [\nabla_t u, \nabla_x u]. \tag{11}$$

Since $\nabla J = 0$, differentiating (6) in space yields $J\nabla_x\nabla_t u - \nabla_x\nabla_x\nabla_x u = 0$, which becomes, using (11) and (8),

$$J(a^j_{,t}e_j + a^j\nabla_t e_j) - a^j_{,xx}e_j - 2a^j_{,x}\nabla_x e_j - a^j\nabla_x\nabla_x e_j = 0. \tag{12}$$

We impose the gauge-fixing condition

$$\nabla_x e_j = 0, \quad j = 1,\dots,2n. \tag{13}$$

The resulting frame along the image is still unitary, since the complex structure commutes with parallel transport. Equation (9) becomes

$$b^j = a^{\bar{j}}_{,x}. \tag{14}$$

Note that $u^\star TN \to \mathbb{R}$ is trivial and that (13) amounts to fixing a trivializing parallel frame. With this choice, the flow on $u^\star TN$ is given by

$$a^j_{,t} e_{\bar{j}} - a^j_{,xx} e_j = -a^j \nabla_t e_{\bar{j}}. \tag{15}$$

Along the image, using (13) and (14), and letting R denote the curvature tensor of (N,h), we have

$$\begin{aligned}
\nabla_x\nabla_t e_{\bar{j}} &= R(\nabla_x u, \nabla_t u)e_{\bar{j}} = a^k b^l R_{kl\bar{j}}{}^q e_q\\
&= \big(a^\alpha b^\beta R_{\alpha\beta\bar{j}}{}^q + a^{\bar\alpha} b^\beta R_{\bar\alpha\beta\bar{j}}{}^q + a^\alpha b^{\bar\beta} R_{\alpha\bar\beta\bar{j}}{}^q + a^{\bar\alpha} b^{\bar\beta} R_{\bar\alpha\bar\beta\bar{j}}{}^q\big)e_q\\
&= \big(a^\alpha a^{\bar\beta}_{,x} R_{\alpha\beta\bar{j}}{}^q + a^{\bar\alpha} a^{\bar\beta}_{,x} R_{\bar\alpha\beta\bar{j}}{}^q - a^\alpha a^\beta_{,x} R_{\alpha\bar\beta\bar{j}}{}^q - a^{\bar\alpha} a^\beta_{,x} R_{\bar\alpha\bar\beta\bar{j}}{}^q\big)e_q\\
&= \sum_{\alpha,\beta}\big[(a^\alpha a^{\bar\beta})_{,x} R_{\alpha\beta\bar{j}}{}^q + \tfrac{1}{2}[(a^{\bar\alpha}a^{\bar\beta})_{,x} + (a^\alpha a^\beta)_{,x}]R_{\bar\alpha\beta\bar{j}}{}^q\big]e_q,
\end{aligned} \tag{16}$$

where we have used the Kähler condition once more:

$$R_{\alpha\beta\bar{J}}{}^{q} = R_{\bar{\alpha}\bar{\beta}\bar{J}}{}^{q}, \quad R_{\bar{\alpha}\beta\bar{J}}{}^{q} = -R_{\alpha\bar{\beta}\bar{J}}{}^{q}.$$

Equation (7) implies $\lim_{x\to\pm\infty} a^i(t,x) = 0$. Therefore, since $\nabla_x h(\nabla_t e_{\bar{J}}, e_q) = h(\nabla_x \nabla_t e_{\bar{J}}, e_q)$, we have

$$\begin{aligned}\nabla_t e_{\bar{J}}(t,x) = \sum_{q=1}^{2n} h(\nabla_t e_{\bar{J}}, e_q)(t,-\infty) e_q(t,x) \\ + \Bigg(\sum_{\alpha,\beta} \big[a^\alpha a^{\bar{\beta}} R_{\alpha\beta\bar{J}}{}^{q} + \tfrac{1}{2}[a^{\bar{\alpha}} a^{\bar{\beta}} + a^\alpha a^\beta] R_{\bar{\alpha}\beta\bar{J}}{}^{q}\big](t,x) \\ - \int_{(-\infty,x]} \sum_{\alpha,\beta} \big[a^\alpha a^{\bar{\beta}} R_{\alpha\beta\bar{J}}{}^{q}{}_{,x} + \tfrac{1}{2}(a^{\bar{\alpha}} a^{\bar{\beta}} + a^\alpha a^\beta) R_{\bar{\alpha}\beta\bar{J}}{}^{q}{}_{,x}\big](t,y)\,dy \Bigg) e_q(t,x). \end{aligned} \tag{17}$$

Defining

$$\begin{aligned} A^q_{\bar{J}}(t,-\infty) &:= h(\nabla_t e_{\bar{J}}, e_q)(t,-\infty), \\ P^q_{\bar{J}}(t,x) &:= \sum_{\alpha,\beta} \big[a^\alpha a^{\bar{\beta}} R_{\alpha\beta\bar{J}}{}^{q} + \tfrac{1}{2}[a^{\bar{\alpha}} a^{\bar{\beta}} + a^\alpha a^\beta] R_{\bar{\alpha}\beta\bar{J}}{}^{q}\big](t,x), \\ Q^q_{\bar{J}}(t,x) &:= -\int_{(-\infty,x]} \sum_{\alpha,\beta} \big[a^\alpha a^{\bar{\beta}} R_{\alpha\beta\bar{J}}{}^{q}{}_{,x} + \tfrac{1}{2}(a^{\bar{\alpha}} a^{\bar{\beta}} + a^\alpha a^\beta) R_{\bar{\alpha}\beta\bar{J}}{}^{q}{}_{,x}\big](t,y)\,dy, \end{aligned}$$

we thus have

$$\nabla_t e_{\bar{J}}(t,x) = A^q_{\bar{J}}(t,-\infty) e_q(t,x) + \big[P^q_{\bar{J}}(t,x) + Q^q_{\bar{J}}(t,x)\big] e_q(t,x). \tag{18}$$

We now estimate these terms. Using (9) we have

$$\nabla_t e_{\bar{J}} = b^k \Gamma^p_{k\bar{J}} e_p = a^{\bar{k}}_{,x} \Gamma^p_{k\bar{J}} e_p. \tag{19}$$

Hence, $h(\nabla_t e_{\bar{J}}, e_q) = a^{\bar{k}}_{,x} \Gamma^q_{k\bar{J}}$ and so we may assume that $A^q_{\bar{J}}$ vanishes at $(t,-\infty)$. To justify this, note that this is indeed the case for the local solution of our equation given by Theorem 1.1; even though this assumption makes use of the finiteness of the $W^{2,2}$ norm of that local solution, the important point is that eventually our estimates will not depend on the $W^{2,2}$ norm of u (equivalently on the $W^{1,2}$ norm of a), and so the proof of the a priori estimate for the system of NLS equations (23) below (for a) goes through, with this assumption. Next, using (8), note that $R_{klp}{}^{q}{}_{,x} = a^s R_{klp}{}^{q}{}_{,s}$. Therefore,

$$|P^q_j(t,x)| < C|\mathrm{a}(t,x)|^2, \qquad |Q^q_j(t,x)| < C\int_{\mathbb{R}} |\mathrm{a}(t,y)|^3 dy, \tag{20}$$

where $C > 0$ depends only on the geometry and where we use the notation $|\mathrm{a}| := \big(\sum_{j=1}^{2n} (a^j)^2\big)^{1/2}$.

To summarize the discussion, we have shown that (15) transforms to the following system of NLS equations

$$-a^{\bar{\gamma}}_{,t} - a^{\gamma}_{,xx} = -a^j P^{\gamma}_j - a^j Q^{\gamma}_j, \qquad \gamma = 1,\dots,n, \tag{21}$$

$$a^{\gamma}_{,t} - a^{\bar{\gamma}}_{,xx} = -a^j P^{\bar{\gamma}}_j - a^j Q^{\bar{\gamma}}_j, \qquad \gamma = 1,\dots,n, \tag{22}$$

or, letting $J_0 = \imath(\sqrt{-1}\, I)$,

$$J_0 \mathrm{a}_{,t} = \mathrm{a}_{,xx} - \mathrm{P}\cdot\mathrm{a} - \mathrm{Q}\cdot\mathrm{a}, \tag{23}$$

where $\mathrm{a} = (a^1, \dots, a^{2n})^T$, $\mathrm{P} = (P^k_j)$, and $\mathrm{Q} = (Q^k_j)$. Equivalently, using the aforementioned identification $\imath$ of $\mathrm{GL}(n, \mathbb{C})$ with a subset of $\mathrm{GL}(2n, \mathbb{R})$,

$$\sqrt{-1}\Phi_{,t} = \Phi_{,xx} - \mathrm{S} \cdot \Phi - \mathrm{T} \cdot \Phi, \tag{24}$$

where

$$\Phi := \imath^{-1}(\mathrm{a}) = (a^1 + \sqrt{-1}\, a^{\bar{1}}, \dots, a^n + \sqrt{-1}\, a^{\bar{n}})^T, \quad \mathrm{S} := (S^\beta_\alpha) = \imath^{-1}(\mathrm{P}), \quad \mathrm{T} := (T^\beta_\alpha) = \imath^{-1}(\mathrm{Q}), \tag{25}$$

and from (20) we have

$$|S^\beta_\alpha| < C|\Phi|^2, \qquad |T^\beta_\alpha| < C \int_{\mathbb{R}} |\Phi|^3 dy. \tag{26}$$

Here we have set $|\Phi| := \left(\sum_{j=1}^n |\Phi^j|^2\right)^{1/2}$.

Remark 2.1. In the case of a variable complete smooth metric on the domain $(M, g) = (\mathbb{R}, \alpha^{-1} dx \otimes dx)$ with $\alpha > 0$ the flow equation (4) becomes

$$b^k e_{\bar{k}} = \alpha a^k_{,x} e_k + \tfrac{1}{2}\alpha_{,x} a^k e_k, \tag{27}$$

which can then be transformed, as before, to

$$J_0 \mathrm{a}_{,t} = \alpha \mathrm{a}_{,xx} + \tfrac{3}{2}\alpha_{,x}\mathrm{a}_{,x} + \tfrac{1}{2}\alpha_{,xx}\mathrm{a} - \mathrm{P} \cdot \mathrm{a} - \mathrm{Q} \cdot \mathrm{a}. \tag{28}$$

Equivalently, again using the map $\imath : \mathrm{GL}(n, \mathbb{C}) \to \mathrm{GL}(2n, \mathbb{R})$,

$$\sqrt{-1}\Phi_{,t} = \alpha\Phi_{,xx} + \tfrac{3}{2}\alpha_{,x}\Phi_{,x} + \tfrac{1}{2}\alpha_{,xx}\Phi - \mathrm{S} \cdot \Phi - \mathrm{T} \cdot \Phi. \tag{29}$$

The only obstacle to treating this equation using the methods below is the first derivative term on the right-hand side (cf. Remark 3.1).

Therefore we have reduced the original flow equation for the map to a system of NLS equations for the frame coefficients of the gradient of the map. Thus we have reduced ourselves to the same situation as in [Chang et al. 2000] (the only difference in (24) from the case where the target is a Riemann surface is that the equation for each Φ^j depends also on the other Φ^k, $k = 1, \dots, n$; however this dependence is only in the nonlinear terms and not in the terms involving derivatives) and their work now implies the following theorem which is the main result of this section.

Theorem 2.2. *Let $(M, g) = (\mathbb{R}, dx \otimes dx)$ and let (N, J, h) be a complete Kähler manifold with bounded geometry. Then for integers $k \geq 2$ the flow equation (4) with $u_0 \in W^{k,2}(\mathbb{R}, N)$ admits a unique solution $u \in C^0(\mathbb{R}, W^{k,2}(\mathbb{R}, N))$.*

For the benefit of the reader that may not be familiar with standard Strichartz estimates techniques we include here the detailed proof of the Chang–Shatah–Uhlenbeck $L^4(\mathbb{R}_{t,loc}, L^\infty(\mathbb{R}_x))$ estimate and how it implies global well-posedness in $W^{k,2}(\mathbb{R}, N)$. No originality is claimed here. This also serves to provide some perspective on the differences between this case and the case of the circle, treated in the following sections. In addition, it serves to explain the three basic steps in obtaining global well-posedness in $W^{k,2}$ that are also (at least schematically) needed in the case of the circle.

Proof. First, we have by Theorem 1.1 local well-posedness of the original Schrödinger map flow (6) in $W^{k,2}(\mathbb{R}, N)$ for $k \geq 2$. The key to obtaining global well-posedness in these spaces will be a local a priori estimate in a space that is morally larger than $C^0(\mathbb{R}, W^{2,2}(\mathbb{R}, N))$. Equivalently, we will prove an estimate for the frame coefficients a in a norm "weaker" than $C^0(\mathbb{R}, W^{1,2}(\mathbb{R}, \mathbb{R}^{2n})) \equiv C^0(\mathbb{R}, W^{1,2}(\mathbb{R}))$.

More precisely, the proof of global well-posedness is divided into three steps:

First, given an initial data u_0 in $W^{2,2}(\mathbb{R}, N)$ (equivalently, $\Phi(0) \in W^{1,2}(\mathbb{R})$) we will prove an a priori estimate on the $L^4(\mathbb{R}_{t,loc}, L^\infty(\mathbb{R}))$ norm of the frame coefficients Φ depending only on the initial energy $\|\Phi\|_{L^2(\mathbb{R})}$ and the geometry. In particular it will imply that $\|\Phi\|_{L^4([0,T],L^\infty(\mathbb{R}))}$ is finite for all $T > 0$. This is the most fundamental step.

Second, taking a derivative of the system of NLS equations (24) for Φ, using the estimate from the first step, and applying similar calculations we prove that $\|\Phi\|_{L^4([0,T],W^{1,\infty}(\mathbb{R}))}$ and $\|\Phi\|_{C^0([0,T],W^{1,2}(\mathbb{R}))}$ are finite for all $T > 0$.

Third, we let $k \geq 3$ and assume our initial data u_0 lies in $W^{k,2}(\mathbb{R}, N)$. Taking further derivatives of the equations (24) and working inductively, one proves that $\|\Phi\|_{C^0([0,T],W^{k-1,2}(\mathbb{R}))}$ is finite for all $T > 0$. This step, sometimes called *propagation of regularity*, is considered as routine once the first two steps have been carried out.

The key feature of the analysis involved here is that while one is interested only in proving that the $W^{1,2}(\mathbb{R})$ norm of $\Phi(t)$ stays finite, one is forced to use the auxiliary space $L^4(\mathbb{R}_t, W^{1,\infty}(\mathbb{R}_x))$.

In fact, although we will not carry this out here, in the first step one may prove local (and hence global) a priori estimates for (24) in other Strichartz spaces (these are by definition the spaces $L^q(\mathbb{R}_t, L^r(\mathbb{R}_x))$ specified by Lemma 2.3 below) as well, for example, $L^6(\mathbb{R} \times \mathbb{R})$.

Having thus outlined the different steps of the proof, we now turn to the proof itself.

Step 1. Suppose a function $c : \mathbb{R} \times \mathbb{R} \to \mathbb{C}$ satisfies the NLS equation

$$\sqrt{-1}\, c_{,t} = c_{,xx} + F \quad \text{for all } t \in [0, T], \qquad c(0) = f, \tag{30}$$

for some function $F : [0, T] \times \mathbb{R} \to \mathbb{C}$ that may depend on c nonlinearly (but not on its derivatives). One then has the integral expression (Duhamel formula)

$$c(t, x) = \int_{\mathbb{R}} f(y) \frac{e^{-\sqrt{-1}|x-y|^2/4t}}{\sqrt{2\pi t}} dy - \sqrt{-1} \int_0^t \int_{\mathbb{R}} F(s, y) \frac{e^{-\sqrt{-1}|x-y|^2/4(t-s)}}{\sqrt{2\pi (t-s)}} dy \wedge ds. \tag{31}$$

Denote the Schrödinger operator by

$$S(t) f := e^{-\sqrt{-1}\, t \partial^2/\partial x^2}; \tag{32}$$

more explicitly, we have for $M = \mathbb{R}$,

$$S(t) f = \int_{\mathbb{R}} f(y) \frac{e^{-\sqrt{-1}|x-y|^2/4t}}{\sqrt{2\pi t}} dy. \tag{33}$$

We now recall the Strichartz estimates (on $\mathbb{R}$). For appropriate q, r we denote by $L^q(\mathbb{R}, L^r(\mathbb{R}))$ the Banach space equipped with the norm

$$\|f\|_{L^q(\mathbb{R},L^r(\mathbb{R}))} := \big\| \|f\|_{L^r(\mathbb{R}_x)} \big\|_{L^q(\mathbb{R}_t)}. \tag{34}$$

Lemma 2.3 [Cazenave 2003, page 33]. *Let q, r satisfy $\frac{2}{q}+\frac{1}{r}=\frac{1}{2}$ with $r\in[2,\infty]$ and let $f\in L^2(\mathbb{R}_x)$. Then the function $t\mapsto S(t)f$ belongs to $L^q(\mathbb{R}_t, L^r(\mathbb{R}_x))\cap C^0(\mathbb{R}_t, L^2(\mathbb{R}_x))$ and there is a constant independent of (q,r) and of $f\in L^2(\mathbb{R})$ such that*

$$\|S(\cdot)f\|_{L^q(\mathbb{R}_t, L^r(\mathbb{R}_x))}\le C\|f\|_{L^2(\mathbb{R})}. \tag{35}$$

In our situation we know that $\|\Phi\|_{L^2}$ is constant in time (recall (10)). Assume also that Φ lies in $L^4([0,T], L^\infty(\mathbb{R}_x))$. We will now show that the $L^4([0,T], L^\infty(\mathbb{R}_x))$ norm of Φ is controlled by its L^2 norm and the geometry. This will imply local and eventually global estimates in $L^4(\mathbb{R}_t, L^\infty(\mathbb{R}_x))$.

Let $F=-\mathrm{S}\cdot\Phi-\mathrm{T}\cdot\Phi$. In what follows we restrict t to the interval $[t_1,t_2]$. Then

$$\Phi^j(t)=S(t-t_1)\Phi^j(t_1)-\sqrt{-1}\int_{t_1}^t S(t-s)F(s,\cdot)\,ds. \tag{36}$$

The first term of (36) is in $L^4([t_1,t_2], L^\infty(\mathbb{R}))$ by the Strichartz estimate (35). We will now show that the second term is also in this space.

First, we consider the term $\mathrm{S}\cdot\Phi\le C\|\Phi\|^3$. We need to estimate

$$\left\|\int_{t_1}^t S(t-s)(\mathrm{S}\cdot\Phi)(s,\cdot)ds\right\|_{L^4([t_1,t_2],L^\infty(\mathbb{R}))}. \tag{37}$$

From (33) we have the dispersive estimate $|S(t)f|\le Ct^{-1/2}\|f\|_{L^1(\mathbb{R})}$. Hence,

$$\begin{aligned}\left\|\int_{t_1}^t S(t-s)(\mathrm{S}\cdot\Phi)(s,\cdot)ds\right\|_{L^\infty(\mathbb{R})} &\le C\int_{t_1}^t (t-s)^{-1/2}\left\||\Phi(s,\cdot)|^3\right\|_{L^1(\mathbb{R})}ds\\ &\le C\cdot E(u_0)\int_{t_1}^t (t-s)^{-1/2}\|\Phi(s,\cdot)\|_{L^\infty(\mathbb{R})}ds\\ &\le C'\left(\int_{t_1}^t ((t-s)^{-1/2})^{4/3}ds\right)^{3/4}\left(\int_{t_1}^t \|\Phi(s,\cdot)\|^4_{L^\infty(\mathbb{R})}ds\right)^{1/4}\\ &= C''|t-t_1|^{1/4}\|\Phi\|_{L^4([t_1,t],L^\infty(\mathbb{R}))},\end{aligned} \tag{38}$$

and it follows that

$$\left\|\int_{t_1}^t S(t-s)(\mathrm{S}\cdot\Phi)(s,\cdot)ds\right\|_{L^4([t_1,t_2],L^\infty(\mathbb{R}))}\le C|t_2-t_1|^{1/2}\|\Phi\|_{L^4([t_1,t_2],L^\infty(\mathbb{R}))}. \tag{39}$$

Next, we consider the term $\mathrm{T}\cdot\Phi\le C\|\Phi\|\int_{\mathbb{R}}\|\Phi\|^3dx$. By applying a Strichartz estimate under the integral sign and using energy conservation we obtain

$$\begin{aligned}&\left\|\int_{t_1}^t S(t-s)(\mathrm{T}\cdot\Phi)(s,\cdot)\,ds\right\|_{L^4([t_1,t_2],L^\infty(\mathbb{R}))}\\ &\qquad\le C'\int_{t_1}^t\left\|\Phi\left\||\Phi|^3\right\|_{L^1(\mathbb{R})}\right\|_{L^2(\mathbb{R})}ds\le C'\int_{t_1}^t\left\|\Phi\|\Phi\|_{L^\infty(\mathbb{R})}\|\Phi\|^2_{L^2(\mathbb{R})}\right\|_{L^2(\mathbb{R})}ds\\ &\qquad\le C''\int_{t_1}^{t_2}\|\Phi\|_{L^\infty(\mathbb{R})}ds\le C''|t_2-t_1|^{3/4}\|\Phi\|_{L^4([t_1,t_2],L^\infty(\mathbb{R}))}.\end{aligned} \tag{40}$$

Combining (39) and (40) we thus obtain, by choosing $|t_2 - t_1|$ small enough (depending only on the initial energy and the geometry of (N, h)), an estimate on $\|\Phi\|_{L^4([t_1,t_2],L^\infty(\mathbb{R}))}$, depending only on the geometry of (N, h) and the initial energy. This then implies that for all $T > 0$ we have the a priori estimate

$$\|\Phi\|_{L^4([0,T],L^\infty(\mathbb{R}_x))} < \infty.$$

Step 2. To prove global well-posedness of the flow equation (6) in $W^{2,2}(\mathbb{R}, N)$, as outlined earlier, one differentiates (24) and follows similar computations as above. The main difference from Step 1 is that now the $L^2(\mathbb{R}_x)$ norm of $\Phi_{,x}$ is no longer preserved and one needs to work in the intersection of the spaces $C^0([0, T], W^{1,2}(\mathbb{R}_x))$ and $L^4([0, T], W^{1,\infty}(\mathbb{R}_x))$. Nevertheless the nonlinearity becomes milder after differentiation and can be readily controlled using the estimate from Step 1. We carry out the details for the sake of completeness.

The differentiated equation takes the form $\sqrt{-1}(\Phi_{,x})_{,t} = (\Phi_{,x})_{,xx} + G$, where

$$|G| < C\big(|\Phi|^4 + |\Phi_{,x}|\big(|\Phi|^2 + \|\Phi\|^3_{L^3(\mathbb{R}_x)}\big)\big).$$

Using the Duhamel formula (31) we have

$$\|\Phi_{,x}\|_{C^0([t_1,t_2],L^2(\mathbb{R}_x))} < C\|\Phi_{,x}(t_1)\|_{L^2(\mathbb{R}_x)} + \|G\|_{L^1([t_1,t_2],L^2(\mathbb{R}_x))}, \tag{41}$$

and using the Duhamel formula together with a Strichartz estimate we have

$$\|\Phi_{,x}\|_{L^4([t_1,t_2],L^\infty(\mathbb{R}_x))} < C\|\Phi_{,x}(t_1)\|_{L^2(\mathbb{R}_x)} + \|G\|_{L^1([t_1,t_2],L^2(\mathbb{R}_x))}. \tag{42}$$

Now, we have rather large freedom in estimating G. For example,

$$\||\Phi|^4\|_{L^1([t_1,t_2],L^2(\mathbb{R}_x))} = \|\Phi\|^4_{L^4([t_1,t_2],L^8(\mathbb{R}_x))} \le \|\Phi\|^4_{L^4([t_1,t_2],L^\infty(\mathbb{R}_x))}, \tag{43}$$

with the latter uniformly bounded from Step 1, while

$$\begin{aligned}\big\||\Phi_{,x}||\Phi|^2\big\|_{L^1([t_1,t_2],L^2(\mathbb{R}_x))} &\le E(u_0)^{1/2}\int_{t_1}^{t_2}\|\Phi_{,x}\|_{L^\infty(\mathbb{R})}\|\Phi\|_{L^\infty(\mathbb{R})}ds\\ &\le C\|\Phi\|_{L^4([t_1,t_2],L^\infty(\mathbb{R}_x))}\|\Phi_{,x}\|_{L^{4/3}([t_1,t_2],L^\infty(\mathbb{R}_x))}\\ &\le C\|\Phi\|_{L^4([t_1,t_2],L^\infty(\mathbb{R}_x))}\|\Phi_{,x}\|_{L^4([t_1,t_2],L^\infty(\mathbb{R}_x))}|t_2-t_1|^{1/2},\end{aligned} \tag{44}$$

and the same calculation applies also to the term $|\Phi_{,x}|\cdot\|\Phi\|^3_{L^3(\mathbb{R})}$. Plugging (43) and (44) back into (42) and choosing $|t_2 - t_1|$ small enough (depending only on the energy and the geometry (here we are using the uniform local estimate found in Step 1)) we obtain an estimate on $\|\Phi_{,x}\|_{L^4([t_1,t_2],L^\infty(\mathbb{R}_x))}$ in terms of $\|\Phi_{,x}(t_1)\|_{L^2(\mathbb{R}_x)}$, the energy, and the geometry. Using this back in (41) we then obtain an estimate on $\|\Phi_{,x}\|_{C^0([t_1,t_2],L^2(\mathbb{R}_x))}$ in terms of $\|\Phi_{,x}(t_1)\|_{L^2(\mathbb{R}_x)}$. This then implies that $\|\Phi_{,x}(t)\|_{L^2(\mathbb{R}_x)}$ increases at most exponentially in t, in particular remains finite for all $t > 0$, as desired. More precisely, for all $T > 0$,

$$\|\Phi_{,x}\|_{C^0([0,T],L^2(\mathbb{R}_x))} < C'e^{CT}, \tag{45}$$

for some uniform constants $C, C' > 0$.

Step 3. The higher derivatives estimates follow similar computations; this step is commonly called *propagation of regularity*. Essentially, each time the system of NLS equations are differentiated we obtain a new system of NLS equations where the nonlinearity is milder than in the previous stage. Hence, for initial data $u_0 \in W^{k,2}(\mathbb{R}, N)$, inductively using the estimates from the previous $k-1$ systems of equations yields an a priori estimate on $\|\Phi\|_{C^0([0,T],W^{k-1,2}(\mathbb{R}))}$ for each $k \geq 3$, again by using the auxiliary spaces $L^4([0, T], W^{k-1,\infty}(\mathbb{R}))$. This concludes the proof of Theorem 2.2. □

Remark 2.4. It is essential to use Theorem 1.1, since even after we reduce the Schrödinger map flow to a system of NLS equations and after proving that a unique global solution for (24) exists it is not completely obvious how to go from such a solution for $\nabla_x u$ to an actual map u into N.

Remark 2.5. The proof of Theorem 2.2 can also be used to obtain uniqueness of solutions for initial data in $W^{1,2}(\mathbb{R}, N)$ intersected with the appropriate Strichartz space. In particular this shows that the uniqueness result of Theorem 1.1 is not optimal.

3. Maps from the circle into a Riemann surface

In this section we consider the Schrödinger map flow with the domain being the round circle. Compared with the previous section, the discussion here is more delicate due to the fact that the domain is no longer simply-connected (introduces holonomy) nor noncompact (lack of dispersion).

Let $u : I \times S^1 \to N$ where $I \subset \mathbb{R}$ is a neighborhood of 0. The bundle $u^\star TN \to I \times S^1$ is no longer trivial and so fixing a frame satisfying (13) does not yield a trivialization. To describe the solution of (13) we work instead with $\mathbb{Z}$-invariant objects over $\mathbb{R}$. We therefore make the identifications

$$\mathrm{Maps}(S^1, N) \cong \mathrm{Maps}(\mathbb{R}, N)^{\mathbb{Z}}, \qquad \Gamma(I \times S^1, u^\star TN) \cong \Gamma(I \times \mathbb{R}, u^\star TN)^{\mathbb{Z}}, \tag{46}$$

the superscript denoting $\mathbb{Z}$-invariant objects, and take the freedom to use either one of these identifications interchangeably. Similar identifications will be made for all the other tensor bundles encountered over $I \times S^1$ (for example, $u^\star(T^\star N \otimes T^\star N \otimes T^\star N \otimes TN)$).

Recall that parallel transport is defined as a map $P : u \mapsto \mathrm{Aut}(u(0)^\star TN, u(1)^\star TN)$ for all $u \in C^\infty([0, 1], N)$, which on any Kähler manifold restricts to an operator $P : C^\infty((S^1, \mathrm{pt}), N) \to \imath(U(n))$ on base-pointed loops. Formally, a solution of (13) is given by $e(t, x) = P(u(t)|_{[0,x]})e(t, 0)$. This can be described somewhat more explicitly as follows.

Let U denote a contractible open set in N and let $e_1, \dots, e_n, e_{n+1} = Je_1, \dots, e_{2n} = Je_n$ denote a local orthonormal frame. Assume $u : I \times \mathbb{R} \to N$ is a solution of (4), a collection of loops in N which we will initially assume to be contained in U (and so, in effect, these loops are all contractible in N). Along the image of our flow we denote by $\alpha^1, \dots, \alpha^{2n}$ the dual 1-forms to $e_1, \dots, e_{2n}$. The Levi-Civita connection along our flow restricted to this patch is represented by a section $A_U = \Gamma^k_{ij}\alpha^i$ of $T^\star N|_U \otimes u(n)$ which pulls back to a connection form $u^\star A_U = \Gamma^k_{ij} a^i dx =: B_U dx$ for the pulled-back bundle. A section $e = E^j e_j$ of the pulled-back bundle (as in (46)) is then (locally) parallel when

$$0 = \nabla e = \frac{\partial E^j}{\partial x} e_j \otimes dx + B_U \cdot e \otimes dx = (E^j_{,x} + B_U{}^j_k E^k) e_j \otimes dx. \tag{47}$$

The solution of this first-order matrix equation simplifies considerably in the case $n = 1$. The matrices B_U then lie in the trivial Lie algebra $so(2) \cong u(1)$ and so their exponentials commute. One may therefore integrate (47) to obtain

$$e(t, 1) = \exp\left(-\int_0^1 B_U dx\right) e(t, 0). \tag{48}$$

If D_u is the disc bounded by u and contained in U, and K denotes the Gaussian curvature of N, then Stokes' Theorem gives

$$e(t,1)=\exp\Big(-\int_{D_u} dA_U\Big)e(t,0)=\exp\Big(-\int_{D_u} K\,dV_{N,h}\Big)e(t,0) \tag{49}$$

(possibly up to a factor of 2π, depending on conventions) from which it becomes evident that one may relax the assumption above (for the moment still restricting to contractible loops) and work globally (one might have two choices for D_u then). Also, we see that the holonomy factor is independent of the starting point on the loop. In fact this last fact is seen to be true for noncontractible loops as well. We have therefore a well-defined holonomy map

$$P : C^\infty((S^1,\mathrm{pt}), N) \to \mathrm{SO}(2) = \imath(U(1)).$$

Next, for general u, since $u(0,S^1)$ and $u(t,S^1)$ are homotopic for any $t\in I$ we may define the surface $D_u = u([0,t]\times S^1)$ and as chains on N $\partial D_u = u(t,S^1) - u(0,S^1)$. Let K denote the Gaussian curvature of (N,h). Then we have once again by Stokes' Theorem

$$e(t,1) = P(u)e(t,0) = \exp\Big(-\int_{D_u} K\,dV_{N,h}\Big)P(u_0)\,e(t,0),$$

or for any $x\in\mathbb{R}$ and $l\in\mathbb{N}$

$$e(t,x+l) = P(u)^l e(t,x) = \exp\Big(-l\int_{D_u} K\,dV_{N,h}\Big)P(u_0)^l e(t,0). \tag{50}$$

Therefore a solution of (13) produces a parallel section of $\Gamma(\mathbb{R},u^\star TN)$ rather than of $\Gamma(\mathbb{R},u^\star TN)^{\mathbb{Z}}$. In expressing our $\mathbb{Z}$-invariant tensors in terms of the frame $\{e_j\}_{j=1}^{2n}$ we therefore use coefficients satisfying a relation appropriately proportional to (50). For example if $v\in\Gamma(\mathbb{R},u^\star TN)^{\mathbb{Z}}$ then we may write $v=v^j e_j$ with $v^j(x+l) = P(u)^{-l}v^j(x)$ (while on the other hand sections of endomorphism tensor bundles require no adjustment when $n=1$).

The main difficulty though is that the lifted frame coefficients that live on $\mathbb{R}$ have infinite energy ($L^2(\mathbb{R})$ norm), and so our goal is to still extract an equation for objects that live on S^1, eventually.

Going through the computations of Section 2 it follows that (15) still holds. We then obtain

$$\begin{aligned}\nabla_t e_{\bar{\jmath}}(t,x) &= \sum_{q=1}^{2n} h(\nabla_t e_{\bar{\jmath}}, e_q)(t,x_0)e_q(t,x)\\ &+\Bigg(\sum_{\alpha,\beta}\Big(\big[a^\alpha a^{\bar\beta}R_{\alpha\beta\bar{\jmath}}{}^q + \tfrac12[a^{\bar\alpha}a^{\bar\beta}+a^\alpha a^\beta]R_{\bar\alpha\beta\bar{\jmath}}{}^q\big](t,x) - \big[a^\alpha a^{\bar\beta}R_{\alpha\beta\bar{\jmath}}{}^q + \tfrac12(a^{\bar\alpha}a^{\bar\beta}+a^\alpha a^\beta)R_{\bar\alpha\beta\bar{\jmath}}{}^q\big](t,x_0)\Big)\\ &\qquad - \int_{[x_0,x]}\sum_{\alpha,\beta}\big[a^\alpha a^{\bar\beta}R_{\alpha\beta\bar{\jmath}}{}^q{}_{,x} + \tfrac12(a^{\bar\alpha}a^{\bar\beta}+a^\alpha a^\beta)R_{\bar\alpha\beta\bar{\jmath}}{}^q{}_{,x}\big](t,y)\,dy\Bigg)e_q(t,x)\end{aligned} \tag{51}$$

The terms depending on the fixed point x_0 are in a sense worse than those that depend on the variable point x since the former must be evaluated in the $L^\infty(\mathbb{R}_x)$ norm. To overcome this apparent obstacle we average over S^1 (namely, x_0 in the range $(x-1,x)$) to obtain

$$\nabla_t e_{\bar{j}}(t,x) = \sum_{q=1}^{2n}\Big(\int_{S^1} h(\nabla_t e_{\bar{j}}, e_q)(t,x_0)\,dx_0\Big)\, e_q(t,x)$$
$$+\Big[\sum_{\alpha,\beta}\Big(\big[a^\alpha a^{\bar\beta} R_{\alpha\beta\bar{j}}{}^q + \tfrac{1}{2}(a^{\bar\alpha}a^{\bar\beta}+a^\alpha a^\beta)R_{\bar\alpha\beta\bar{j}}{}^q\big](t,x) - \int_{S^1}\big[a^\alpha a^{\bar\beta} R_{\alpha\beta\bar{j}}{}^q + \tfrac{1}{2}(a^{\bar\alpha}a^{\bar\beta}+a^\alpha a^\beta)R_{\bar\alpha\beta\bar{j}}{}^q\big](t,x_0)\,dx_0\Big)$$
$$-\int_{[x_0,x]}\sum_{\alpha,\beta}\big[a^\alpha a^{\bar\beta} a^s R_{\alpha\beta\bar{j}}{}^q{}_{,s} + \tfrac{1}{2}a^s(a^{\bar\alpha}a^{\bar\beta}+a^\alpha a^\beta)R_{\bar\alpha\beta\bar{j}}{}^q{}_{,s}\big](t,y)\,dy\Big] e_q(t,x), \quad (52)$$

which, upon setting

$$O_{\bar{j}}^q(t,x) := \int_{S^1} h(\nabla_t e_{\bar{j}}, e_q)(t,x_0)\,dx_0,$$
$$P_{\bar{j}}^q(t,x) := \sum_{\alpha,\beta}\big[a^\alpha a^{\bar\beta} R_{\alpha\beta\bar{j}}{}^q + \tfrac{1}{2}(a^{\bar\alpha}a^{\bar\beta}+a^\alpha a^\beta)R_{\bar\alpha\beta\bar{j}}{}^q\big](t,x),$$
$$Q_{\bar{j}}^q(t,x) := -\int_{[x_0,x]}\sum_{\alpha,\beta}\big[a^\alpha a^{\bar\beta} a^s R_{\alpha\beta\bar{j}}{}^q{}_{,s} + \tfrac{1}{2}a^s(a^{\bar\alpha}a^{\bar\beta}+a^\alpha a^\beta)R_{\bar\alpha\beta\bar{j}}{}^q{}_{,s}\big](t,y)\,dy,$$

becomes

$$\nabla_t e_{\bar{j}}(t,x) = \Big(O_{\bar{j}}^q + P_{\bar{j}}^q - \int_{S^1} P_{\bar{j}}^q(t,x_0)\,dx_0 + Q_{\bar{j}}^q\Big) e_q. \quad (53)$$

Switching to complex notation, as in (24), we have

$$\sqrt{-1}\Phi_{,t} = \Phi_{,xx} - \mathrm{U}\cdot\Phi - \mathrm{S}\cdot\Phi + \mathrm{W}\cdot\Phi - \mathrm{T}\cdot\Phi, \quad (54)$$
$$\Phi^\alpha(t,x+l) = P(u(t)|_{[0,1]})^{-l}\Phi^\alpha(t,x), \quad (55)$$

where we have set

$$\mathrm{U} := \iota^{-1}(O_{\bar{j}}^q), \quad \mathrm{S} := \iota^{-1}(P_{\bar{j}}^q), \quad \mathrm{T} := \iota^{-1}(Q_{\bar{j}}^q), \quad \mathrm{W} := \iota^{-1}\Big(\int_{S^1} P_{\bar{j}}^q(t,x_0)\,dx_0\Big).$$

To estimate U we note that according to (19) we have $\nabla_t e_{\bar{j}} = a^{\bar{k}}_{,x}\Gamma^p_{k\bar{j}} e_p$, hence

$$h(\nabla_t e_{\bar{j}}, e_q) = a^{\bar{k}}_{,x}\Gamma^q_{k\bar{j}}. \quad (56)$$

Note that in (56) the left-hand side, hence also the right-hand side, are bona fide functions on S^1 (even though each term separately in the product on the right-hand side is not). Therefore, using (8),

$$\int_{S^1} h(\nabla_t e_{\bar{j}}, e_q)(t,x_0)\,dx_0 = \int_{S^1} a^{\bar{k}}_{,x}\Gamma^q_{k\bar{j}}dx_0 = -\int_{S^1} a^{\bar{k}}\Gamma^q_{k\bar{j},x}dx_0 = -\int_{S^1} a^{\bar{k}}a^p\Gamma^q_{k\bar{j},p}dx_0.$$

Hence we have the estimates

$$\|\mathrm{U}\| < C\int_{S^1}\|\Phi\|^2 dx, \quad \|\mathrm{W}\| < C\int_{S^1}\|\Phi\|^2 dx, \quad \|\mathrm{S}\| < C\|\Phi\|^2, \quad \|\mathrm{T}\| < C\int_{S^1}\|\Phi\|^3 dx, \quad (57)$$

where $C > 0$ depends only on the geometry of (N,h). We stress that (54) and (55) are equations on $I\times\mathbb{R}$.

Next, we try to derive equations that will be defined on $I \times S^1$. Define a real-valued function θ by

$$P(u(t)|_{[0,1]}) =: e^{\sqrt{-1}\,\theta(t)} \in U(1), \tag{58}$$

and set

$$a := \Phi^1 = a^1 + \sqrt{-1}\, a^{\bar{1}}.$$

Note that the holonomy factor (58) is independent of x as noted after (49). Also note that we cannot restrict θ to $[-\pi, \pi)$ in order not to violate continuity of θ.

As remarked in the paragraph after (50), the functions $\mathrm{U} = U_1^1$, $\mathrm{S} = S_1^1$, $\mathrm{W} = W_1^1$, $\mathrm{T} = T_1^1$ are $\mathbb{Z}$-invariant. Also

$$\varphi(t, x) := e^{\sqrt{-1}\,\theta x} a(t, x) \tag{59}$$

is $\mathbb{Z}$-invariant. Moreover, so are all of its x-derivatives. To wit,

$$\varphi(t, x)_{,x} = \sqrt{-1}\,\theta\varphi(t, x) + e^{\sqrt{-1}\,\theta x} a(t, x)_{,x} = \varphi(t, x + 1)_{,x} \tag{60}$$

since $(e^{\sqrt{-1}\,\theta} a(t, x + 1))_{,x} = a(t, x)_{,x}$, and the claim now follows by induction. It follows that the estimates we will obtain for φ will imply the same estimates for a.

Equation (54) becomes, after the change of variable (59),

$$\sqrt{-1}\,\varphi_{,t} = \varphi_{,xx} - 2\sqrt{-1}\,\theta\varphi_{,x} - \big(\theta^2 + x\theta_{,t} + Q_1^1 + S_1^1 - W_1^1 + T_1^1\big)\varphi. \tag{61}$$

Let $\beta : I \times \mathbb{R} \to I \times \mathbb{R}$ be given by

$$\beta(t, x) = \left(t,\ x - 2\int_{[0,t]} \theta\, ds\right).$$

Let

$$\tilde{x} := x + 2\int_{[0,t]} \theta\, ds.$$

Writing $(t, x) = \beta\big(t, x + 2\int_{[0,t]} \theta\, ds\big) = \beta(t, \tilde{x})$, (54) becomes

$$\begin{aligned}\sqrt{-1}(\varphi \circ \beta)_{,t}(t, \tilde{x}) &= (\varphi \circ \beta)_{,\tilde{x}\tilde{x}}(t, \tilde{x}) \\ &\quad - \big(\theta^2(t) + (\tilde{x} - 2\textstyle\int_{[0,t]}\theta\, ds)\theta_{,t}(t) + (Q_1^1 \circ \beta + S_1^1 \circ \beta - W_1^1 \circ \beta + T_1^1 \circ \beta)(t, \tilde{x})\big)(\varphi \circ \beta)(t, \tilde{x}).\end{aligned} \tag{62}$$

This equation is on $I \times S^1$.

Remark 3.1. Note that here it was crucial that θ does not depend on x in order to have $\partial\tilde{x}/\partial x = 1$. This is also the difference from the situation in (29).

The main result of this section is the following a priori estimate:

Theorem 3.2. *Let $(M, g) = (S^1, dx \otimes dx)$ and let (N, J, h) be a complete Riemann surface with bounded geometry. Given $u_0 \in W^{2,2}(S^1, N)$, the solution $\varphi(t, x)$ of the system of NLS equations (62) satisfies for all $T > 0$ the a priori estimate*

$$\|\varphi\|_{L^4([0,T], L^4(S^1, \mathbb{R}^{2n}))} < \infty.$$

This will be shown to imply:

Corollary 3.3. *Let $(M, g) = (S^1, dx \otimes dx)$ and let (N, J, h) be a complete Riemann surface with bounded geometry. Then for integers $k \geq 2$ the flow equation (4) with $u_0 \in W^{k,2}(S^1, N)$ admits a unique solution $u \in C^0(\mathbb{R}, W^{k,2}(S^1, N))$.*

Proof of Theorem 3.2. We will use (62) to obtain a priori estimates on

$$\tilde{\varphi}(t,\tilde{x}) := \varphi \circ \beta(t,\tilde{x}).$$

The estimates on $\tilde{\varphi}$ and on φ are equivalent since the two functions only differ by a time-dependent translation in the space direction. We will localize in time: Indeed it is enough to prove local (in time) a priori estimates for solutions of (62) in $C^0(\mathbb{R}_{t,loc}, L^2(S^1)) \cap L^4(\mathbb{R}_{t,loc} \times S^1)$ depending in a good manner only on $\|\tilde{\varphi}\|_{L^2(S^1)} = E(u_0)^{1/2}$ and a bounded constant depending on time, since that will rule out finite-time blow-up.

We now recall some work of Bourgain that will be of central importance later (see also [Ginibre 1996] for an exposition). We start with some Fourier restriction estimates:

Lemma 3.4 [Bourgain 1993, page 112]. *Let φ be a periodic solution of the linear Schrödinger equation on S^1. Then*

$$\|\varphi\|_{L^4(S^1\times S^1)} \le \sqrt{2}\|\varphi(0)\|_{L^2(S^1)},$$

and dually

$$\|\varphi\|_{L^2(S^1\times S^1)} \le \sqrt{2}\|\varphi\|_{L^{4/3}(S^1\times S^1)}.$$

More generally, Bourgain proved the following fundamental result that allows for the same estimate — now with appropriate weights — even for an arbitrary function whose Fourier modes are not necessarily restricted to the parabola $\{(p, p^2) : p \in \mathbb{Z}\}$. We state the result although we will only directly use a consequence of it.

Lemma 3.5 [Bourgain 1993, Proposition 2.33]. *Let $f(x,t) = \sum_{m,n\in\mathbb{Z}} a_{m,n} e^{\sqrt{-1}(mx+nt)}$ be a function on $S^1 \times S^1$. Then*

$$\left(\sum_{m,n\in\mathbb{Z}} \big(|n-m^2|+1\big)^{-3/4} |a_{m,n}|^2\right)^{1/2} \le C\|f\|_{L^{4/3}(S^1\times S^1)}.$$

In addition, if $|\lambda_{m,n}| \le (1+|n-m^2|)^{-3/4}$, then

$$\left\|\sum_{m,n\in\mathbb{Z}} \lambda_{m,n} a_{m,n} e^{\sqrt{-1}(mx+nt)}\right\|_{L^4(S^1\times S^1)} \le C\|f\|_{L^{4/3}(S^1\times S^1)}.$$

In both estimates $C > 0$ is some universal constant.

Using this estimate, Bourgain obtains the following L^4 estimate for the nonlinear contribution in Duhamel's formula. This estimate will play a central role below. Let $f(x) = \sum_{m\in\mathbb{Z}} a_m e^{\sqrt{-1}\,mx} \in L^2(S^1)$. On $M = S^1$ the Schrödinger operator (see (32)) takes the form

$$(S(t)f)(x) = \sum_{m\in\mathbb{Z}} a_m e^{\sqrt{-1}(mx+m^2 t)}. \tag{63}$$

Lemma 3.6 [Bourgain 1993, §4]. *Let $F \in L^{4/3}(S^1 \times S^1)$. For any $0 < \delta < 1/8$ and $0 < B < \frac{1}{100\delta}$ there holds*

$$\left\|\int_0^{2\delta} S(t-\tau)F(\tau,x)\,d\tau\right\|_{L^4(S^1\times S^1)} \le C(B^{-1/4}+\delta B)\|F\|_{L^{4/3}(S^1\times S^1)},$$

where $C > 0$ is some universal constant.

The constant B can be thought of as a Fourier mode cut-off parameter, measuring distance of a lattice point in $\mathbb{Z}^2$ from the parabola $\{(m, m^2) : m \in \mathbb{Z}\}$. The constant δ is the time cut-off parameter.

Equation (62) is equivalent to the integral equation

$$\tilde{\varphi}(t, \tilde{x}) = S(t)\tilde{\varphi}(0, \tilde{x}) - \sqrt{-1} \int_0^t S(t-\tau) F(\tau, \tilde{x})\, d\tau, \tag{64}$$

with

$$F(\tau, \tilde{x}) = -\big(\theta^2(\tau) + (\tilde{x} - 2\textstyle\int_{[0,\tau]} \theta ds)\theta_{,t}(\tau) + (Q_1^1 \circ \beta + S_1^1 \circ \beta - W_1^1 \circ \beta + T_1^1 \circ \beta)(\tau, \tilde{x})\big)\tilde{\varphi}(\tau, \tilde{x}). \tag{65}$$

There is a subtlety here: The time derivative of $\tilde{\varphi}$ (or of φ) is not necessarily $\mathbb{Z}$-invariant (in $\tilde{x}$). However, (62) holds on $I \times S^1$ and it is equivalent to the integral equation (64).

We would like to obtain an a priori L^4 estimate on $\tilde{\varphi}$. We localize in time, namely multiply (64) by a smooth cut-off function in time $\psi(t)$ satisfying $\psi = 1$ on $[-\delta, \delta]$ and $\psi = 0$ for $|t| \geq 2\delta$. Here δ is a positive number smaller than $1/8$ to be specified later. We may thus regard $\psi\tilde{\varphi}$ as a function on $S^1 \times S^1$ with period 1 in both the t and $\tilde{x}$ variables and Bourgain's estimates apply.

First, the linear term satisfies

$$\|\psi S(t)\tilde{\varphi}(0, \tilde{x})\|_{L^4(S^1 \times S^1)} \leq \sqrt{2}\|\tilde{\varphi}(0, \cdot)\|_{L^2(S^1)} = \sqrt{2E(u_0)},$$

according to Lemma 3.4.

Next, we estimate the integral term. The terms involving Q and W are simpler since |Q| and |W| are uniformly bounded according to (57) and conservation of energy.

We now turn to the other terms. First, using Lemma 3.4 under the integral sign, and assuming $\|\theta\|_{L^\infty} \leq C$, we have

$$\left\| \psi \int_0^t S(t-\tau)\big(\theta^2(\tau)\tilde{\varphi}(\tau, \tilde{x})\big)\, d\tau \right\|_{L^4(S^1 \times S^1)} \leq \int_0^{2\delta} \|\theta^2(\tau)\tilde{\varphi}(\tau, \cdot)\|_{L^2(S^1)} d\tau \leq 2C^2\delta\|\varphi(0)\|_{L^2(S^1)}. \tag{66}$$

To show that this assumption holds, use the representation of the holonomy given by (48): $|\theta(t)| \leq \int_0^1 |\Gamma_{ij}^k||a^i| dx \leq C' E(u_0)^{1/2}$, where we have used the assumption of bounded geometry — indeed it implies that Christoffel symbols are uniformly bounded [Eichhorn 1991].

Second, $|\tilde{x}| \leq 1$ and so $|\tilde{x} - 2\int_{[0,\tau]} \theta\, ds| \leq 1 + 2 \cdot 1 \cdot C$. Let $\{\alpha_1, \alpha_{\bar{1}}\}$ be an orthonormal coframe dual to $\{e_1, e_{\bar{1}}\}$. To compute the time derivative of θ, recall that by (50) we have

$$\theta(t) = \int_{D_u} K\, dV_{N,h} = \int_{D_u} K \alpha_1 \wedge \alpha_{\bar{1}} = \int_{I \times S^1} K \circ u(t, x)[a^1 b^{\bar{1}} - a^{\bar{1}} b^1]\, dx \wedge dt,$$

since $u^\star\alpha_1 = a^1 dx + b^1 dt$, $u^\star\alpha_{\bar{1}} = a^{\bar{1}} dx + b^{\bar{1}} dt$. Combining this with the equality $b^k = a^{\bar{k}}_{,x}$ given by (9), we have

$$\theta_{,t} = \int_{S^1} K \circ u(t, x)(a^1 b^{\bar{1}} - a^{\bar{1}} b^1)\, dx = -\tfrac{1}{2} \int_{S^1} K \circ u(t, x)((a^1)^2 + (a^{\bar{1}})^2)_{,x} dx.$$

Integrating by parts this becomes

$$\theta_{,t} = \tfrac{1}{2} \int_{S^1} (K \circ u(t, x))_{,x}((a^1)^2 + (a^{\bar{1}})^2)\, dx = \tfrac{1}{2} \int_{S^1} K_{,s} \circ u(t, x) a^s ((a^1)^2 + (a^{\bar{1}})^2)\, dx.$$

By bounded geometry we therefore have

$$\|\theta_{,t}\|_{L^\infty} \le C\|a\|^3_{L^3(S^1)}. \tag{67}$$

Therefore the term $\big(\tilde{x} - 2\int_{[0,\tau]}\theta\, ds\big)\theta_{,t}(\tau)\tilde{\varphi}(\tau,\tilde{x})$ behaves in the same way as the term $T_1^1 \circ \beta\tilde{\varphi}(\tau,\tilde{x})$ in (65), and so it's enough to treat the latter. We will do that shortly.

Third, $|S_1^1 \circ \beta \cdot \tilde{\varphi}| < C|\tilde{\varphi}|^3$, and therefore this term may be estimated in $L^4(S^1 \times S^1)$ just like in Bourgain's estimates for a cubic nonlinearity. More precisely, by Lemma 3.6 we have

$$\left\| \psi \int_0^t S(t-\tau)\big(|\tilde{\varphi}|^3(\tau,\tilde{x})\big)\, d\tau \right\|_{L^4(S^1\times S^1)} \le C(\delta B + B^{-1/4})\|\psi\tilde{\varphi}\|^3_{L^4(S^1\times S^1)}, \tag{68}$$

with $B > 0$ as in the lemma, δ is as before the time cut-off parameter, and $C > 0$ is a uniform constant.

Fourth, using Lemma 3.4 and energy conservation we have

$$\begin{aligned}\left\| \psi \int_0^t S(t-\tau)\Big(\tilde{\varphi}\int_{S^1}|\tilde{\varphi}|^3 d\tilde{x}(\tau,\tilde{x})\Big)\, d\tau \right\|_{L^4(S^1\times S^1)} &\le C\int_0^{2\delta} \left\| \tilde{\varphi}(\tau,\,\cdot\,)\int_{S^1}|\tilde{\varphi}(\tau,\,\cdot\,)|^3 d\tilde{x}\right\|_{L^2(S^1)} d\tau \\ &\le C'\int_0^{2\delta}\int_{S^1}|\tilde{\varphi}(\tau,\,\cdot\,)|^3 d\tilde{x}\, d\tau \\ &\le C''\delta^{1/4}\|\varphi\|^3_{L^4(S^1\times S^1)}. \end{aligned} \tag{69}$$

Combining Equations (66)–(69) we have

$$\|\psi\tilde{\varphi}\|_{L^4([0,2\delta]\times S^1)} \le C\Big((1+\delta)\|\varphi\|_{L^2(S^1)} + (\delta B + B^{-1/4} + \delta^{1/4})\|\psi\tilde{\varphi}\|^3_{L^4([0,2\delta]\times S^1)}\Big). \tag{70}$$

In fact, due to energy conservation, the time interval may be taken to be $[t_1, t_1+2\delta]$ for any $t_1 \in \mathbb{R}$. Now, by choosing B large enough and then choosing δ small enough, in such a manner that δB is also small enough (all of these choices depend only on the initial energy and the geometry) we therefore may argue similarly to Bourgain to obtain a uniform estimate on $\|\tilde{\varphi}\|_{L^4([t_1,t_1+2\delta]\times S^1)} = \|\varphi\|_{L^4([t_1,t_1+2\delta]\times S^1)} = \|a\|_{L^4([t_1,t_1+2\delta]\times S^1)}$:

$$\|\varphi\|_{L^4([t_1,t_1+2\delta]\times S^1)} < C, \tag{71}$$

where $C > 0$ is a uniform constant depending only on the initial energy and the geometry. We therefore obtain the global a priori estimate

$$\|a\|_{L^4([0,T]\times S^1)} < \infty,$$

for all $T > 0$. This concludes the proof of Theorem 3.2. □

Remark 3.7. A difference between our situation and that of Bourgain [1993, page 139] is that while Bourgain actually proves the existence (and uniqueness) of a local solution of the periodic cubic NLS in $L^4(\mathbb{R}_{t,loc} \times S^1)$ using energy conservation, we only need to prove an a priori estimate in this norm for the unique local solution given by Theorem 1.1.

Conclusion of the proof of Corollary 3.3. To obtain global well-posedness for our original flow equation in $W^{k,2}$ we take $k-1$ derivatives of (62). In fact, we obtain certain terms that are worse than those that arise when one differentiates the cubic NLS. For example, for $k = 2$ we obtain several extra terms the worst of which are of order $|\tilde{\varphi}|^4$ and $|\tilde{\varphi}_{,x}| \cdot \|\tilde{\varphi}\|^3_{L^3(S^1)}$. Such terms may be handled nevertheless. We carry out the computations in detail in the case $k=2$, omitting the details in the case $k \geq 3$ as they are similar (see the remarks in Step 3 of Section 2).

Set $w := \tilde{\varphi}_{,\tilde{x}}$. Taking a derivative of (62) we obtain

$$\sqrt{-1}\, w_{,t} = w_{,\tilde{x}\tilde{x}} + G, \tag{72}$$

where

$$|G| < C\big(|\tilde{\varphi}|^4 + |\tilde{\varphi}|^2 + |w|\big(1 + |\tilde{\varphi}|^2 + \|\tilde{\varphi}\|^3_{L^3(S^1)}\big)\big). \tag{73}$$

As before we would like to obtain an $L^4(\mathbb{R}_{t,loc} \times S^1)$ estimate, this time for w. We use Lemma 3.4 in order to handle the term $|\tilde{\varphi}|^4$ (more precisely, the corresponding term in the Duhamel formula); the corresponding contribution is bounded by

$$C \int_{t_1}^{t_1+2\delta} \|\varphi(\tau,\,\cdot\,)^4\|_{L^2(S^1)} d\tau.$$

Using the Gagliardo–Nirenberg inequality [Aubin 1998, page 93], we have

$$\|\varphi\|^4_{L^8(S^1)} \leq C\big(1 + \|\varphi_{,x}\|^{1/2}_{L^4(S^1)}\big)\big(1 + \|\varphi\|^{7/2}_{L^4(S^1)}\big), \tag{74}$$

Note that the Gagliardo–Nirenberg inequality as cited requires $\int_{S^1} \varphi\, dx = 0$; nevertheless we know that $\|\varphi\|_{L^1(S^1)}$ is uniformly bounded in time due to the Cauchy–Schwarz inequality and conservation of energy, and so (74) holds in our case. It is enough to treat the worst term on the right-hand side of (74), namely the term $\|\varphi_{,x}\|^{1/2}_{L^4(S^1)} \|\varphi\|^{7/2}_{L^4(S^1)}$. The Hölder inequality and (71) give

$$\int_{t_1}^{t_1+2\delta} \|\varphi_{,x}\|^{1/2}_{L^4(S^1)} \|\varphi\|^{7/2}_{L^4(S^1)} ds \leq \|\varphi_{,x}\|^{1/2}_{L^4([t_1,t_1+2\delta]\times S^1)} \|\varphi\|^{7/2}_{L^4([t_1,t_1+2\delta]\times S^1)} \leq C\|\varphi_{,x}\|^{1/2}_{L^4([t_1,t_1+2\delta]\times S^1)}. \tag{75}$$

Since this is sublinear in the norm we are estimating it will be possible to use this inequality to obtain the a priori estimate we are after. Next, of course the nonlinear term $|\tilde{\varphi}|^2$ in (73) is even easier to handle:

$$\|\varphi^2\|_{L^1([t_1,t_1+2\delta],L^2(S^1))}. \leq C\sqrt{\delta}\|\varphi\|_{L^4([t_1,t_1+2\delta],L^4(S^1))} \leq C'\sqrt{\delta}. \tag{76}$$

Let us now treat the other nonlinearities. To handle the contribution of the term $|\tilde{\varphi}^2\tilde{\varphi}_{,x}|$ to the Duhamel formula, we apply Lemma 3.6 and the Hölder inequality

$$\begin{aligned}\Big\|\int_{t_1}^{t} S(t-\tau)(\varphi^2\varphi_{,x})(\tau,\,\cdot\,)\,d\tau\Big\|_{L^4([t_1,t_1+2\delta]\times S^1)} &\leq C(\delta B + B^{-1/4})\|\varphi^2\varphi_{,x}\|_{L^{4/3}([t_1,t_1+2\delta]\times S^1)}\\ &\leq C(\delta B + B^{-1/4})\|\varphi_{,x}\|_{L^4([t_1,t_1+2\delta]\times S^1)}\|\varphi\|^2_{L^4([t_1,t_1+2\delta]\times S^1)};\end{aligned} \tag{77}$$

this term can be controlled by a small uniform constant times $\|\varphi_{,x}\|_{L^4([t_1,t_1+2\delta]\times S^1)}$, by choosing δ, B appropriately.

Next, using Lemma 3.4, the contribution of the term $|\tilde{\varphi}_{,x}| \cdot \|\tilde{\varphi}\|^3_{L^3(S^1)}$ to the Duhamel formula can be bounded as follows:

$$\Big\|\int_{t_1}^{t} S(t-\tau)(|\tilde{\varphi}_{,x}| \cdot \|\tilde{\varphi}\|^3_{L^3(S^1)})(\tau,\,\cdot\,)\,d\tau\Big\|_{L^4([t_1,t_1+2\delta]\times S^1)} \leq C\int_{t_1}^{t_1+2\delta} \big\||\varphi_{,x}| \cdot |\varphi\|^3_{L^3(S^1)}\big\|_{L^2(S^1)} dt, \tag{78}$$

and using the interpolation inequality $\|\varphi\|^3_{L^3(S^1)} \le \|\varphi\|^2_{L^4(S^1)}\|\varphi\|_{L^2(S^1)} \le C\|\varphi\|^2_{L^4(S^1)}$, this may be estimated as follows:

$$\begin{aligned}
(78) &\le C' \int_{t_1}^{t_1+2\delta} \|\varphi_{,x}\|_{L^2(S^1)}\|\varphi\|^2_{L^4(S^1)} dt \le C'' \int_{t_1}^{t_1+2\delta} \|\varphi_{,x}\|_{L^4(S^1)}\|\varphi\|^2_{L^4(S^1)} dt \\
&\le C'' ||\varphi_{,x}||_{L^4([t_1,t_1+2\delta]\times S^1)} \Big(\int_{t_1}^{t_1+2\delta} \|\varphi\|^{8/3}_{L^4(S^1)} dt \Big)^{3/4} \\
&\le C''(2\delta)^{1/4}\|\varphi_{,x}\|_{L^4([t_1,t_1+2\delta]\times S^1)}\|\varphi\|^2_{L^4([t_1,t_1+2\delta]\times S^1)} \le C'''\delta^{1/4}\|\varphi_{,x}\|_{L^4([t_1,t_1+2\delta]\times S^1)}.
\end{aligned} \tag{79}$$

To summarize, combining (75), (76), (77), and (79), we may thus find $\delta, C > 0$, depending only on the initial energy and the geometry, for which (71) still holds and for which we also have the a priori estimate

$$\|\varphi_{,x}\|_{L^4([t_1,t_1+2\delta]\times S^1)} \le C(1 + \|\varphi_{,x}(t_1)\|_{L^2(S^1)}). \tag{80}$$

The main point here is that δ does not depend on $\|\varphi_{,x}(t_1)\|_{L^2(S^1)}$.

We now need to "close" the argument by estimating the $L^\infty([t_1, t_1+2\delta], L^2(S^1))$ norm of $\varphi_{,x}$, making use of the auxiliary estimate (80). For each $t \in [t_1, t_1 + 2\delta]$ we have (see (72))

$$\begin{aligned}
\|\varphi_{,x}\|_{L^\infty([t_1,t_1+2\delta],L^2(S^1))} &\le C\|\varphi_{,x}(t_1)\|_{L^2(S^1)} + \big\| \|G\|_{L^1([t_1,t],L^2(S^1))} \big\|_{L^\infty([t_1,t_1+2\delta])} \\
&= C\|\varphi_{,x}(t_1)\|_{L^2(S^1)} + \|G\|_{L^1([t_1,t_1+2\delta],L^2(S^1))}.
\end{aligned} \tag{81}$$

With the exception of the term $|\varphi|^2|\varphi_{,x}|$, we have already estimated all of the nonlinearities occurring in G in $L^1([t_1, t_1 + 2\delta], L^2(S^1))$ in the process of proving (80) (see (75), (76), and (79)). Let us now estimate that term in this norm. Using the Gagliardo–Nirenberg inequality (together with the remark following (74)), energy conservation and the Hölder inequality we have

$$\begin{aligned}
\int_{t_1}^{t_1+2\delta} \|\varphi^2\varphi_{,x}\|_{L^2(S^1)} d\tau &\le \int_{t_1}^{t_1+2\delta} \|\varphi\|_{L^\infty(S^1)}\|\varphi\|_{L^4(S^1)}\|\varphi_{,x}\|_{L^4(S^1)} d\tau \\
&\le C \int_{t_1}^{t_1+2\delta} \big(1 + \|\varphi\|^{1/2}_{L^2(S^1)}\big)\big(1 + \|\varphi_{,x}\|^{1/2}_{L^2(S^1)}\big)\|\varphi\|_{L^4(S^1)}\|\varphi_{,x}\|_{L^4(S^1)} d\tau.
\end{aligned} \tag{82}$$

As earlier, it suffices to estimate the worst term on the right-hand side. Namely, we estimate

$$\begin{aligned}
\int_{t_1}^{t_1+2\delta} \|\varphi\|^{1/2}_{L^2(S^1)}\|\varphi_{,x}\|^{1/2}_{L^2(S^1)}\|\varphi\|_{L^4(S^1)}\|\varphi_{,x}\|_{L^4(S^1)} d\tau &\le C' \int_{t_1}^{t_1+2\delta} \|\varphi\|_{L^4(S^1)}\|\varphi_{,x}\|^{3/2}_{L^4(S^1)} d\tau \\
&\le C''\delta^{3/8}\|\varphi\|_{L^4([t_1,t_1+2\delta]\times S^1)}\|\varphi_{,x}\|^{3/2}_{L^4([t_1,t_1+2\delta]\times S^1)}.
\end{aligned} \tag{83}$$

It follows that

$$\|\varphi_{,x}\|_{L^\infty([t_1,t_1+2\delta],L^2(S^1))} \le C\big(1 + \|\varphi_{,x}(t_1)\|_{L^2(S^1)} + \|\varphi_{,x}(t_1)\|^{3/2}_{L^2(S^1)}\big). \tag{84}$$

Therefore there exists uniform constants $C, C', C'' > 0$ such that for all $T > 0$,

$$\|\varphi_{,x}\|_{L^\infty([0,T],L^2(S^1))} < C'' e^{C' e^{CT}} < \infty. \tag{85}$$

This concludes the proof of Corollary 3.3. □

4. Maps of the circle into a Kähler manifold

In this section we explain the difficulties encountered when one tries to apply the same methods to treat the case of maps from the circle to Kähler manifolds of arbitrary dimension $n \geq 1$.

For general n, one gets an expression for a solution of (13) given by the chronological exponential

$$\begin{aligned} e(t,x+1) &= A^{-1}(t,x)e(t,x) \\ &:= \lim_{n\to\infty} \exp\big(-\tfrac{1}{n}B_U(t,x+1)\big)\exp\big(-\tfrac{1}{n}B_U(t,x+\tfrac{n-1}{n})\big)\cdots\exp\big(-\tfrac{1}{n}B_U(t,x+\tfrac{1}{n})\big)e(t,0); \quad (86)\end{aligned}$$

see for example [Dubrovin et al. 1985]. Applying ∇_x to (86) and using the fact that $\nabla_x e = 0$, we obtain $A_{,x} = 0$, that is, $A(t,x)$ does not depend on x. From now on we simply write $A(t)$.

Now $\Phi(t,x+1) = A(t)\Phi(t,x)$. Since $A(t)$ is unitary (and hence normal) it is unitarily diagonalizable and we set

$$A(t) = U(t)^\star D(t) U(t), \qquad \text{with } U(t)\in U(n) \text{ and } D(t) = \operatorname{diag}(e^{\sqrt{-1}\,\theta_1},\dots,e^{\sqrt{-1}\,\theta_n}).$$

The vector-valued function $\tilde\Phi$ defined by

$$\tilde\Phi(t,x) := U(t)^\star D(t)^{-x} U(t)\Phi(t,x) = A(t)^{-x}\Phi(t,x).$$

is periodic in x. Moreover, by a computation similar to (60), so are all of its x-derivatives. We have

$$\begin{aligned} \Phi_{,t} &= (A(t)^x)_{,t}\tilde\Phi + A(t)^x\tilde\Phi_{,t}, \\ \Phi_{,x} &= U(t)^\star D(t)^x \operatorname{diag}(\sqrt{-1}\,\theta_i)U(t)\tilde\Phi + U(t)^\star D(t)^x U(t)\tilde\Phi_{,x}, \\ \Phi_{,xx} &= U(t)^\star D(t)^x \operatorname{diag}(-\theta_i^2)U(t)\tilde\Phi + 2U(t)^\star D(t)^x\operatorname{diag}(\sqrt{-1}\,\theta_i)U(t)\tilde\Phi_{,x} + U(t)^\star D(t)^x U(t)\tilde\Phi_{,xx}. \end{aligned}$$

It follows that

$$\begin{aligned}\sqrt{-1}\Phi_{,t} - \Phi_{,xx} = A(t)^x\big[&\sqrt{-1}\tilde\Phi_{,t} - \tilde\Phi_{,xx} + (A(t)^x)_{,t}\tilde\Phi \\ &- U(t)^\star\operatorname{diag}(-\theta_i^2)U(t)\tilde\Phi - 2U(t)^\star\operatorname{diag}(\sqrt{-1}\,\theta_i)U(t)\tilde\Phi_{,x}\big]. \quad (87)\end{aligned}$$

Therefore equations (54)–(55) may be rewritten as

$$\begin{aligned}\sqrt{-1}\,\tilde\Phi_{,t} = \tilde\Phi_{,xx} &- (A(t)^x)_{,t}\tilde\Phi + U(t)^\star\operatorname{diag}(-\theta_i^2)U(t)\tilde\Phi \\ &+ 2U(t)^\star\operatorname{diag}(\sqrt{-1}\,\theta_i)U(t)\tilde\Phi_{,x} - A(t)^{-x}\big(\mathrm{Q}\cdot\Phi + \mathrm{S}\cdot\Phi - \mathrm{W}\cdot\Phi + \mathrm{T}\cdot\Phi\big). \quad (88)\end{aligned}$$

Note that the last term is expressed in terms of Φ instead of $\tilde\Phi$. However as far as the estimates are concerned this is not important since it involves no derivatives and the two vectors differ by a unitary transformation. Two problems now arise. First, one needs to obtain an estimate on the variation of the holonomy matrix $A(t)$ along the flow. Such an estimate was available in the one-dimensional setting due to the Gauss-Bonnet theorem. Second, the matrix multiplying the first derivative term is not diagonal and so it is not clear how to eliminate this term.

Although this requires some work, and we will not attempt to provide the details here, the first difficulty may likely be overcome using the theory developed by Chacon and Fomenko [1991] for a noncommutative version of the Stokes' Theorem for product integrals (see also the classical references [Nijenhuis

1953; Schlesinger 1928]). To approach the second difficulty one may consider $\hat{\Phi} := U(t)^{\star} D(t)^{-x} \Phi$ instead of $\tilde{\Phi}$. Then the matrix multiplying the first derivative of $\hat{\Phi}$ is diagonal. Therefore, we may apply the space-time transformation as in the Riemann surface case, however for each equation in the system separately. However, this introduces a new obstacle. Indeed, then one needs to control the time derivative of $D(t)$ as well as of $U(t)$. The main difficulty comes from the latter. In general, the unitary diagonalizing matrix does not vary smoothly (or even continuously) even when a family of matrices does [Kato 1966, page 111]. Instead one may try to diagonalize $A(t)$ smoothly. However, to the best of our knowledge, even given such a diagonalization, the problem is that even though the diagonalizing matrix is then smooth one has essentially no control over its derivatives (that is, estimates on these derivatives in terms of derivatives of $A(t)$). We hope to come back to this problem in the future. In some sense the two transformations (to $\tilde{\Phi}$ and to $\hat{\Phi}$) are dual to each other, and one may ask whether for higher-dimensional domains the two troublesome terms, namely the first derivative term and the derivative of the holonomy, may be a source for finite-time blow-up.

Acknowledgments

The authors are grateful to the referee for a careful reading and helpful suggestions that improved the article. They also thank B. Dai for a careful reading and for correcting an error in the computation of the holonomy in an earlier version of this article. This material is based upon work supported in part under National Science Foundation Graduate and Postdoctoral Research Fellowships and grants DMS-0406627, 0702270, 0602678. Rubinstein was also supported by graduate fellowships at MIT and at Princeton University, and by a Clay Mathematics Institute Liftoff Fellowship. This work was mostly carried out while Rodnianski was visiting MIT in Spring 2006, and while Rubinstein was visiting Princeton University in Spring 2008, and the authors thank these institutions for their hospitality.

References

[Aubin 1998] T. Aubin, *Some nonlinear problems in Riemannian geometry*, Springer, Berlin, 1998. MR 99i:58001 Zbl 0896.53003

[Bejenaru et al. 2007] I. Bejenaru, A. D. Ionescu, and C. E. Kenig, "Global existence and uniqueness of Schrödinger maps in dimensions $d \geq 4$", *Adv. Math.* **215**:1 (2007), 263–291. MR 2008i:35124 Zbl 1152.35049

[Bejenaru et al. 2008] I. Bejenaru, A. D. Ionescu, C. E. Kenig, and D. Tataru, "Global Schrödinger maps", preprint, 2008. arXiv 0807.0265

[Bourgain 1993] J. Bourgain, "Fourier transform restriction phenomena for certain lattice subsets and applications to nonlinear evolution equations. I. Schrödinger equations", *Geom. Funct. Anal.* **3**:2 (1993), 107–156. MR 95d:35160a Zbl 0787.35097

[Cazenave 2003] T. Cazenave, *Semilinear Schrödinger equations*, Courant Lecture Notes in Mathematics **10**, Courant Institute, New York, 2003. MR 2004j:35266 Zbl 1055.35003

[Chacon and Fomenko 1991] R. V. Chacon and A. T. Fomenko, "Stokes' formula for Lie algebra valued connection and curvature forms", *Adv. Math.* **88**:2 (1991), 258–300. MR 92m:53116 Zbl 0741.53021

[Chang et al. 2000] N.-H. Chang, J. Shatah, and K. Uhlenbeck, "Schrödinger maps", *Comm. Pure Appl. Math.* **53**:5 (2000), 590–602. MR 2001a:35161 Zbl 1028.35134

[Chihara 2008] H. Chihara, "Schrödinger flow into almost Hermitian manifolds", preprint, 2008. arXiv 0807.3395

[Ding 2002] W.-Y. Ding, "On the Schrödinger flows", pp. 283–291 in *Proceedings of the International Congress of Mathematicians* (Beijing, 2002), vol. 2, Higher Education Press, 2002.

[Ding and Wang 1998] W. Ding and Y. Wang, "Schrödinger flow of maps into symplectic manifolds", *Sci. China Ser. A* **41**:7 (1998), 746–755. MR 99m:58062 Zbl 0918.53017

[Ding and Wang 2001] W. Ding and Y. Wang, "Local Schrödinger flow into Kähler manifolds", *Sci. China Ser. A* **44**:11 (2001), 1446–1464. MR 2002k:53129 Zbl 1019.53032

[Dubrovin et al. 1985] B. A. Dubrovin, S. P. Novikov, and A. T. Fomenko, *Modern geometry—methods and applications*, vol. II, Graduate Texts in Math. **104**, Springer, New York, 1985. MR 86m:53001 Zbl 0565.57001

[Eells and Sampson 1964] J. Eells, Jr. and J. H. Sampson, "Harmonic mappings of Riemannian manifolds", *Amer. J. Math.* **86** (1964), 109–160. MR 29 #1603 Zbl 0122.40102

[Eichhorn 1991] J. Eichhorn, "The boundedness of connection coefficients and their derivatives", *Math. Nachr.* **152** (1991), 145–158. MR 92k:53069 Zbl 0736.53031

[Eidelman and Zhitarashu 1998] S. D. Eidelman and N. V. Zhitarashu, *Parabolic boundary value problems*, Operator Theory: Advances and Applications **101**, Birkhäuser, Basel, 1998. MR 99h:35080 Zbl 0893.35001

[Ginibre 1996] J. Ginibre, "Le problème de Cauchy pour des EDP semi-linéaires périodiques en variables d'espace (d'après Bourgain)", pp. 163–187 (Exp. 796) in *Séminaire Bourbaki, 1994/95*, Astérisque **237**, Soc. Mat. de France, Paris, 1996. MR 98e:35154 Zbl 0870.35096

[Kato 1966] T. Kato, *Perturbation theory for linear operators*, Die Grundlehren der mathematischen Wissenschaften **132**, Springer, New York, 1966. MR 34 #3324

[Landau and Lifschitz 1935] L. D. Landau and E. M. Lifschitz, "On the theory of the dispersion of magnetic permeability in ferromagnetic bodies", *Phys. Z. Sow.* **8** (1935), 153–169. Reprinted as pp. 101–114 in *Collected papers of L. D. Landau*, Gordon and Breach, New York (1967), and as pp. 14–22 in a special issue of *Ukr. J. Physics* **53** (2008). Zbl 0565.57001

[McGahagan 2004] H. McGahagan, *Some existence and uniqueness results for Schroedinger maps and Landau–Lifshitz–Maxwell equations*, Ph.D. thesis, New York University, 2004.

[McGahagan 2007] H. McGahagan, "An approximation scheme for Schrödinger maps", *Comm. Partial Differential Equations* **32**:1-3 (2007), 375–400. MR 2009e:35265 Zbl 1122.35138

[Nijenhuis 1953] A. Nijenhuis, "On the holonomy groups of linear connections. IA. General properties of affine connections", *Nederl. Akad. Wetensch. Proc. Ser. A.* **56** = *Indagationes Math.* **15** (1953), 233–240. MR 16,171c Zbl 0051.13203

[Pang et al. 2001] P. Y. H. Pang, H.-Y. Wang, and Y.-D. Wang, "Schrödinger flow for maps into Kähler manifolds", *Asian J. Math.* **5**:3 (2001), 509–533. MR 2003h:53092 Zbl 1021.53039

[Pang et al. 2002] P. Y. H. Pang, H.-Y. Wang, and Y.-D. Wang, "Schrödinger flow on Hermitian locally symmetric spaces", *Comm. Anal. Geom.* **10**:4 (2002), 653–681. MR 2003i:53102 Zbl 1032.53057

[Schlesinger 1928] L. Schlesinger, "Parallelverschiebung und Krümmungstensor", *Math. Ann.* **99** (1928), 413–434. MR 1512459 JFM 54.0766.04

[Sulem et al. 1986] P.-L. Sulem, C. Sulem, and C. Bardos, "On the continuous limit for a system of classical spins", *Comm. Math. Phys.* **107**:3 (1986), 431–454. MR 87k:82024 Zbl 0614.35087

[Terng and Uhlenbeck 2006] C.-L. Terng and K. Uhlenbeck, "Schrödinger flows on Grassmannians", pp. 235–256 in *Integrable systems, geometry, and topology*, edited by C.-L. Terng, AMS/IP Stud. Adv. Math. **36**, Amer. Math. Soc., Providence, RI, 2006. MR 2007a:37079 Zbl 1110.37056

[Zhou et al. 1991] Y. L. Zhou, B. L. Guo, and S. B. Tan, "Existence and uniqueness of smooth solution for system of ferromagnetic chain", *Sci. China Ser. A* **34**:3 (1991), 257–266. MR 92i:82030 Zbl 0752.35074

Received 13 Nov 2008. Revised 25 Feb 2009. Accepted 4 May 2009.

IGOR RODNIANSKI: irod@math.princeton.edu
Department of Mathematics, Princeton University, Fine Hall, Princeton, NJ 08544, United States

YANIR A. RUBINSTEIN: yanir@member.ams.org
Department of Mathematics, Johns Hopkins University, 3400 N Charles St., Baltimore, MD 21218, United States

GIGLIOLA STAFFILANI: gigliola@math.mit.edu
Department of Mathematics, Massachusetts Institue of Technology, 77 Massachusetts Avenue, 2-246, Cambridge, MA 02139, United States
http://www-math.mit.edu/~gigliola/

ANALYSIS AND PDE
Vol. 2, No. 2, 2009

ROTH'S THEOREM IN $\mathbb{Z}_4^n$

TOM SANDERS

We show that if $A \subset \mathbb{Z}_4^n$ contains no three-term arithmetic progressions in which all the elements are distinct then $|A| = o(4^n/n)$.

1. Introduction

Let G be a finite abelian group. A three-term arithmetic progression in G is a triple $(x, x+d, x+2d)$ with $x, d \in G$; a proper progression is one in which all the elements are different, that is, $2d \neq 0_G$.

Roth [1953] famously proved that any subset of $\mathbb{Z}/N\mathbb{Z}$ of sufficiently large density contains a proper three-term arithmetic progression, a result which was generalised by Meshulam:

Theorem 1.1 [Meshulam 1995]. *Suppose that G is a finite abelian group of odd order and $A \subset G$ contains no proper three-term arithmetic progressions. Then*

$$|A| = O\big(|G|/\log^{\Omega(1)}|G|\big).$$

An explicit value for the $\Omega(1)$ constant can be read out of the proof, and it seems that in light of [Bourgain 2008] (itself improving on [Bourgain 1999; Szemerédi 1990; Heath-Brown 1987]) one could probably take any constant strictly less than $2/3$. While this appears to be the limit in general, for certain groups one can do better. Indeed, for $\mathbb{Z}_3^n$ (or, more generally, any abelian group of odd order and bounded exponent), Roth's original argument simplifies considerably to give the following result, which is qualitatively due to Brown and Buhler [1984].

Theorem 1.2 Roth–Meshulam. *Suppose that $G = \mathbb{Z}_3^n$ and $A \subset G$ contains no proper three-term arithmetic progressions. Then*

$$|A| = O\big(|G|/\log|G|\big).$$

The question of what the true bounds on $|A|$ are arises in many different studies [Frankl et al. 1987; Yekhanin and Dumer 2004; Edel 2004; Edel et al. 2007] and improving the bound is a well known open problem, as reported in [Green 2005; Croot and Lev 2007; Tao 2008, Section 3.1]; the closest anyone has come is in [Croot 2007; 2008]. While we are not able to make progress on this question, it is the purpose of this paper to show an improvement for a different class of groups.

It was quite natural in Theorem 1.1 to insist that G be of odd order: in the group $\mathbb{Z}_2^n$ every arithmetic progression is easily seen to be of the form (x, y, x), so no set contains a proper progression. Not all groups of even order are as trivial as $\mathbb{Z}_2^n$ and, as part of a more general corpus of results, Lev resolved the question of which abelian groups Meshulam's theorem could be extended to.

MSC2000: 42A05.
Keywords: Roth–Meshulam, cap set problem, Fourier, Freĭman, Balog–Szemerédi, characteristic 2, $\mathbb{Z}_4^n$, three-term arithmetic progressions.

Theorem 1.3 [Lev 2004]. *Suppose that $G = \mathbb{Z}_4^n$ and $A \subset G$ contains no proper three-term arithmetic progressions. Then*

$$|A| = O\big(|G|/\log|G|\big).$$

This special case of Lev's work follows rather easily from the method used to prove the Roth–Meshulam theorem coupled with a positivity observation. At considerable further expense we are able to establish a minor improvement:

Theorem 1.4. *Suppose that $G = \mathbb{Z}_4^n$ and $A \subset G$ contains no proper three-term arithmetic progressions. Then*

$$|A| = O\big(|G|/\log|G|\log\log^{\Omega(1)}|G|\big).$$

The requirement that *all* the elements of our progressions be distinct is essential in our work. It is easy to see by the Cauchy–Schwarz inequality that any set $A \subset G := \mathbb{Z}_4^n$ has at least $\alpha^2|G|^{3/2}$ progressions. It follows that if $\alpha^2|G|^{3/2} > |G|$ then A contains a progression in which not all the elements are the same; however, this may well be a degenerate one of the form (x, y, x).

The paper now splits as follows. In Section 2, we record the necessary information about the Fourier transform. In Section 3 and Section 4, we outline our approach to counting progressions and compare it with the Roth–Meshulam–Lev method to give some indication of where we are able to make gains. In Section 5 we define the notion of a family which we shall work with for the bulk of the paper and the proof of Theorem 1.4, which are in Sections 6–11. We close in Section 12 with a conjecture and a discussion of lower bounds.

2. The Fourier transform

We shall make considerable use of the Fourier transform, for which the classic [Rudin 1962] serves as the standard reference. Having said this, the style of our work has more in common with [Tao and Vu 2006], which is also to be recommended.

Suppose that G is a finite abelian group. $\widehat{G}$ denotes the *dual group* of G, that is the group of homomorphisms $\gamma : G \to S^1$, where $S^1 := \{z \in \mathbb{C} : |z| = 1\}$. G is endowed with a natural Haar probability measure, denoted $\mathbb{P}_G$, assigning mass $|G|^{-1}$ to each element of G; we denote integration against $\mathbb{P}_G$ by $\mathbb{E}_{x\in G}$ and, in general, $\mathbb{E}_{x\in S}$ corresponds to integration against the probability measure $\mathbb{P}_S$ assigning mass $|S|^{-1}$ to each $s \in S$.

For $p \in [1, \infty]$ we define the spaces $L^p(G)$ and $\ell^p(G)$ to be the vector space of functions $f : G \to \mathbb{C}$ endowed with the norms

$$\|f\|_{L^p(G)} := \big(\mathbb{E}_{x\in G}|f(x)|^p\big)^{1/p} \quad \text{and} \quad \|f\|_{\ell^p(G)} := \Big(\sum_{x\in G}|f(x)|^p\Big)^{1/p},$$

with the usual conventions when $p = \infty$. As vector spaces these are all the same (since G is finite), although the norms are different. A specific consequence of this normalisation is that

$$\langle f, g\rangle_{L^2(G)} = \mathbb{E}_{x\in G} f(x)\overline{g(x)} \quad \text{and} \quad \langle f, g\rangle_{\ell^2(G)} = \sum_{x\in G} f(x)\overline{g(x)}.$$

We define the Fourier transform in the usual way, mapping a function $f \in L^1(G)$ to $\widehat{f} \in \ell^\infty(\widehat{G})$, where

$$\widehat{f}(\gamma) := \mathbb{E}_{x\in G} f(x)\overline{\gamma(x)} = \frac{1}{|G|}\sum_{x\in G} f(x)\overline{\gamma(x)}.$$

The significance of the Fourier transform is, in no small part, determined by the effect it has on convolution: recall that if $f, g \in L^1(G)$ then their convolution $f * g$ is defined by

$$(f * g)(x) := \mathbb{E}_{x\in G} f(x)g(y-x).$$

The Fourier transform functions as an algebra isomorphism from $L^1(G)$ under convolution to $\ell^\infty(\widehat{G})$ under pointwise multiplication: $\widehat{f*g} = \widehat{f}\,\widehat{g}$.

(Note that both convolution and the Fourier transform are used on different groups at the same time through this work, and although it is always made clear, the reader should be alert to this.)

We are particularly interested in finite (abelian) groups of exponent 2, all of which are isomorphic to $\mathbb{Z}_2^n$ for some n; to avoid introducing an unnecessary parameter we shall refer to them in the former terms. On these groups the characters correspond to maps $x \mapsto (-1)^{r\cdot x}$, where $r \cdot x$ is the usual bilinear form on $\mathbb{Z}_2^n$ considered as a vector space over $\mathbb{F}_2$.

3. Counting progressions and analytic statement of results

It has been observed in many places that one may estimate the size of the largest subset of an abelian group not containing a three-term arithmetic progression by establishing a lower bound on the number of three-term arithmetic progressions. It should, therefore, come as little surprise that we are interested in the quantity

$$\Lambda(A) := \mathbb{E}_{x,d\in G} 1_A(x)1_A(x+d)1_A(x+2d).$$

which counts three-term arithmetic progressions: specifically $\Lambda(A)|G|^2$ is the number of three-term arithmetic progressions in A.

Denoting by $T(G)$ the number of trivial (that is, nonproper) three-term arithmetic progressions in G, we see that if $\Lambda(A)|G|^2 > T(G)$ then we must have a nontrivial three-term arithmetic progression. This perspective is, perhaps, inspired by an equivalence established in [Varnavides 1959], but we shall not dwell on this relationship here.

Meshulam's theorem is a simple corollary of the following result.

Theorem 3.1. *Suppose that G is a finite abelian group of odd order and $A \subset G$ has density $\alpha > 0$. Then*

$$\Lambda(A) \geqslant \exp(-\alpha^{-O(1)}).$$

To see how Meshulam's theorem follows, note that if G is of odd order then $(x, x+d, x+2d)$ is a proper progression if and only if $d \neq 0_G$. Thus $T(G) = |G|$ and so if $\Lambda(A)|G|^2 > |G|$ then A contains a proper progression; the result follows on inserting the bound for $\Lambda(A)$ from the theorem and rearranging.

Lev effectively removed the odd-order condition from Theorem 3.1:

Theorem 3.2 [Lev 2004]. *Suppose that G is a finite abelian group and $A \subset G$ has density $\alpha > 0$. Then*

$$\Lambda(A) \geqslant \exp(-\alpha^{-O(1)}).$$

In general abelian groups $T(G)$ may be comparable to $|G|^2$ which is why we are not able to conclude Meshulam's theorem without the odd order condition. Indeed, as noted before it is not always true.

It is instructive to consider two examples. First, in $G=\mathbb{Z}_2^n$ one sees that $T(G)=|G|^2$ — all progressions are trivial — so although we have many progressions,[1] none are proper.

Second, the group $G=\mathbb{Z}_4^n$ has $T(G)=|G|^{3/2}+O(|G|)$: any trivial progression (x,y,z) with $x+z=2y$ has $x=z$, $x=y$ or $y=z$. In the first case this implies that $x-y\in\{x'\in G:2x'=0_G\}$; in the second and third cases this implies that all three elements are equal. Thus, in the first case we have $|G|\,\big|\{x'\in G:2x'=0_G\}\big|$ progressions and in the second and third $|G|$ each. This leads to the claimed bound which in turn allows us to establish Meshulam's theorem for $\mathbb{Z}_4^n$.

In this particular case, however, one may proceed directly along the lines of the proof of the Roth–Meshulam theorem (coupled with the aforementioned positivity observation) to establish a stronger bound than in Theorem 3.2.

Theorem 3.3. *Suppose that $G=\mathbb{Z}_4^n$ and $A\subset G$ has density $\alpha>0$. Then*

$$\Lambda(A)\geqslant \exp(-O(\alpha^{-1})).$$

On arranging α large enough so that $\Lambda(A)|G|^2>|G|^{3/2}+O(|G|)$ is guaranteed by the above theorem we get Theorem 1.3; the main result of this paper is the following refinement of Theorem 3.3 which by a similar arrangement implies Theorem 1.4.

Theorem 3.4. *Suppose that $G=\mathbb{Z}_4^n$ and $A\subset G$ has density $\alpha>0$. Then*

$$\Lambda(A)\geqslant \exp\big(-O(\alpha^{-1}\log^{-1/6}\alpha^{-1}\log\log^{5/3}\alpha^{-1})\big).$$

4. Outline of the proof

Our work is strongly influenced by the original Roth–Meshulam–Lev argument; to explain our extra purchase we shall recall a sketch of this. There are basically three ingredients. First, one has a lemma passing from a large Fourier coefficient to increased density on a subgroup.

Lemma 4.1. *Suppose that G is a group of bounded exponent, that $A\subset G$ has density $\alpha>0$, and that $\sup_{\gamma\neq 0_{\widehat{G}}}|\widehat{1_A}(\gamma)|\geqslant\epsilon\alpha$. Then there is a subgroup $G'\leqslant G$ of bounded index such that $\|1_A*\mathbb{P}_{G'}\|_{L^\infty(G)}\geqslant \alpha+\Omega(\alpha\epsilon)$.*

The proof of this is easy and we shall use some similar results in Section 6; we make no improvement on this ingredient and, indeed, the lemma is in many ways best possible.

The core of the argument is the following lemma and it is here that we shall do better. The lemma expresses the fact that either a set A is "uniform" having about the right number of three-term arithmetic progressions or else it has increased density on a subgroup of bounded index.

Lemma 4.2. *Suppose that G is a group of bounded exponent and $A\subset G$ has density $\alpha>0$. Then either $\Lambda(A)=\Omega(\alpha^3)$ or there is a subgroup $G'\leqslant G$ of bounded index such that $\|1_A*\mathbb{P}_{G'}\|_{L^\infty(G)}\geqslant\alpha+\Omega(\alpha^2)$.*

[1] It is easy to see this without Theorem 3.2: $A\subset\mathbb{Z}_2^n$ clearly contains $|A|^2$ progressions since every pair $(x,y)\in A^2$ generates a triple (x,y,x) which is a three-term arithmetic progression in $\mathbb{Z}_2^n$.

Sketch of proof. By the usual application of the inversion formula one has

$$\Lambda(A) = \sum_{\gamma \in \widehat{G}} \widehat{1_A}(\gamma)^2 \, \widehat{1_A}(2\gamma).$$

We write $H := \{\gamma \in \widehat{G} : 2\gamma = 0_{\widehat{G}}\}$, so that

$$\Lambda(A) = \alpha \sum_{\gamma \in H} \widehat{1_A}(\gamma)^2 + O\big(\sup_{\gamma \neq 0_{\widehat{G}}} |\widehat{1_A}(\gamma)|\alpha \big)$$

by Parseval's theorem. Note that if $\gamma \in H$ then γ is a real character, so $|\widehat{1_A}(\gamma)|^2 = \widehat{1_A}(\gamma)^2$; thus we certainly have

$$\sum_{\gamma \in H} \widehat{1_A}(\gamma)^2 = \sum_{\gamma \in H} |\widehat{1_A}(\gamma)|^2 \geqslant |\widehat{1_A}(0_{\widehat{G}})|^2 = \alpha^2.$$

This is the previously mentioned positivity observation of Lev. It follows that either $\Lambda(A) \geqslant \alpha^3/2$ and we are done or $\sup_{\gamma \neq 0_{\widehat{G}}} |\widehat{1_A}(\gamma)| = \Omega(\alpha^2)$ in which case we apply Lemma 4.1 and are done. □

Lemma 4.2 can be iterated to get Theorem 3.3, and again we shall use essentially the same style of iteration in Section 11 to prove Theorem 3.4.

Sketch of proof of Theorem 3.3. We apply the preceding lemma repeatedly, incrementing the density at each stage that we are in the second case of the lemma and terminating if we are in the first case.

At each stage we have $\alpha \mapsto \alpha + \Omega(\alpha^2)$. Thus, after $O(\alpha^{-1})$ iterations the density will have doubled. Since density cannot increase above 1, the iteration terminates after

$$O(\alpha^{-1}) + O((2\alpha)^{-1}) + O((4\alpha)^{-1}) + \cdots = O(\alpha^{-1})$$

steps.

When the iteration terminates we have some group $G' \leqslant G$ with $|G : G'| = \exp(O(\alpha^{-1}))$ such that $\Lambda(A) = \Omega(\alpha^3 |G : G'|^2) = \exp(O(\alpha^{-1}))$. The result follows. □

We shall exploit some of the additional structure of $\mathbb{Z}_4^n$ to effectively improve Lemma 4.2 and thereby gain our strengthening of the Roth–Meshulam–Lev argument.

In $G = \mathbb{Z}_4^n$ a triple (x, y, z) with $x + z = 2y$ must have x and z in the same coset of $\operatorname{im} 2$, where 2 denotes the map $x \mapsto 2x$. Thus it is natural to partition A by the cosets of $\operatorname{im} 2$, because when counting three-term arithmetic progressions we only ever need to consider sums $x + z$ with x and z in the same coset.

Since $\operatorname{im} 2 = \ker 2$ we shall index the elements of this partition of A by elements of $\ker 2$ and, for simplicity later, translate them all so that they lie in $\operatorname{im} 2$. Specifically, then, we proceed as follows.

Suppose that G is a finite abelian group and $A \subset G$. Define

$$f_A : \operatorname{im} 2 \to [0, 1];\, u \mapsto \mathbb{E}_{z \in G : 2z = u} 1_A(z),$$

and note that

$$\begin{aligned}\Lambda(A) &= \mathbb{E}_{x,d \in G} 1_A(x) 1_A(x + 2d) \mathbb{E}_{d' \in G : 2d' = 2d} 1_A(x + d') \\ &= \mathbb{E}_{x,d \in G} 1_A(x) 1_A(x + 2d) f_A(2(x + d)) \\ &= \mathbb{E}_{x,u \in G} 1_A(x) 1_A(2u - x) f_A(2u).\end{aligned}$$

Now, for each $y \in \operatorname{im} 2$ let $t_y \in G$ be such that $2t_y = y$, and let $A_y := A \cap (t_y + \ker 2) - t_y \subset \ker 2$. Furthermore, for each $y \in \operatorname{im} 2$ let τ_y be "translation by y", defined by

$$\tau_y : L^1(\operatorname{im} 2) \to L^1(\operatorname{im} 2);\ f \mapsto (x \mapsto f(x+y)).$$

In this notation we have

$$\begin{aligned}\Lambda(A) &= \mathbb{E}_{y\in \operatorname{im} 2}\mathbb{E}_{x\in t_y+\ker 2, v\in \operatorname{im} 2} 1_{A_y}(x-t_y)1_{A_y}(v-x-t_y)f_A(v)\\ &= \mathbb{E}_{y\in \operatorname{im} 2}\mathbb{E}_{z\in \ker 2, v\in \operatorname{im} 2} 1_{A_y}(z)1_{A_y}(v-y-z)f_A(v)\\ &= \mathbb{E}_{y\in \operatorname{im} 2}\langle \tau_y(1_{A_y} * 1_{A_y}), f_A\rangle_{L^2(\operatorname{im} 2)}. \end{aligned} \tag{4-1}$$

Note that $*$ here denotes convolution on $\ker 2$, since this is where the sets A_y have been arranged to live. In $\mathbb{Z}_4^n$ we have $\operatorname{im} 2 = \ker 2$, which simplifies this expression so that it only involves one group.

Our argument will consider two cases depending on whether or not f_A supports large L^2-mass.

(i) Large L^2-mass: Suppose that $\|f_A\|^2_{L^2(\operatorname{im} 2)} \geqslant \alpha^{5/3}$. Then, on average, A has density $\alpha^{2/3}$ on the fibres of the points in the set $2.A := \{2a : a \in A\}$. We wish to estimate inner products of the form $\langle \tau_y(1_{A_y} * 1_{A_y}), f_A\rangle_{L^2(\operatorname{im} 2)}$, where the set A_y is the fibre of y. Plancherel's theorem tells us that

$$\big\langle \tau_y(1_{A_y} * 1_{A_y}), f_A\big\rangle_{L^2(\operatorname{im} 2)} = \sum_{\gamma\in\widehat{\operatorname{im} 2}} |\widehat{1_{A_y}}(\gamma)|^2\gamma(y)\widehat{f_A}(\gamma) = \alpha_y^2\alpha + O\big(\sup_{\gamma\neq 0_{\widehat{\operatorname{im} 2}}} |\widehat{f_A}(\gamma)|\alpha_y\big),$$

where α_y is the density of the fibre. If $\alpha_y \geqslant \alpha^{2/3}$ then we get a nontrivial character at which $\widehat{f_A}(\gamma) = \Omega(\alpha^{5/3})$. This leads to a corresponding density increment which could only be iterated $O(\alpha^{-2/3})$ times before the density would have to exceed 1.

(ii) Small L^2-mass: Suppose that $\|f_A\|^2_{L^2(\operatorname{im} 2)} \leqslant \alpha^{5/3}$. Then $2.A$ has density at least $\alpha^{1/3}$. We now replace f_A with $1_{2.A}$ and find, in much the same way as above, that we have a nontrivial Fourier mode (this time of a fibre) of size $\Omega(\alpha^{5/3})$. If one could now perform a density increment in a way that was simultaneous for all fibres then this could only happen $O(\alpha^{-2/3})$ times.

These two cases would combine to suggest that A contained $\exp\big(-O(\alpha^{-2/3})\big)$ three-term arithmetic progressions. Unfortunately the second is too optimistic; the content of this paper is in making a version of the sketch above work and, in particular, dealing with the harder case of small L^2-mass.

5. Families

We make a new definition for the remainder of the paper; it will help simplify some later inductive steps and should seem fairly natural given the discussion of the previous section.

Suppose that H is a finite (abelian) group of exponent 2. A *family* on H is a vector $\mathscr{A} = (A_h)_{h\in H}$, where $A_h \subset H$ for all $h \in H$; we call the set A_h a *fibre* of $\mathscr{A}$. We define the *density function* of $\mathscr{A}$ to be

$$f_{\mathscr{A}} : H \to [0, 1];\ h \mapsto \mathbb{P}_H(A_h),$$

and refer to $\mathbb{E}_{x\in H} f_{\mathscr{A}}(x)$ as the *density* of $\mathscr{A}$ denoted $\mathbb{P}_H(\mathscr{A})$.

We are interested in the quantity

$$\Lambda(\mathscr{A}) := \mathbb{E}_{h\in H}\langle \tau_h(1_{A_h} * 1_{A_h}), f_{\mathscr{A}}\rangle_{L^2(H)},$$

and it is useful to note that $|H|^4\Lambda(\mathscr{A})$ is the number of quadruples (a, a', y, h) with $a, a' \in A_h$ and $y \in A_{a+a'-h}$.

If $A \subset \mathbb{Z}_4^n$ then the family $\mathscr{A} := (A_y)_{y\in \mathrm{im}\,2}$ defined earlier for use in (4-1) has $\Lambda(\mathscr{A}) = \Lambda(A)$ and density α. Conversely, given any family $\mathscr{A}$ on $\mathbb{Z}_2^n$ we can clearly construct a set A in $\mathbb{Z}_4^n$ such that $\Lambda(A) = \Lambda(\mathscr{A})$; families are simply a notational convenience. The bulk of the paper now concerns the proof that $\Lambda(\mathscr{A})$ is large in terms of the density of $\mathscr{A}$.

6. Density increments on families

The arguments of this section are straightforward and encode the various ways in which we shall try to increment the density of our family under certain circumstances. The simplest of these is the standard ℓ^∞-density increment lemma which follows.

Lemma 6.1. *Suppose that H is a finite abelian group of exponent 2, $f : H \to [0, 1]$ and γ is a nontrivial character. Then the subgroup $H' := \{\gamma\}^\perp$ has index 2 and $\|f * \mathbb{P}_{H'}\|_{L^\infty(H)} = \mathbb{E}_{h\in H} f(h) + |\widehat{f}(\gamma)|$.*

Proof. Let $h_0 \in H \setminus H'$ so that $h_0 + H'$ is the coset of H' in H not equal to H'. By definition

$$|\widehat{f}(\gamma)| = \big|\mathbb{E}_{h\in H} 1_{H'}(h) f(h) - \mathbb{E}_{h\in H} 1_{h_0+H'}(h) f(h)\big|.$$

We also have

$$\mathbb{E}_{h\in H} f(h) = \mathbb{E}_{h\in H} 1_{H'}(h) f(h) + \mathbb{E}_{h\in H} 1_{h_0+H'}(h) f(h),$$

which on being added to the previous tells us that

$$2\max\big\{\mathbb{E}_{h\in H} 1_{H'}(h) f(h), \mathbb{E}_{h\in H} 1_{h_0+H'}(h) f(h)\big\} = \mathbb{E}_{h\in H} f(h) + |\widehat{f}(\gamma)|.$$

Since the index of H' in H is 2 we have $2(\mathbb{P}_H)|_{H'} = \mathbb{P}_{H'}$, whence

$$\max\big\{(f * \mathbb{P}_{H'})(0_H), (f * \mathbb{P}_{H'})(h_0)\big\} = \mathbb{E}_{h\in H} f(h) + |\widehat{f}(\gamma)|,$$

and the result follows. □

The next lemma is a sort of simultaneous version of the above. If a family has a large number of its fibres having a large Fourier coefficient at the same nontrivial character γ then there is a related family with increased density.

Lemma 6.2. *Suppose that H is a finite abelian group of exponent 2, $\mathscr{A} = (A_h)_{h\in H}$ is a family on H and γ is a nontrivial character. Then there is a subgroup $H' \leqslant H$ of index 2 and a family $\mathscr{A}'$ on H' such that*

$$\Lambda(\mathscr{A}) \geqslant 2^{-4}\Lambda(\mathscr{A}') \quad \textit{and} \quad \mathbb{P}_{H'}(\mathscr{A}') \geqslant \mathbb{P}_H(\mathscr{A}) + \mathbb{E}_{h\in H}|\widehat{1_{A_h}}(\gamma)|.$$

Proof. Let $H' := \{\gamma\}^\perp$ and let $h_0 \in H \setminus H'$ so that $h_0 + H'$ is the coset of H' in H not equal to H'. For each $h \in H$ apply Lemma 6.1 to see that

$$\|1_{A_h} * \mathbb{P}_{H'}\|_{L^\infty(H)} \geqslant \mathbb{E}_{\tilde{h}\in H} 1_{A_h}(\tilde{h}) + |\widehat{1_{A_h}}(\gamma)|.$$

Now let $x_h \in H$ be such that $\|1_{A_h} * \mathbb{P}_{H'}\|_{L^\infty(H)} = (1_{A_h} * \mathbb{P}_{H'})(x_h)$ and define $B_h := A_h \cap (x_h + H') - x_h \subset H'$, whence

$$\mathbb{P}_{H'}(B_h) \geqslant \mathbb{E}_{\tilde{h}\in H} 1_{A_h}(\tilde{h}) + |\widehat{1_{A_h}}(\gamma)| = f_{\mathscr{A}}(h) + |\widehat{1_{A_h}}(\gamma)|.$$

It follows that

$$\mathbb{E}_{h\in H}\mathbb{P}_{H'}(B_h) \geqslant \mathbb{E}_{h\in H} f_{\mathscr{A}}(h) + \mathbb{E}_{h\in H}|\widehat{1_{A_h}}(\gamma)|,$$

whence by averaging we deduce there is a coset $h_1 + H'$ of H such that

$$\mathbb{E}_{h\in h_1+H'}\mathbb{P}_{H'}(B_h) \geqslant \mathbb{E}_{h\in H} f_{\mathscr{A}}(h) + \mathbb{E}_{h\in H}|\widehat{1_{A_h}}(\gamma)|.$$

Now we define a family $\mathscr{A}'$ on H' as follows: for each $h' \in H'$ let $A'_{h'} := B_{h_1+h'}$. Clearly $\mathscr{A}'$ has the required density; it remains to show that $\Lambda(\mathscr{A}) \geqslant 2^{-4}\Lambda(\mathscr{A}')$, which is a relatively simple counting exercise.

There are $|H'|^4\Lambda(\mathscr{A}')$ quadruples (a'_0, a'_1, y', h') with $a'_0, a'_1 \in A'_{h'}$ and $y' \in A'_{a'_0+a'_1-h'}$. Every such quadruple corresponds uniquely to a quadruple

$$(a_0, a_1, y, h) := (a'_0 + x_{h_1+h'}, a'_1 + x_{h_1+h'}, y' + x_{a'_0+a'_1-h'+h_1}, h_1 + h');$$

unique since there is an obvious inverse on the image taking (a_0, a_1, y, h) to

$$(a'_0, a'_1, y', h') = (a_0 - x_h, a_1 - x_h, y - x_{a_0+a_1-h-2x_h+2h_1}, h - h_1).$$

Now,

$$a_0 = a'_0 + x_{h_1+h'} \in A'_h + x_{h_1+h'} = B_{h_1+h'} + x_{h_1+h'} \subset A_{h_1+h'} = A_h,$$

and similarly $a_1 \in A_h$. Furthermore

$$y = y' + x_{a'_0+a'_1-h'+h_1} \in A'_{a'_0+a'_1-h'} + x_{a'_0+a'_1-h'+h_1} = B_{a'_0+a'_1-h'+h_1} + x_{a'_0+a'_1-h'+h_1} \subset A_{a'_0+a'_1-h'+h_1} = A_{a_0+a_1-h},$$

since $2x_{h_1+h'} = 0_H$ and $2h_1 = 0_H$. It follows that every quadruple (a'_0, a'_1, y', h') with $a'_0, a'_1 \in A'_{h'}$ and $y' \in A'_{a'_0+a'_1-h'}$ corresponds to a unique quadruple (a_0, a_1, y, h) with $a_0, a_1 \in A_h$ and $y \in A_{a_0+a_1-h}$, whence

$$|H'|^4\Lambda(\mathscr{A}) \leqslant |H|^4\Lambda(\mathscr{A}).$$

The result follows on noting that $|H|^4 = 2^4|H'|^4$. □

The last part of this proof was a rather fiddly verification of a type which we shall have to do repeatedly, and while we were comprehensive in the details above, in the future we shall include fewer of them.

The final lemma of the section takes a family where the density function is nonuniform and produces a new family with a larger density, again very much in the spirit of the previous two lemmas.

Lemma 6.3. *Suppose that H is a finite abelian group of exponent 2, $\mathscr{A} = (A_h)_{h\in H}$ is a family on H and γ is a nontrivial character. Then there is a subgroup $H' \leqslant H$ of index 2 and a family $\mathscr{A}'$ on H' such that*

$$\Lambda(\mathscr{A}) \geqslant 2^{-4}\Lambda(\mathscr{A}') \quad \textit{and} \quad \mathbb{P}_{H'}(\mathscr{A}') \geqslant \mathbb{P}_H(\mathscr{A}) + |\widehat{f_{\mathscr{A}}}(\gamma)|.$$

Proof. Let $H' := \{\gamma\}^{\perp}$ and let $h_0 \in H \setminus H'$ so that $h_0 + H'$ is the coset of H' in H not equal to H'. Apply Lemma 6.1 so that we have

$$\|f_{\mathscr{A}} * \mathbb{P}_{H'}\|_{L^\infty(H)} = \mathbb{P}_H(\mathscr{A}) + |\widehat{f_{\mathscr{A}}}(\gamma)|.$$

Let $h_1 \in H'$ be such that $(f_{\mathscr{A}} * \mathbb{P}_{H'})(h_1) = \|f_{\mathscr{A}} * \mathbb{P}_{H'}\|_{L^\infty(H)}$. Now, define a family $\mathscr{A}'$ as follows: for each $h' \in H'$

(i) if $A_{h_1+h'} \cap H'$ is larger than $A_{h_1+h'} \cap (h_0 + H')$ then put $x_{h_1+h'} := 0_H$ and $A'_{h'} := A_{h_1+h'} \cap H'$;

(ii) otherwise put $x_{h_1+h'} := h_0$ and $A'_{h'} := A_{h_1+h'} \cap (h_0 + H') - h_0$.

By averaging we have $\mathbb{P}_{H'}(A'_{h'}) \geqslant \mathbb{P}_H(A_{h_1+h'})$, whence

$$\mathbb{E}_{h'\in H'}\mathbb{P}_{H'}(A'_{h'}) \geqslant (f_{\mathscr{A}} * \mathbb{P}_{H'})(h_1) = \mathbb{P}_H(\mathscr{A}) + |\widehat{f_{\mathscr{A}}}(\gamma)|,$$

which yields the required density condition. It remains, as before, to show that $\Lambda(\mathscr{A}) \geqslant 2^{-4}\Lambda(\mathscr{A}')$; we proceed as in the previous lemma.

There are $|H'|^4\Lambda(\mathscr{A}')$ quadruples (a'_0, a'_1, y', h') with $a'_0, a'_1 \in A'_{h'}$ and $y' \in A'_{a'_0+a'_1-h'}$. Every such quadruple corresponds uniquely to a quadruple

$$(a_0, a_1, y, h) := (a'_0 + x_{h_1+h'}, a'_1 + x_{h_1+h'}, y' + x_{a'_0+a'_1-h'+h_1}, h_1 + h')$$

with $a_0, a_1 \in A_h$ and $y \in A_{a_0+a_1-h}$, whence $|H'|^4\Lambda(\mathscr{A}) \leqslant |H|^4\Lambda(\mathscr{A})$ and the result follows on noting that $|H|^4 = 2^4|H'|^4$. □

7. Families with large mean square density

In this section we show how a family $\mathscr{A}$ for which $\|f_{\mathscr{A}}\|_{L^2(H)}$ is large (compared with its trivial lower bound of $\mathbb{P}_H(\mathscr{A})^2$) has $\Lambda(\mathscr{A})$ large. The basic idea is that if $\|f_{\mathscr{A}}\|_{L^2(H)}$ is large then most of the fibres A_h have large density and so are more easily "uniformised". When they are uniform the count $\Lambda(\mathscr{A})$ is easily seen to be large.

It is instructive to consider a simplified situation. Suppose that $\mathscr{A}$ is a family which is assumed to have fibres of density either 0 or δ and the support of $f_{\mathscr{A}}$ has density σ. This family has density $\delta\sigma$ and $\|f_{\mathscr{A}}\|^2_{L^2(H)} = \delta^2\sigma$, which is large compared with the trivial lower bound of $(\delta\sigma)^2$ if σ is small. Now, the standard Roth–Meshulam argument can be used to show that $\Lambda(\mathscr{A}) = \exp\big(O(\delta^{-1}\sigma^{-1})\big)$, and the proposition below asserts that this can be improved when σ is small.

Proposition 7.1. *Suppose that H is a finite abelian group of exponent 2 and $\mathscr{A} = (A_h)_{h\in H}$ is a family on H such that $f_{\mathscr{A}} = \delta 1_S$ for some $\delta \in (0, 1]$ and $S \subset H$ of density σ. Then $\Lambda(\mathscr{A}) = \exp\big(-O(\delta^{-1}\log\sigma^{-1})\big)$.*

Naturally the proof is iterative with the following lemma acting as the driver.

Lemma 7.2. *Suppose that H is a finite abelian group of exponent 2, $\mathscr{A} = (A_h)_{h\in H}$ is a family on H and $f_{\mathscr{A}} = \delta 1_S$ for some $\delta \in (0, 1]$ and $S \subset H$ of density σ. Then either*

$$\Lambda(\mathscr{A}) \geqslant \delta^3\sigma^2/2$$

or there is a subgroup $H' \leqslant H$ of index 2, a family $\mathscr{A}'$ on H' and set $S' \subset H'$ such that $f_{\mathscr{A}'} = \delta 1_{S'}$ and

$$\mathbb{P}_{H'}(S') \geqslant \sigma(1 + \delta/2) \quad \textit{and} \quad \Lambda(\mathscr{A}) \geqslant 2^{-4}\Lambda(\mathscr{A}').$$

Proof. Since $f_{\mathscr{A}} = \delta 1_S$ we have

$$\Lambda(\mathscr{A}) = \delta\mathbb{E}_{h\in H}\langle\tau_h(1_{A_h} * 1_{A_h}), 1_S\rangle_{L^2(H)}.$$

Applying Plancherel's theorem to the inner products we get

$$\Lambda(\mathscr{A}) = \delta\mathbb{E}_{h\in H}\sum_{\gamma\in\widehat{H}} |\widehat{1_{A_h}}(\gamma)|^2\widehat{1_S}(\gamma)\gamma(h).$$

The triangle inequality may be used on these inner sums to separate out the trivial mode. Indeed, since $\widehat{1_{A_h}}(0_{\widehat{H}}) = f_{\mathcal{A}}(h)$ and $\widehat{1_S}(0_{\widehat{H}}) = \sigma$ we get, after a little manipulation,

$$\mathbb{E}_{h\in H}\sum_{\gamma\neq 0_{\widehat{H}}} |\widehat{1_{A_h}}(\gamma)|^2|\widehat{1_S}(\gamma)| \geqslant \mathbb{E}_{h\in H} f_{\mathcal{A}}(h)^2\sigma - \delta^{-1}\Lambda(\mathcal{A}) = \delta^2\sigma^2 - \delta^{-1}\Lambda(\mathcal{A}).$$

Now, we are done unless $\Lambda(\mathcal{A}) \leqslant \delta^3\sigma^2/2$ (in fact, unless $\Lambda(\mathcal{A}) < \delta^3\sigma^2/2$, but we shall not use this), whence

$$\mathbb{E}_{h\in H}\sum_{\gamma\neq 0_{\widehat{H}}} |\widehat{1_{A_h}}(\gamma)|^2|\widehat{1_S}(\gamma)| \geqslant \delta^2\sigma^2/2.$$

On the other hand,

$$\mathbb{E}_{h\in H}\sum_{\gamma\neq 0_{\widehat{H}}} |\widehat{1_{A_h}}(\gamma)|^2 = \mathbb{E}_{h\in H}(f_{\mathcal{A}}(h) - f_{\mathcal{A}}(h)^2) = \delta(1-\delta)\sigma \leqslant \delta\sigma,$$

by Parseval's theorem. Using this with the triangle inequality in the previous expression tells us that S is linearly biased:

$$\sup_{\gamma\neq 0_{\widehat{H}}} |\widehat{1_S}(\gamma)| \geqslant \delta\sigma/2.$$

Thus, by Lemma 6.1 there is a subgroup $H' \leqslant H$ of index 2 such that

$$\|1_S * \mathbb{P}_{H'}\|_{L^\infty(H)} \geqslant \sigma(1+\delta/2). \tag{7-1}$$

Let $h_1 \in H$ be such that $(1_S*\mathbb{P}_{H'})(h_1) = \|1_S*\mathbb{P}_{H'}\|_{L^\infty(H)}$ and define a family $\mathcal{A}' := (A'_{h'})_{h'\in H'}$ as follows. For each $h' \in H'$ let $x_{h'+h_1}$ be such that $(1_{A_{h'+h_1}}*\mathbb{P}_{H'})(x_{h'+h_1})$ is maximal. If $(1_{A_{h'+h_1}}*\mathbb{P}_{H'})(x_{h'+h_1}) > 0$ then

$$0 < (1_{A_{h'+h_1}}*\mathbb{P}_{H'})(x_{h'+h_1})/2 \leqslant f_{\mathcal{A}}(h'+h_1) = \delta 1_S(h'+h_1),$$

whence

$$1_{A_{h'+h_1}}*\mathbb{P}_{H'}(x_{h'+h_1}) \geqslant \mathbb{E}_{h\in H} 1_{A_{h'+h_1}}(h) = f_{\mathcal{A}}(h'+h_1) = \delta,$$

and $A_{h'+h_1} \cap (x_{h'+h_1} + H') - x_{h'+h_1}$ contains a set of density δ; let $A'_{h'}$ be such a set. If

$$(1_{A_{h'+h_1}}*\mathbb{P}_{H'})(x_{h'+h_1}) = 0$$

then let $A'_{h'} = \varnothing$. Finally, we write $S' := S \cap (h_1 + H') - h_1$ and it remains to check that we have the required properties.

First, note that

$$f_{\mathcal{A}'}(h') = \mathbb{P}_{H'}(A'_{h'}) \leqslant 2\mathbb{P}_H(A_{h'+h_1}) = 2f_{\mathcal{A}}(h'+h_1),$$

thus if $f_{\mathcal{A}'}(h') > 0$ then $h'+h_1 \in S$ and so $h' \in S'$. Similarly,

$$f_{\mathcal{A}'}(h') = \mathbb{P}_{H'}(A'_{h'}) \geqslant \mathbb{P}_H(A_{h'+h_1}) = 2f_{\mathcal{A}}(h'+h_1),$$

so if $f_{\mathcal{A}'}(h') = 0$ then $h'+h_1 \notin S$, whence $h' \notin S'$. By design, $f_{\mathcal{A}'}$ takes only the values 0 and δ and so we have the representation $f_{\mathcal{A}'} = \delta 1_{S'}$.

Secondly, we have $\mathbb{P}_{H'}(S') = 1_S * \mathbb{P}_{H'}(h_1)$, whence $\mathbb{P}_{H'}(S') \geqslant \sigma(1+\delta/2)$ by (7-1). Lastly, we check that $\Lambda(\mathcal{A}) \geqslant 2^{-4}\Lambda(\mathcal{A}')$ in the usual fashion.

There are $|H'|^4\Lambda(\mathscr{A}')$ quadruples (a_0', a_1', y', h') with $a_0', a_1' \in A'_{h'}$ and $y' \in A'_{a_0'+a_1'-h'}$. Every such quadruple corresponds uniquely to a quadruple

$$(a_0, a_1, y, h) := (a_0' + x_{h_1+h'}, a_1' + x_{h_1+h'}, y' + x_{a_0'+a_1'-h'+h_1}, h_1 + h')$$

with $a_0, a_1 \in A_h$ and $y \in A_{a_0+a_1-h}$, whence $|H'|^4\Lambda(\mathscr{A}) \leqslant |H|^4\Lambda(\mathscr{A})$ and the result follows on noting that $|H|^4 = 2^4|H'|^4$. □

Proof of Proposition 7.1. Let $H_0 := H$, $\mathscr{A}_0 := \mathscr{A}$, $\alpha_0 := \delta\sigma$, $S_0 := S$ and $\sigma_0 := \sigma$. Suppose that we have a finite abelian group H_i of exponent 2 with a family $\mathscr{A}_i$ on H_i of density α_i and a set S_i of density σ_i such that $f_{\mathscr{A}_i} = \delta 1_{S_i}$. Apply Lemma 7.2 to see that either

$$\Lambda(\mathscr{A}_i) \geqslant \delta^3\sigma_i^2/2,$$

or there is a subgroup H_{i+1} of index 2 in H_i, a family $\mathscr{A}_{i+1}$ and a set S_{i+1} such that

$$f_{\mathscr{A}_{i+1}} = \delta 1_{S_{i+1}}, \sigma_{i+1} \geqslant \sigma_i(1+\delta/2) \quad \text{and} \quad \Lambda(\mathscr{A}_i) \geqslant 2^{-4}\Lambda(\mathscr{A}_{i+1}).$$

Since $\sigma_i \leqslant 1$ we see that this iteration must terminate at some stage i with $(1+\delta/2)^i \leqslant \sigma^{-1}$, that is, with $i \leqslant 2\delta^{-1}\log\sigma^{-1}$. It follows that

$$\Lambda(\mathscr{A}) \geqslant 2^{-8\delta^{-1}\log\sigma^{-1}}\delta^3\sigma^2/2,$$

which is the result. □

Proposition 7.1 will be used again in Section 9 but it may seem like the rather special form of the family considered is too restrictive. However, a standard dyadic decomposition lets us apply this proposition to an arbitrary family; we gain precisely in the case when $\|f_{\mathscr{A}}\|_{L^2(H)}^2\alpha^{-2} \to \infty$.

Corollary 7.3. *Suppose that H is a finite abelian group of exponent 2, $\mathscr{A} = (A_h)_{h\in H}$ is a family on H of density α and $\|f_{\mathscr{A}}\|_{L^2(H)} = K\alpha^2$ for some $K \geqslant 2$. Then*

$$\Lambda(\mathscr{A}) = \exp\bigl(-O(\alpha^{-1}K^{-1}\log^2 K)\bigr).$$

Proof. Let $S_i := \{h \in H : 2^{-(i+1)} \leqslant f_{\mathscr{A}}(h) \leqslant 2^{-i}\}$ and $S' := \{h \in H : f_{\mathscr{A}}(h) \leqslant \alpha/2\}$. We may use these sets to partition the range of summation in $\|f_{\mathscr{A}}\|_{L^2(H)}^2$: by the triangle inequality

$$\sum_{i\leqslant\lceil\log_2\alpha^{-1}\rceil} 2^{-2i}\mathbb{P}_H(S_i) + (\alpha/2)^2 \geqslant \|f_{\mathscr{A}}\|_{L^2(H)}^2.$$

The Cauchy–Schwarz inequality tells us that $\|f_{\mathscr{A}}\|_{L^2(H)}^2 \geqslant \alpha^2$, whence

$$\sum_{i\leqslant\lceil\log_2\alpha^{-1}\rceil} 2^{-2i}\mathbb{P}_H(S_i) \geqslant 3\|f_{\mathscr{A}}\|_{L^2(H)}^2/4. \tag{7-2}$$

Now let $\epsilon \in (0, 1]$ be a parameter to be chosen later and note that

$$\sum_{i\leqslant\lceil\log_2\alpha^{-1}\rceil} 2^{\epsilon i} \leqslant 2\alpha^{-\epsilon}\sum_{i\leqslant\lceil\log_2\alpha^{-1}\rceil} 2^{\epsilon(i-\lceil\log_2\alpha^{-1}\rceil)} \leqslant 2\alpha^{-\epsilon}\sum_{j=0}^{\infty}2^{-\epsilon j} = \frac{2\alpha^{\epsilon}}{1-2^{-\epsilon}} \leqslant 2\epsilon^{-1}\alpha^{-\epsilon}.$$

Returning to (7-2) we see that

$$\sum_{i \leqslant \lceil \log_2 \alpha^{-1} \rceil} 2^{\epsilon i}\, 2^{-(2+\epsilon)i} \mathbb{P}_H(S_i) \geqslant 3\|f_{\mathscr{A}}\|_{L^2(H)}^2/4,$$

and so by averaging from our previous calculation there is some $i \leqslant \lceil \log_2 \alpha^{-1} \rceil$ such that

$$(2\epsilon^{-1}\alpha^{-\epsilon})\, 2^{-(2+\epsilon)i} \mathbb{P}_H(S_i) \geqslant 3\|f_{\mathscr{A}}\|_{L^2(H)}^2/4. \tag{7-3}$$

Moreover $2^{-(i+1)}\mathbb{P}_H(S_i) \leqslant \mathbb{E}_{h\in H} f_{\mathscr{A}}(h) = \alpha$; so, recalling that $\|f_{\mathscr{A}}\|_{L^2(H)}^2 = K\alpha^2$ we have

$$2^{-(1+\epsilon)i} \geqslant 3\epsilon K\alpha^{1+\epsilon}/16.$$

If we take $\epsilon = 1/(1+\log K)$, we get

$$2^{-(i+1)} = \Omega(\alpha K/\log K). \tag{7-4}$$

Let $\mathscr{A}'$ be a family defined as follows. If $h \in S_i$ then A'_h is a subset of A_h of density $2^{-(i+1)}$ and A'_h is empty otherwise. By comparison of the terms in $\Lambda(\mathscr{A})$ with those in $\Lambda(\mathscr{A}')$ we see that $\Lambda(\mathscr{A}) \geqslant \Lambda(\mathscr{A}')$.

We now apply Proposition 7.1 to $\mathscr{A}'$; it is easy to see from (7-3) and (7-4) that $\delta^{-1} = 2^{(i+1)} = O(\alpha^{-1}K^{-1}\log K)$ and

$$\log \sigma^{-1} = \log \mathbb{P}_H(S_i)^{-1} = O\big(\log((\delta\alpha^{-1})^{2+\epsilon}K^{-1}\log K)\big) = O(\log K\delta\alpha^{-1}).$$

The result follows on noting that

$$\Lambda(\mathscr{A}) = \exp\big(-O(\delta^{-1}\log K\delta\alpha^{-1})\big)$$

increases as δ decreases. □

It should be noted that one cannot completely remove the logarithmic term in this corollary. We might have $K \sim \alpha^{-1}$, but $\Lambda(\mathscr{A})$ may still be $\exp(-\Omega(\log K))$. To see this consider, for example, the family $\mathscr{A}$, where every fibre A_h is a random set of density α. Of course, the logarithmic power will not significantly affect our final result and is only critical when K is much smaller than α^{-1}, in which case it may be possible to remove it entirely.

8. A quasirandom Balog–Szemerédi–Gowers–Freĭman theorem

The Balog–Szemerédi–Gowers–Freĭman theorem is a now ubiquitous result in additive combinatorics introduced by Gowers [1998]. It combines (a refined proof of) the Balog–Szemerédi theorem [1994] with the structure theorem of Freĭman [1973] concerning sets with small sum set. Since we are working in finite abelian groups of exponent 2 we actually require the far easier torsion version of Freĭman's theorem proved in [Ruzsa 1999]. In fact, in this setting a version of the Balog–Szemerédi–Gowers–Freĭman theorem is known with relatively good bounds.

Theorem 8.1 [Green and Tao 2009, Theorem 1.7]. *Suppose that H is a group of exponent 2, $A \subset H$ has density α and $\|1_A * 1_A\|_{L^2(H)}^2 \geqslant c\alpha^3$. Then there is an element $x \in H$ and a subgroup $H' \leqslant H$ such that*

$$\mathbb{P}_H(H') = \exp\big(-O(c^{-1}\log c^{-1})\big)\alpha \quad \textit{and} \quad (1_A * \mathbb{P}_{H'})(x) \geqslant c/2.$$

We actually require a slightly modified version of this result which also ensures that A' behaves uniformly on H'. This can essentially be read out of the proof in [Green and Tao 2009]; however, for completeness, we include a "decoupled" proof here.

Corollary 8.2. *Suppose that H is a group of exponent 2, $A \subset H$ has density α and $\|1_A * 1_A\|_{L^2(H)}^2 \geqslant c\alpha^3$, and $\epsilon \in (0, 1]$ is a parameter. Then there is an element $x \in H$ and a subgroup $H' \leqslant H$ such that*

$$\mathbb{P}_H(H') = \exp\bigl(-O((c^{-1}+\epsilon^{-1})\log c^{-1})\bigr)\alpha \quad \textit{and} \quad (1_A * \mathbb{P}_{H'})(x) \geqslant c/2,$$

and writing $A' := A \cap (x + H') - x \subset H'$ one has

$$\sup_{\gamma \neq 0_{\widehat{H'}}} |\widehat{1_{A'}}(\gamma)| \leqslant \epsilon \mathbb{P}_{H'}(A').$$

Proof. We apply Theorem 8.1 to get an element $x_0 \in H$ and a subgroup $H_0 \leqslant H$ such that

$$\mathbb{P}_H(H_0) = \exp\bigl(-O(c^{-1}\log c^{-1})\bigr) \quad \text{and} \quad (1_A * \mathbb{P}_{H_0})(x_0) \geqslant c/2.$$

Put $A_0 := A \cap (x_0 + H_0) - x_0 \subset H_0$ and $\alpha_0 := \mathbb{P}_{H_0}(A_0)$. Now, suppose that we have been given an element $x_i \in H$, a subgroup H_i and a subset A_i of H_i of density α_i. If

$$\sup_{\gamma \neq 0_{\widehat{H_i}}} |\widehat{1_{A_i}}(\gamma)| \leqslant \epsilon \alpha_i, \tag{8-1}$$

then we terminate the iteration; otherwise we apply Lemma 6.1 to get a subgroup H_{i+1} of index 2 in H_i such that

$$\|1_{A_i} * \mathbb{P}_{H_{i+1}}\|_{L^\infty(H_i)} \geqslant \alpha_i(1+\epsilon).$$

Let x_{i+1} be such that $(1_{A_i} * \mathbb{P}_{H_{i+1}})(x_{i+1}) = \|1_{A_i} * \mathbb{P}_{H_{i+1}}\|_{L^\infty(H_i)}$, and $A_{i+1} = A_i \cap (x_{i+1} + H_{i+1}) - x_{i+1} \subset H_{i+1}$.

Since $\alpha_i \leqslant 1$ we see that this iteration must terminate at some stage i with $(1+\epsilon)^i \leqslant \alpha_0^{-1}$, that is, with $i \leqslant \epsilon^{-1}\log \alpha_0^{-1} = O(\epsilon^{-1}\log c^{-1})$. We put $x := x_0 + \cdots + x_i$ and $H' := H_i$ so that H_i has index $O(\epsilon^{-1}\log c^{-1})$ in H_0 and $A' = A_i$ has density at least $c/2$. Thus

$$\mathbb{P}_H(H') = \mathbb{P}_H(H_0)\,\mathbb{P}_{H_0}(H_k) = \exp\bigl(-O((c^{-1}+\epsilon^{-1})\log c^{-1})\bigr)\alpha,$$

and it remains to note that the final condition of the corollary holds in view of the fact that we must have (8-1) for the iteration to terminate. □

The iteration in this proof is essentially the iteration at the core of the usual Roth–Meshulam argument (given in the sketch proof in Section 4) and consequently if one could improve the ϵ-dependence in the above result one could probably improve the Roth–Meshulam argument directly. Unfortunately in our use of this corollary ϵ and c are comparable; thus, even in the presence of Marton's conjecture, more commonly called the polynomial Freĭman–Ruzsa conjecture [Green 2005], we would see no significant improvement in our final result.

9. Families with high fibered energy

In this section we use our previous work to show that if a family $\mathscr{A}$ has large additive energy in its fibres then $\Lambda(\mathscr{A})$ is large. The actual statement of the result is rather technical so we take a moment now to sketch the approach.

The key tool is the corollary of the Balog–Szemerédi–Gowers–Freĭman theorem established in Section 8. This may be applied individually to the fibres of $\mathscr{A}$ in each case, producing a subgroup on which the fibre is very dense. If all of these subgroups are very different then it is easy to see that $\Lambda(\mathscr{A})$ must be large; if not then by expanding them a little bit we find one subgroup on which a lot of fibres of $\mathscr{A}$ are very dense and we may use Proposition 7.1 to get that $\Lambda(\mathscr{A})$ is large.

Concretely, then, the purpose of this section is to prove the following.

Lemma 9.1. *Suppose that H is a finite abelian group of exponent 2, and $\mathscr{A} = (A_h)_{h\in H}$ is a family on H of density α such that*

$$\sup_{\gamma\neq 0_{\widehat{H}}} |\widehat{f_{\mathscr{A}}}(\gamma)| \leqslant L\alpha^2,$$

for some parameter $L \geqslant 1$. Suppose further that S is a set of density σ and $K \geqslant 1$ is a parameter such that

(i) *$K\alpha \geqslant f_{\mathscr{A}}(h) \geqslant K\alpha/2$ for all $h \in S$;*

(ii) *and $\|1_{A_h} * 1_{A_h}\|^2_{L^2(H)} \geqslant cf_{\mathscr{A}}(h)^3$ for all $h \in S$.*

Then

$$\Lambda(\mathscr{A}) \geqslant \exp\Big(-O\Big(L(\log^2\alpha^{-1} + \log\sigma^{-1})\,\exp\big(O\big((c^{-1} + K^{1/2}c^{-1/2})\log c^{-1}\big)\big)\Big)\Big).$$

Proof. By Corollary 8.2 (with $\epsilon = 2^{-2}\sqrt{c/K}$) we see that for each $h \in S$ there is an element $x_h \in H$ and a subgroup $H_h \leqslant H$ such that the set $A'_h := A_h \cap (x_h + H_h) - x_h$ has

$$\mathbb{P}_H(H_h) = \exp\big(-O\big((c^{-1} + K^{1/2}c^{-1/2})\log c^{-1}\big)\big) f_{\mathscr{A}}(h) \quad \text{and} \quad \mathbb{P}_{H_h}(A'_h) \geqslant c/2,$$

and, furthermore,

$$\sup_{\gamma\neq 0_{\widehat{H_h}}} |\widehat{1_{A'_h}}(\gamma)| \leqslant \epsilon\mathbb{P}_{H_h}(A'_h). \tag{9-1}$$

Now, let $S_0 := \{h \in S : (f_{\mathscr{A}} * \mathbb{P}_{H_h})(h) \geqslant \alpha/2\}$ and $S_1 := S \setminus S_0$; we shall now split into two cases according to which of S_0 or S_1 is larger.

Case 1. Suppose that $\mathbb{P}_H(S_0) \geqslant \sigma/2$. Then

$$\Lambda(\mathscr{A}) \geqslant \alpha^3\sigma\exp\big(-O\big((c^{-1} + K^{1/2}c^{-1/2})\log c^{-1}\big)\big).$$

Proof. By nonnegativity of the terms in $\Lambda(\mathscr{A})$ we have

$$\Lambda(\mathscr{A}) \geqslant \mathbb{E}_{h\in H} 1_{S_0}(h)\langle\tau_h(1_{A_h} * 1_{A_h}), f_{\mathscr{A}}\rangle_{L^2(H)}.$$

We analyse these inner products individually. Suppose that $h \in S_0$ and note that

$$\big\langle\tau_h(1_{A_h} * 1_{A_h}), f_{\mathscr{A}}\big\rangle_{L^2(H)} \geqslant \mathbb{P}_H(H_h)^2\langle\tau_h(1_{A'_h} * 1_{A'_h}), f_{\mathscr{A}}\rangle_{L^2(h+H_h)}.$$

As usual this inner product is analysed using the Fourier transform: by Plancherel's theorem we have

$$\big\langle 1_{A'_h} * 1_{A'_h}, \tau_{-h}(f_{\mathscr{A}})\big\rangle_{L^2(H_h)} = \sum_{\gamma\in\widehat{H_h}} |\widehat{1_{A'_h}}(\gamma)|^2\,\widehat{\tau_{-h}(f_{\mathscr{A}})}(\gamma).$$

Separating out the contribution from the trivial character we get

$$\langle 1_{A'_h} * 1_{A'_h}, \tau_{-h}(f_{\mathscr{A}})\rangle_{L^2(H_h)} \geqslant \mathbb{P}_{H_h}(A'_h)^2 (f_{\mathscr{A}} * \mathbb{P}_{H_h})(h) - \sum_{\gamma \neq 0_{\widehat{H_h}}} \left|\widehat{1_{A'_h}}(\gamma)\right|^2 \left|\widehat{\tau_{-h}(f_{\mathscr{A}})}(\gamma)\right|. \tag{9-2}$$

This last term sum can be estimated as follows using Hölder's inequality and the Cauchy–Schwarz inequality:

$$\begin{aligned}\sum_{\gamma \neq 0_{\widehat{H_h}}} \left|\widehat{1_{A'_h}}(\gamma)\right|^2 \left|\widehat{\tau_{-h}(f_{\mathscr{A}})}(\gamma)\right| &\leqslant \sup_{\gamma \neq 0_{\widehat{H_h}}} \left|\widehat{1_{A'_h}}(\gamma)\right| \sum_{\gamma \in \widehat{H_h}} \left|\widehat{1_{A'_h}}(\gamma)\right| \left|\widehat{\tau_{-h}(f_{\mathscr{A}})}(\gamma)\right| \\ &\leqslant \sup_{\gamma \neq 0_{\widehat{H_h}}} \left|\widehat{1_{A'_h}}(\gamma)\right| \left(\sum_{\gamma \in \widehat{H_h}} \left|\widehat{1_{A'_h}}(\gamma)\right|^2\right)^{1/2} \left(\sum_{\gamma \in \widehat{H_h}} \left|\widehat{\tau_{-h}(f_{\mathscr{A}})}(\gamma)\right|^2\right)^{1/2}\end{aligned}$$

By Parseval's theorem,

$$\sum_{\gamma \in \widehat{H_h}} \left|\widehat{1_{A'_h}}(\gamma)\right|^2 = \mathbb{P}_{H_h}(A'_h) \quad \text{and} \quad \sum_{\gamma \in \widehat{H_h}} \left|\widehat{\tau_{-h}(f_{\mathscr{A}})}(\gamma)\right|^2 = \left|f_{\mathscr{A}}\right|^2_{L^2(h+H_h)},$$

and combining all this with (9-1) tells us that

$$\sum_{\gamma \neq 0_{\widehat{H_h}}} \left|\widehat{1_{A'_h}}(\gamma)\right|^2 \left|\widehat{\tau_{-h}(f_{\mathscr{A}})}(\gamma)\right| \leqslant \epsilon \mathbb{P}_{H_h}(A'_h)^{3/2} \|f_{\mathscr{A}}\|_{L^2(h+H_h)} \leqslant \epsilon \mathbb{P}_{H_h}(A'_h)^{3/2} \sqrt{2K} (f_{\mathscr{A}} * \mathbb{P}_{H_h})(h).$$

The last inequality here follows from the fact that $h \in S_0$ ensures that $f_{\mathscr{A}}(h) \leqslant K\alpha$ and $(f_{\mathscr{A}} * \mathbb{P}_{H_h})(h) \geqslant \alpha/2$. Finally, our choice of ϵ tells us that

$$\sum_{\gamma \neq 0_{\widehat{H_h}}} |\widehat{1_{A'_h}}(\gamma)|^2 |\widehat{\tau_{-h}(f_{\mathscr{A}})}(\gamma)| \leqslant \mathbb{P}_{H_h}(A'_h)^2 (f_{\mathscr{A}} * \mathbb{P}_{H_h})(h)/2,$$

whence, inserting this in (9-2), we get

$$\langle 1_{A'_h} * 1_{A'_h}, \tau_{-h}(f_{\mathscr{A}})\rangle_{L^2(H_h)} \geqslant \mathbb{P}_{H_h}(A'_h)^2 (f_{\mathscr{A}} * \mathbb{P}_{H_h})(h)/2.$$

Thus, our earlier averaging tells us that

$$\Lambda(\mathscr{A}) \geqslant \mathbb{E}_{h \in H} 1_{S_0}(h) \mathbb{P}_H(H_h)^2\, \mathbb{P}_{H_h}(A'_h)^2 (f_{\mathscr{A}} * \mathbb{P}_{H_h})(h)/2,$$

and hence immediately that

$$\Lambda(\mathscr{A}) \geqslant 2^{-3} c^2 \alpha \mathbb{E}_{h \in H} 1_{S_0}(h) \mathbb{P}_H(H_h)^2 = \alpha^3 \sigma \exp\Big(-O\big((c^{-1} + K^{1/2} c^{-1/2}) \log c^{-1}\big)\Big).$$

The case is complete. □

Case 2. Suppose that $\mathbb{P}_H(S_1) \geqslant \sigma/2$. Then

$$\Lambda(\mathscr{A}) \geqslant \exp\Big(-O\Big(L(\log^2 \alpha^{-1} + \log \sigma^{-1}) \exp\big(O\big((c^{-1} + K^{1/2} c^{-1/2}) \log c^{-1}\big)\big)\Big)\Big).$$

Proof. Suppose that $h \in S_1$ so that $(f_{\mathscr{A}} * \mathbb{P}_{H_h})(h) \leqslant \alpha/2$. By the Fourier inversion formula we have

$$\sum_{\gamma \in H_h^{\perp}} \widehat{f_{\mathscr{A}}}(\gamma)\gamma(h) = (f_{\mathscr{A}} * \mathbb{P}_{H_h})(h) \leqslant \alpha/2.$$

Separating out the trivial mode where $\widehat{f_{\mathscr{A}}}(0_{\widehat{H}}) = \alpha$ and applying the triangle inequality we have

$$\sum_{0_{\widehat{H}} \neq \gamma \in H_h^{\perp}} |\widehat{f_{\mathscr{A}}}(\gamma)| \geqslant \alpha/2.$$

Write $\mathscr{L}' := \{\gamma : |\widehat{f_{\mathscr{A}}}(\gamma)| \geqslant \mathbb{P}_H(H_h)\alpha/4\}$ and note that since $|H_h^{\perp}| = \mathbb{P}_H(H_h)^{-1}$ we have

$$\sum_{0_{\widehat{H}} \neq \gamma \in \mathscr{L}' \cap H_h^{\perp}} |\widehat{f_{\mathscr{A}}}(\gamma)| \geqslant \sum_{0_{\widehat{H}} \neq \gamma \in H_h^{\perp}} |\widehat{f_{\mathscr{A}}}(\gamma)| - \sum_{\gamma \in H_h^{\perp} \setminus \mathscr{L}'} |\widehat{f_{\mathscr{A}}}(\gamma)| \geqslant \alpha/4.$$

Since $\sup_{\gamma \neq 0_{\widehat{H}}} |\widehat{f_{\mathscr{A}}}(\gamma)| \leqslant L\alpha^2$ we conclude that

$$|\mathscr{L}' \cap H_h^{\perp}| \geqslant \alpha^{-1}/4L.$$

Let $I_h \subset \mathscr{L}' \cap H_h^{\perp}$ be a set of $d := \lfloor \log_2(\alpha^{-1}/4L) \rfloor$ independent elements — possible since $2^d \leqslant |\mathscr{L}' \cap H_h^{\perp}|$ — and put $H_h' := I_h^{\perp}$. Since $I_h \subset H_h^{\perp}$, it follows that $H_h' = I_h^{\perp} \supset H_h$, whence $H_h \leqslant H_h'$. Since the elements of I_h are independent, we have

$$\mathbb{P}_H(H_h') = |I_h|^{-1} = 2^{-d} \leqslant 8L\alpha.$$

Since $h \in S_1$ we also have

$$\mathbb{P}_H(H_h) = \exp\Big(-O\big((c^{-1} + K^{1/2}c^{-1/2}) \log c^{-1}\big)\Big)\alpha,$$

whence

$$|H_h' : H_h| \leqslant L \exp\Big(O\big((c^{-1} + K^{1/2}c^{-1/2}) \log c^{-1}\big)\Big),$$

and it follows that

$$\mathbb{P}_{H_h'}(A_h') \geqslant L^{-1} \exp\Big(-O\big((c^{-1} + K^{1/2}c^{-1/2}) \log c^{-1}\big)\Big).$$

Thus there is some δ with $\delta|H|$ an integer and

$$\delta \geqslant L^{-1} \exp\Big(-O\big((c^{-1} + K^{1/2}c^{-1/2}) \log c^{-1}\big)\Big)$$

such that $\mathbb{P}_{H_h'}(A_h') \geqslant \delta$ for all $h \in S_1$; for each $h \in S_1$ let A_h'' be a subset of A_h' of density δ.

Each H_h' is defined by the set $I_h \subset \mathscr{L}'$ and there are at most $\binom{|\mathscr{L}'|}{d}$ such sets. Hence there is some space $H' \leqslant H$ such that $H_h' = H'$ for at least a proportion $\binom{|\mathscr{L}'|}{d}^{-1}$ of the elements of S_1; call this set S_2.

We now turn to estimating the density of S_2. First, by Parseval's theorem

$$|\mathscr{L}'|(\mathbb{P}_H(H_h)\alpha/4)^2 \leqslant \sum_{\gamma \in \widehat{H}} |\widehat{f_{\mathscr{A}}}(\gamma)|^2 = \|f_{\mathscr{A}}\|_{L^2(H)}^2 \leqslant K\alpha^2.$$

It follows from the lower bounds in $\mathbb{P}_H(H_h)$ that

$$|\mathscr{L}'| \leqslant \alpha^{-2} \exp\Big(O\big((c^{-1} + K^{1/2}c^{-1/2}) \log c^{-1}\big)\Big),$$

whence

$$\binom{|\mathscr{L}'|}{d} \leqslant \exp\Big(O\big((c^{-1} + K^{1/2}c^{-1/2}) \log^2 \alpha^{-1} \log c^{-1}\big)\Big).$$

This tells us that

$$\mathbb{P}_H(S_2) \geqslant \mathbb{P}_H(S_1)\big/\tbinom{|\mathcal{L}'|}{d} \geqslant \sigma \exp\Bigl(-O\bigl((c^{-1} + K^{1/2}c^{-1/2})\log^2\alpha^{-1}\log c^{-1}\bigr)\Bigr).$$

Finally, by averaging let $h_1 + H'$ be a coset of H' on which S_2 has at least the above density and define a new family $\mathcal{A}'''$ on H' as follows. For each $h' \in S_2 - h_1$, let $A'''_{h'} := A''_{h_1+h'}$; if $h' \in H' \setminus (S_2 - h_1)$ then let $A'''_{h'} := \emptyset$. By the definition of S_2 for each $h' \in S_2 - h_1$ $A''_{h_1+h'}$ is a subset of $H' = H'_{h_1+h'}$ of density δ. Thus by Proposition 7.1 we have

$$\begin{aligned}\Lambda(\mathcal{A}''') &= \exp\Bigl(-O\Bigl(\delta^{-1}\bigl(\log^2\alpha^{-1}(c^{-1}+K^{1/2}c^{-1/2})\log c^{-1} + \log\sigma^{-1}\bigr)\Bigr)\Bigr)\\ &= \exp\Bigl(-O\Bigl(L(\log^2\alpha^{-1}+\log\sigma^{-1})\exp\bigl(O\bigl((c^{-1}+K^{1/2}c^{-1/2})\log c^{-1}\bigr)\bigr)\Bigr)\Bigr).\end{aligned}$$

Finally it remains for us to check that $|H|^4\Lambda(\mathcal{A}) \geqslant |H'|^4\Lambda(\mathcal{A}''')$ from which the case follows; we proceed in the usual manner.

There are $|H'|^4\Lambda(\mathcal{A}')$ quadruples (a_0', a_1', y', h') with $a_0', a_1' \in A'''_{h'}$ and $y' \in A'''_{a_0'+a_1'-h'}$. Every such quadruple corresponds uniquely to a quadruple

$$(a_0, a_1, y, h) := (a_0' + x_{h_1+h'}, a_1' + x_{h_1+h'}, y' + x_{a_0'+a_1'-h'+h_1}, h_1 + h')$$

with $a_0, a_1 \in A_h$ and $y \in A_{a_0+a_1-h}$, whence $|H'|^4\Lambda(\mathcal{A}) \leqslant |H|^4\Lambda(\mathcal{A})$ and the result follows. □

Having concluded both cases it remains to note that certainly one of $\mathbb{P}_H(S_1)$ and $\mathbb{P}_H(S_0)$ is at least $\sigma/2$ and so at least one of the cases occurs. □

10. Families with small mean square density

In this section we use our previous work to establish the following lemma which is the main driver in the proof of Theorem 3.4 in the case when the density function has small mean square.

Lemma 10.1. *Suppose that H is a finite abelian group of exponent 2, $\mathcal{A} = (A_h)_{h\in H}$ is a family on H of density α, $\|f_{\mathcal{A}}\|^2_{L^2(H)} = K\alpha^2$ and $L \geqslant \max\{K, 2\}$ is a parameter. Then there is an absolute constant $C_{\mathcal{G}} > 0$ such that either*

$$\Lambda(\mathcal{A}) \geqslant \exp\bigl(-(1+\log^2\alpha^{-1})\exp(C_{\mathcal{G}}L^3\log^2 L)\bigr)$$

or there is a subgroup $H' \leqslant H$ of index 2 and a family $\mathcal{A}'$ on H' such that

$$\mathbb{P}_{H'}(\mathcal{A}') \geqslant \alpha + L\alpha^2/4K \quad \textit{and} \quad \Lambda(\mathcal{A}) \geqslant 2^{-4}\Lambda(\mathcal{A}').$$

Proof. Let $S_L := \{h \in H : f_{\mathcal{A}}(h) \geqslant 4K\alpha\}$ and $S_S := \{h \in H : f_{\mathcal{A}}(h) \leqslant \alpha/4\}$. Now,

$$\mathbb{E}_{h\in H} 1_{S_L}(h) f_{\mathcal{A}}(h) \leqslant \frac{1}{4K\alpha}\mathbb{E}_{h\in H} 1_{S_L}(h) f_{\mathcal{A}}(h)^2 \leqslant \frac{\alpha}{4},$$

and

$$\mathbb{E}_{h\in H} 1_{S_S}(h) f_{\mathcal{A}}(h) \leqslant \alpha/4$$

trivially, whence, putting $S := H \setminus (S_L \cup S_S)$, we have that

$$\mathbb{E}_{h\in H} 1_S(h) f_{\mathscr{A}}(h) \geqslant \alpha/2.$$

Let $S_i := \{h \in S : 2^{i-2}\alpha \leqslant f_{\mathscr{A}}(h) \leqslant 2^{i-1}\alpha\}$ and note that

$$\sum_{i\leqslant\lceil\log K\rceil+1} \mathbb{E}_{h\in H} 1_{S_i}(h)\, 2^{i-1}\alpha \geqslant \alpha/2,$$

and thus by averaging there is some $i \leqslant \lceil\log K\rceil + 1$ such that

$$\mathbb{E}_{h\in H} 1_{S_i}(h)\, 2^{i-1}\alpha \geqslant \alpha/2\big(\lceil\log K\rceil + 1\big).$$

As a byproduct note that $\mathbb{P}_H(S_i) = \Omega(1/K(1+\log K))$. We write $K_i = 2^{i-1}$, so that

$$K_i\alpha \geqslant f_{\mathscr{A}}(h) \geqslant K_i\alpha/2 \quad \text{for all } h \in S_i$$

and

$$\mathbb{E}_{h\in H} 1_{S_i}(h) f_{\mathscr{A}}(h) = \Omega\big(\alpha/(1+\log K)\big).$$

Now suppose that $\mathbb{E}_{h\in H} 1_{S_i}(h)|\widehat{1_{A_h}}(\gamma)|^2 \geqslant L\alpha^3$. Since $|\widehat{1_{A_h}}(\gamma)| \leqslant 4K\alpha$ if $h \in S$, we then conclude that

$$\mathbb{E}_{h\in H} 1_{S_i}(h)|\widehat{1_{A_h}}(\gamma)| \geqslant L\alpha^2/4K.$$

Applying Lemma 6.2 we find we are in the second case of Lemma 10.1. Similarly, by Lemma 6.3 we are done if $|\widehat{f_{\mathscr{A}}}(\gamma)| \geqslant L\alpha^2/4K$. Thus we may assume that

$$\sup_{\gamma\neq 0_{\widehat{H}}} \mathbb{E}_{h\in H} 1_{S_i}(h)|\widehat{1_{A_h}}(\gamma)|^2 \leqslant L\alpha^3, \tag{10-1}$$

$$\sup_{\gamma\neq 0_{\widehat{H}}} \mathbb{E}_{h\in H} 1_{S_i}(h)|\widehat{1_{A_h}}(\gamma)| \leqslant L\alpha^2/4K, \tag{10-2}$$

$$\sup_{\gamma\neq 0_{\widehat{H}}} |\widehat{f_{\mathscr{A}}}(\gamma)| \leqslant L\alpha^2/4K. \tag{10-3}$$

As usual, by the nonnegativity of the terms in $\Lambda(\mathscr{A})$, we have

$$\Lambda(\mathscr{A}) \geqslant \mathbb{E}_{h\in H} 1_{S_i}(h)\big\langle \tau_h(1_{A_h} * 1_{A_h}), f_{\mathscr{A}}\big\rangle_{L^2(H)}.$$

We apply Plancherel's theorem to the inner products on the right to get

$$\big\langle \tau_h(1_{A_h} * 1_{A_h}), f_{\mathscr{A}}\big\rangle_{L^2(H)} = \sum_{\gamma\in\widehat{H}} |\widehat{1_{A_h}}(\gamma)|^2 \widehat{f_{\mathscr{A}}}(\gamma)\gamma(h).$$

Separating out the trivial mode and applying the triangle inequality then tells us that

$$\big\langle \tau_h(1_{A_h} * 1_{A_h}), f_{\mathscr{A}}\big\rangle_{L^2(H)} \geqslant f_{\mathscr{A}}(h)^2\alpha - \sum_{\gamma\neq 0_{\widehat{H}}} |\widehat{1_{A_h}}(\gamma)|^2 |\widehat{f_{\mathscr{A}}}(\gamma)|.$$

Thus

$$\sum_{\gamma\neq 0_{\widehat{H}}} \mathbb{E}_{h\in H} 1_{S_i}(h)|\widehat{1_{A_h}}(\gamma)|^2 |\widehat{f_{\mathscr{A}}}(\gamma)| \geqslant \alpha\mathbb{E}_{h\in H} 1_{S_i}(h) f_{\mathscr{A}}(h)^2 - \Lambda(\mathscr{A}).$$

It follows that either

$$\Lambda(\mathscr{A}) \geqslant \alpha\mathbb{E}_{h\in H} 1_{S_i}(h) f_{\mathscr{A}}(h)^2/2 = \Omega\big(\alpha^3/(1+\log K)\big),$$

and we are done or

$$\sum_{\gamma \neq 0_{\widehat{H}}} \mathbb{E}_{h\in H} 1_{S_i}(h)|\widehat{1_{A_h}}(\gamma)|^2|\widehat{f_{\mathscr{A}}}(\gamma)| \geqslant \alpha \mathbb{E}_{h\in H} 1_{S_i}(h) f_{\mathscr{A}}(h)^2/2, \tag{10-4}$$

which we now assume. Let

$$\mathscr{L} := \left\{ \gamma \in \widehat{H} : \mathbb{E}_{h\in H} 1_{S_i}(h)|\widehat{1_{A_h}}(\gamma)|^2 \geqslant \frac{(\mathbb{E}_{h\in H} 1_{S_i}(h) f_{\mathscr{A}}(h)^2)^2}{2^4 K \mathbb{E}_{h\in H} 1_{S_i}(h) f_{\mathscr{A}}(h)} \right\}.$$

We shall now show that

$$\sum_{\gamma \notin \mathscr{L}} \mathbb{E}_{h\in H} 1_{S_i}(h)|\widehat{1_{A_h}}(\gamma)|^2|\widehat{f_{\mathscr{A}}}(\gamma)| \leqslant \alpha \mathbb{E}_{h\in H} 1_{S_i}(h) f_{\mathscr{A}}(h)^2/4. \tag{10-5}$$

We apply the triangle inequality to the left-hand side after swapping the order of summation to get that it is at most

$$\sup_{\gamma \notin \mathscr{L}} \big(\mathbb{E}_{h\in H} 1_{S_i}(h)|\widehat{1_{A_h}}(\gamma)|^2\big)^{1/2} \sum_{\gamma \in \widehat{H}} \big(\mathbb{E}_{h\in H} 1_{S_i}(h)|\widehat{1_{A_h}}(\gamma)|^2\big)^{1/2}|\widehat{f_{\mathscr{A}}}(\gamma)|.$$

Now apply the Cauchy–Schwarz inequality to this to see that the sum is at most

$$\bigg(\sum_{\gamma \in \widehat{H}} \mathbb{E}_{h\in H} 1_{S_i}(h)|\widehat{1_{A_h}}(\gamma)|^2\bigg)^{1/2} \bigg(\sum_{\gamma \in \widehat{H}} |\widehat{f_{\mathscr{A}}}(\gamma)|^2\bigg)^{1/2} = \sqrt{\mathbb{E}_{h\in H} 1_{S_i}(h) f_{\mathscr{A}}(h) K \alpha^2}$$

by Parseval's theorem, after interchanging the order of summation again. The bound (10-5) now follows from the definition of $\mathscr{L}$. Combining this with (10-4) we see that

$$\sum_{0_{\widehat{H}} \neq \gamma \in \mathscr{L}} \mathbb{E}_{h\in H} 1_{S_i}(h)\big|\widehat{1_{A_h}}(\gamma)\big|^2 \big|\widehat{f_{\mathscr{A}}}(\gamma)\big| \geqslant \alpha \mathbb{E}_{h\in H} 1_{S_i}(h) f_{\mathscr{A}}(h)^2/4 = \Omega\big(K_i \alpha^3/(1+\log K)\big).$$

Write

$$\mathscr{L}_j := \big\{\gamma \in \widehat{H} : 2^{-j} L\alpha^3 \geqslant \mathbb{E}_{h\in H} 1_{S_i}(h)|\widehat{1_{A_h}}(\gamma)|^2 \geqslant 2^{-(j+1)} L\alpha^3\big\},$$

and note that by (10-1) we have $\mathscr{L} \setminus \{0_{\widehat{H}}\} = \bigcup_{j=0}^{j_0} \mathscr{L}_j$, where j_0 is the smallest integer such that

$$2^{-(j_0+1)} L\alpha^3 \leqslant \big(\mathbb{E}_{h\in H} 1_{S_i}(h) f_{\mathscr{A}}(h)^2\big)^2 / 2^4 K \mathbb{E}_{h\in H} 1_{S_i}(h) f_{\mathscr{A}}(h);$$

crucially,

$$j_0 = O(\log L) \quad \text{and} \quad 2^{-(j_0+1)} L\alpha^3 = \Omega\big(K_i^2 \alpha^3 / K(1+\log K)\big).$$

It follows by averaging (and since $L \geqslant \max\{2, K\}$) that there is some $j \leqslant j_0$ such that

$$\sum_{0_{\widehat{H}} \neq \gamma \in \mathscr{L}_j} \mathbb{E}_{h\in H} 1_{S_i}(h)|\widehat{1_{A_h}}(\gamma)|^2|\widehat{f_{\mathscr{A}}}(\gamma)| = \Omega\big(K_i \alpha^3/(1+\log K)\log L\big).$$

Inserting (10-3) and dividing gives that

$$\sum_{0_{\widehat{H}} \neq \gamma \in \mathscr{L}_j} \mathbb{E}_{h\in H} 1_{S_i}(h)|\widehat{1_{A_h}}(\gamma)|^2 = \Omega\big(K_i K \alpha/(1+\log K) L \log L\big).$$

Now, the usual convexity of L^p-norms tells us that

$$\mathbb{E}_{h\in H}1_{S_i}(h)|\widehat{1_{A_h}}(\gamma)|^2 \leqslant \big(\mathbb{E}_{h\in H}1_{S_i}(h)|\widehat{1_{A_h}}(\gamma)|\big)^{2/3}\big(\mathbb{E}_{h\in H}1_{S_i}(h)|\widehat{1_{A_h}}(\gamma)|^4\big)^{1/3}.$$

Thus, by (10-2) we have

$$\big(\mathbb{E}_{h\in H}1_{S_i}(h)|\widehat{1_{A_h}}(\gamma)|^2\big)^3 \leqslant \frac{L^2\alpha^4}{2^4K^2}\mathbb{E}_{h\in H}1_{S_i}(h)|\widehat{1_{A_h}}(\gamma)|^4.$$

Dividing out and summing over $\mathscr{L}_j$, using the fact that it is a dyadic range, tells us that

$$\sum_{\gamma\in\widehat{H}}\mathbb{E}_{h\in H}1_{S_i}(h)|\widehat{1_{A_h}}(\gamma)|^4 = \Omega\big(\alpha^3K_i^5K/(1+\log K)^3L^3\log L\big).$$

Thus Parseval's theorem reveals that

$$\mathbb{E}_{h\in H}1_{S_i}(h)\|1_{A_h}*1_{A_h}\|_{L^2(H)}^2 = \Omega\big(\alpha^3K_i^5K/(1+\log K)^3L^3\log L\big).$$

Finally, let

$$S_i' := \big\{h\in S_i : \|1_{A_h}*1_{A_h}\|_{L^2(H)}^2 \geqslant \mathbb{E}_{h\in H}1_{S_i}(h)\|1_{A_h}*1_{A_h}\|_{L^2(H)}^2/2\big\}$$

and note that if $h\in S_i'$ then $f_{\mathscr{A}}(h)\leqslant K_i\alpha$, whence

$$\|1_{A_h}*1_{A_h}\|_{L^2(H)}^2 = \Omega\big(\alpha^3K_i^2K/(1+\log K)^3L^3\log L\big).$$

Furthermore

$$\mathbb{E}_{h\in H}1_{S_i'}(h)\|1_{A_h}*1_{A_h}\|_{L^2(H)}^2 \geqslant \Omega\big(\alpha^3K_i^5K/(1+\log K)^3L^3\log L\big),$$

whence $\mathbb{P}_H(S) = \Omega(L^4)$. We now apply Lemma 9.1 to see that

$$\Lambda(\mathscr{A}) \geqslant \exp\Big(-(1+\log^2\alpha^{-1})\,\exp\big(O\big(K^{-1}(1+\log K)^3L^3\log L\big)\big)\Big).$$

However, $K^{-1}(1+\log K)^3 = O(1)$, whence we get the result. □

It may seem bizarre to have thrown away the extra strength of the $K^{-1}(1+\log K)^3$ term at the very end of this proof. However, in applications we shall have a dichotomy between the case when K is large and when K is small. In the latter we shall not, in fact, be able to guarantee that K is much bigger than 1 whence the above estimate of $K^{-1}(1+\log K)^3 = O(1)$ is tight.

11. Proof of Theorem 3.4

As will have become clear the proof of Theorem 3.4 is iterative and is driven by Lemma 10.1 and Corollary 7.3.

Proof of Theorem 3.4. Let $H_0 := \operatorname{im} 2$ and $\mathscr{A}_0$ be the family corresponding to the set A, which has density $\alpha_0 = \alpha$. We shall define a sequence of families $(\mathscr{A}_i)_i$ on subgroups $(H_i)_i$ with density α_i and the properties:

$$\Lambda(\mathscr{A}_{i+1}) \leqslant 2^{-4}\Lambda(\mathscr{A}_i) \leqslant 2^{-4i}\Lambda(A), \qquad \alpha_{i+1} \geqslant \alpha_i\big(1+\Omega(\alpha_i\log^{1/6}\alpha_i^{-1}\log\log^{-5/3}\alpha_i^{-1})\big).$$

It is useful to define auxiliary variables K_i and L_i such that

$$C_{\mathcal{G}} L_i^3 \log^2 L_i = \log \alpha_i^{-1}/2 \quad \text{and} \quad K_i := \alpha_i^{-2} \| f_{\mathcal{A}_i} \|_{L^2(H_i)}^2.$$

Suppose that we are at stage i of the iteration; we consider two cases:

(i) If $L_i \leqslant 2 + K_i^2/(1+\log K_i)^2$ then apply Corollary 7.3 and terminate the iteration with

$$\Lambda(\mathcal{A}_i) = \exp\bigl(-O(\alpha_i^{-1} K_i^{-1} \log^2 K_i)\bigr) = \exp\bigl(-O(\alpha^{-1} \log^{-1/6} \alpha^{-1} \log\log^{5/3} \alpha^{-1})\bigr).$$

(ii) If $L_i > 2 + K_i^2/(1+\log K_i)^2$ then apply Lemma 10.1 with parameter L_i. If we have the first conclusion of the lemma then

$$\Lambda(\mathcal{A}_i) \geqslant \exp\bigl(-(1+\log \alpha_i^{-1})^2 \exp(C_{\mathcal{G}} L_i^3 \log^2 L_i)\bigr).$$

In view of the definition of L_i and the fact that $\alpha_i \geqslant \alpha$ we conclude that $\exp(C_{\mathcal{G}} L_i^3 \log^2 L_i) \leqslant \alpha^{-1/2}$, whence we certainly have

$$\Lambda(\mathcal{A}_i) = \exp\bigl(-O(\alpha^{-1} \log^{-1/6} \alpha^{-1} \log\log^{5/3} \alpha^{-1})\bigr)$$

again. The other conclusion of Lemma 10.1 tells us that we have a new subgroup $H_{i+1} \leqslant H_i$, and a family $\mathcal{A}_{i+1}$ on H_{i+1} with

$$\alpha_{i+1} \geqslant \alpha_i(1 + (L_i/4K_i)\alpha^{-1}) \quad \text{and} \quad \Lambda(\mathcal{A}_{i+1}) \geqslant 2^{-4} \Lambda(\mathcal{A}_i);$$

this has the desired property for the iteration.

In view of the lower bound on α_i we see that the density doubles in

$$F(\alpha) = O(\alpha^{-1} \log^{-1/6} \alpha^{-1} \log\log^{5/3} \alpha^{-1})$$

steps, whence the iteration must terminate in at most $F(\alpha) + F(2\alpha) + F(2^2\alpha) + \cdots$ steps. Of course $F(2\alpha') \leqslant F(\alpha')/\sqrt{2}$ whenever $\alpha' \in (0, c_0]$ for some absolute constant c_0. Thus, on summing the geometric progression we see that the iteration terminates in $O(F(\alpha))$ steps. It follows that at the time of termination we have

$$\Lambda(A) \geqslant \exp\bigl(-O(\alpha^{-1} \log^{-1/6} \alpha^{-1} \log\log^{5/3} \alpha^{-1})\bigr) \Lambda(\mathcal{A}_i),$$

and we get the result. □

12. Concluding remarks

No doubt some improvement could be squeezed out of our arguments by more judicious averaging, but there is a natural limit placed on the method by Corollary 8.2, and it seems that to move the 1/6 in Theorem 3.4 past 1 would require a new idea. This, however, is a little frustrating for the following reason.

The well-known Erdős–Turán conjecture is essentially equivalent to asking for Roth's theorem in $\mathbb{Z}/N\mathbb{Z}$ for any set of density $\delta(N)$, where $\delta(N)$ is a function with $\sum_N N^{-1}\delta(N) = \infty$. In particular, $\delta(N) = 1/\log N \log\log N \log\log\log N$ satisfies this hypothesis and so to have the analogue of the Erdős–Turán conjecture in $\mathbb{Z}_4^n$ we would need to push the constant 1/6 past 1.

In light of the heuristic in Section 4 one might reasonably conjecture the following much stronger result.

Conjecture 12.1. *Suppose that $G = \mathbb{Z}_4^n$ and $A \subset G$ contains no proper three-term arithmetic progressions. Then $|A| = O\big(|G|/\log^{3/2}|G|\big)$.*

Of course much more may be true. We were able to find the following lower bound; as with $\mathbb{Z}_3^n$, where the best lower bound is due to Edel [2004] (see also [Lin and Wolf 2009]), its density is of power shape.

Proposition 12.2. *Suppose that $G = \mathbb{Z}_4^n$. Then there is a set $A \subset G$ with no proper three-term arithmetic progressions and $|A| = \Omega(|G|^{2/3})$.*

Proof. The set

$$A_0 = \left\{ \begin{array}{l} (0,0,0),\ (0,0,1),\ (0,1,0),\ (0,1,2),\ (0,2,1),\ (0,2,2),\ (1,0,0),\ (1,0,2), \\ (1,2,0),\ (1,2,2),\ (2,0,1),\ (2,0,2),\ (2,1,0),\ (2,1,2),\ (2,2,0),\ (2,2,1) \end{array} \right\}$$

in $\mathbb{Z}_4^3$ has size 16 and contains no proper three-term arithmetic progressions. The result now follows on noting that the product of two sets not containing any proper three-term arithmetic progressions does, itself, not contain any proper three-term arithmetic progressions:

Suppose that B and C are such sets and $(x_0, x_1), (y_0, y_1), (z_0, z_1) \in B \times C$ have $x + y = 2z$. Then $x_i + y_i = 2z_i$ for $i \in \{0, 1\}$. However since B and C do not contain any proper progressions we have $x_i = y_i$ for all $i \in \{0, 1\}$ whence $x = y$ and so the progression is not proper. □

We are unaware of any serious search for better choices of A_0, though they may exist. Indeed, recently Elsholtz observed that a more general construction designed for Moser's cube problem can be used.

Moser asked for large subsets of $\{0, 1, 2\}^n$ not containing three points on a line; Komlós and Chvátal [Chvátal 1972] note that the sets

$$S_n := \big\{x \in \{0, 1, 2\}^n : x_i = 1 \text{ for } \lfloor n/3 \rfloor \text{ values of } i \in [n]\big\}$$

have size $\Omega(3^n/\sqrt{n})$ by Stirling's formula and satisfy Moser's requirement. Our set A_0 is equal to S_3. Embedding S_n in $\mathbb{Z}_4^n$ in the obvious way it may be checked that the lack of lines in S_n yields a set containing no proper three-term arithmetic progressions and hence the following theorem.

Theorem 12.3 [Elsholtz 2008, Theorem 3]. *Suppose that $G = \mathbb{Z}_4^n$. Then there is a set $A \subset G$ with no proper three-term arithmetic progressions and*

$$|A| = \Omega\big(|G|^{\log 3/\log 4}/\sqrt{\log|G|}\big).$$

The reader may wish to know that $\log 3/\log 4 = 0.792\ldots$. The details along with some other results and generalisations are supplied in Elsholtz's paper.

Acknowledgments

The author thanks Ernie Croot for a number of very useful conversations, Christian Elsholtz for supplying the preprint [Elsholtz 2008], Olof Sisask for writing a program to find the example in Proposition 12.2, Terry Tao for useful comments and two anonymous referees for useful comments and careful reading.

References

[Balog and Szemerédi 1994] A. Balog and E. Szemerédi, "A statistical theorem of set addition", *Combinatorica* **14**:3 (1994), 263–268. MR 95m:11019 Zbl 0812.11017

[Bourgain 1999] J. Bourgain, "On triples in arithmetic progression", *Geom. Funct. Anal.* **9**:5 (1999), 968–984. MR 2001h:11132 Zbl 0959.11004

[Bourgain 2008] J. Bourgain, "Roth's theorem on progressions revisited", *J. Anal. Math.* **104** (2008), 155–192. MR 2009g:11011 Zbl 1155.11011

[Brown and Buhler 1984] T. C. Brown and J. P. Buhler, "Lines imply spaces in density Ramsey theory", *J. Combin. Theory Ser. A* **36**:2 (1984), 214–220. MR 85d:05068 Zbl 0532.05002

[Chvátal 1972] V. Chvátal, "Remarks on a problem of Moser", *Canad. Math. Bull.* **15** (1972), 19–21. MR 47 #1642 Zbl 0232.05002

[Croot 2007] E. S. Croot, "On the decay of the Fourier transform and three term arithmetic progressions", *Online J. Anal. Comb.* **2** (2007), Art. 6, 10 pp. MR 2008h:11011 Zbl 1145.11015

[Croot 2008] E. S. Croot, "Subsets of $\mathbb{F}_p^n$ without three term arithmetic progressions have several large Fourier coefficients.", preprint, 2008.

[Croot and Lev 2007] E. S. Croot and V. F. Lev, "Open problems in additive combinatorics", pp. 207–233 in *Additive combinatorics*, edited by A. Granville et al., CRM Proc. Lecture Notes **43**, Amer. Math. Soc., Providence, RI, 2007. MR 2009d:11038 Zbl 05219607

[Edel 2004] Y. Edel, "Extensions of generalized product caps", *Des. Codes Cryptogr.* **31**:1 (2004), 5–14. MR 2004m:51021 Zbl 1057.51005

[Edel et al. 2007] Y. Edel, C. Elsholtz, A. Geroldinger, S. Kubertin, and L. Rackham, "Zero-sum problems in finite abelian groups and affine caps", *Q. J. Math.* **58**:2 (2007), 159–186. MR 2008g:11017 Zbl 05190260

[Elsholtz 2008] C. Elsholtz, "Lower bounds for Roth's theorem in $\mathbb{Z}_4^n$", preprint, 2008.

[Frankl et al. 1987] P. Frankl, R. L. Graham, and V. Rödl, "On subsets of abelian groups with no 3-term arithmetic progression", *J. Combin. Theory Ser. A* **45**:1 (1987), 157–161. MR 88f:05018 Zbl 0613.10043

[Freĭman 1973] G. A. Freĭman, *Foundations of a structural theory of set addition*, Translations of Mathematical Monographs **37**, American Mathematical Society, Providence, R. I., 1973. MR 50 #12944 Zbl 0271.10044

[Gowers 1998] W. T. Gowers, "A new proof of Szemerédi's theorem for arithmetic progressions of length four", *Geom. Funct. Anal.* **8**:3 (1998), 529–551. MR 2000d:11019 Zbl 0907.11005

[Green 2005] B. Green, "Finite field models in additive combinatorics", pp. 1–27 in *Surveys in combinatorics 2005*, edited by B. S. Webb, London Math. Soc. Lecture Note Ser. **327**, Cambridge Univ. Press, 2005. MR 2006j:11030 Zbl 1155.11306

[Green and Tao 2009] B. Green and T. Tao, "A note on the Freiman and Balog–Szemerédi–Gowers theorems in finite fields", *J. Aust. Math. Soc.* **86**:1 (2009), 61–74. MR 2495998 Zbl 05550054

[Heath-Brown 1987] D. R. Heath-Brown, "Integer sets containing no arithmetic progressions", *J. London Math. Soc.* (2) **35**:3 (1987), 385–394. MR 88g:11005 Zbl 0589.10062

[Lev 2004] V. F. Lev, "Progression-free sets in finite abelian groups", *J. Number Theory* **104**:1 (2004), 162–169. MR 2004k:11023 Zbl 1043.11022

[Lin and Wolf 2009] Y. Lin and J. Wolf, "Subsets of $\mathbb{F}_q^n$ containing no k-term progressions", preprint, 2009.

[Meshulam 1995] R. Meshulam, "On subsets of finite abelian groups with no 3-term arithmetic progressions", *J. Combin. Theory Ser. A* **71**:1 (1995), 168–172. MR 96g:20033 Zbl 0832.11006

[Roth 1952] K. Roth, "Sur quelques ensembles d'entiers", *C. R. Acad. Sci. Paris* **234** (1952), 388–390. MR 13,724d Zbl 0046.04302

[Roth 1953] K. F. Roth, "On certain sets of integers", *J. London Math. Soc.* **28** (1953), 104–109. MR 14,536g Zbl 0050.04002

[Rudin 1962] W. Rudin, *Fourier analysis on groups*, Wiley, New York, 1962. Reprinted 1990 in the Wiley Classics Library. MR 27 #2808 Zbl 0107.096

[Ruzsa 1999] I. Z. Ruzsa, "An analog of Freiman's theorem in groups", pp. 323–326 in *Structure theory of set addition*, edited by J.-M. Deshouillers et al., Astérisque **258**, Soc. Math. de France, Paris, 1999. MR 2000h:11111 Zbl 0946.11007

[Szemerédi 1990] E. Szemerédi, "Integer sets containing no arithmetic progressions", *Acta Math. Hungar.* **56**:1-2 (1990), 155–158. MR 92c:11100 Zbl 0721.11007

[Tao 2008] T. Tao, *Structure and randomness: pages from year one of a mathematical blog*, American Mathematical Society, Providence, RI, 2008. MR 2459552

[Tao and Vu 2006] T. Tao and V. Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics **105**, Cambridge University Press, 2006. MR 2008a:11002 Zbl 1127.11002

[Varnavides 1959] P. Varnavides, "On certain sets of positive density", *J. London Math. Soc.* **34** (1959), 358–360. MR 21 #5595 Zbl 0088.25702

[Yekhanin and Dumer 2004] S. Yekhanin and I. Dumer, "Long nonbinary codes exceeding the Gilbert–Varshamov bound for any fixed distance", *IEEE Trans. Inform. Theory* **50**:10 (2004), 2357–2362. MR 2097052

Received 14 Nov 2008. Revised 30 Mar 2009. Accepted 4 May 2009.

TOM SANDERS: t.sanders@dpmms.cam.ac.uk
Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge, CB3 0WA, United Kingdom
http://www.dpmms.cam.ac.uk/~tws22/

ANALYSIS AND PDE
Vol. 2, No. 2, 2009

THE HIGH EXPONENT LIMIT $p \to \infty$ FOR THE ONE-DIMENSIONAL NONLINEAR WAVE EQUATION

TERENCE TAO

We investigate the behaviour of solutions $\phi = \phi^{(p)}$ to the one-dimensional nonlinear wave equation $-\phi_{tt} + \phi_{xx} = -|\phi|^{p-1}\phi$ with initial data $\phi(0, x) = \phi_0(x)$, $\phi_t(0, x) = \phi_1(x)$, in the high exponent limit $p \to \infty$ (holding ϕ_0, ϕ_1 fixed). We show that if the initial data ϕ_0, ϕ_1 are smooth with ϕ_0 taking values in $(-1, 1)$ and obey a mild nondegeneracy condition, then ϕ converges locally uniformly to a piecewise limit $\phi^{(\infty)}$ taking values in the interval $[-1, 1]$, which can in principle be computed explicitly.

1. Introduction

Consider solutions $\phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ to the defocusing nonlinear wave equation

$$-\phi_{tt} + \phi_{xx} = |\phi|^{p-1}\phi \tag{1-1}$$

where $p > 1$ is a parameter. From standard energy methods (see, for example, [Sogge 1995]), relying in particular on the conserved energy

$$E(\phi)(t) = \int_{\mathbb{R}} \frac{1}{2}|\phi_t|^2 + \frac{1}{2}|\phi_x|^2 + \frac{1}{p+1}|\phi|^{p+1}\, dx \tag{1-2}$$

and on the Sobolev embedding $H^1_x(\mathbb{R}) \subset L^\infty_x(\mathbb{R})$, we know that given any initial data $\phi_0 \in H^1_x(\mathbb{R})$, $\phi_1 \in L^2_x(\mathbb{R})$, there exists a unique global energy class solution $\phi \in C^0_t H^1_x \cap C^1_t L^2_x(\mathbb{R} \times \mathbb{R})$ to (1-1) with initial data $\phi(0) = \phi_0$, $\phi_t(0) = \phi_1$. One has a similar theory for data that is only locally of finite energy, thanks to finite speed of propagation.

In this paper we investigate the asymptotic behaviour of this solution $\phi = \phi^{(p)}$ in the high exponent limit[1] $p \to \infty$, while keeping the initial data fixed. To avoid technicalities, let us suppose that ϕ_0, ϕ_1 are smooth and compactly supported, and that $|\phi_0(x)| < 1$ for all x. Formally, we expect $\phi^{(p)}$ to converge in some sense to some solution $\phi = \phi^{(\infty)}$ of the *infinitely nonlinear defocusing wave equation*

$$-\phi_{tt} + \phi_{xx} = |\phi|^{\infty}\phi \tag{1-3}$$

with initial data $\phi(0) = \phi_0$, $\phi_t(0) = \phi_1$.

MSC2000: 35L15.
Keywords: nonlinear wave equation.
The author is supported by a grant from the MacArthur Foundation, by NSF grant DMS-0649473, and by the NSF Waterman Award.

[1] We are indebted to Tristan Roy for posing this question.

Of course, (1-3) does not make rigorous sense. But, motivated by analogy with infinite barrier potentials[2], one might wish to interpret the infinite nonlinearity $|\phi|^\infty \phi$ as a "barrier nonlinearity" which is constraining ϕ to have magnitude at most 1, but otherwise has no effect. Intuitively, we thus expect the limiting wave $\phi^{(\infty)}$ to evolve like the linear wave equation until it reaches the threshold $\phi^{(\infty)} = +1$ or $\phi^{(\infty)} = -1$, at which point it should "reflect" off the nonlinear barrier[3]. The purpose of this paper is to make the above intuition rigorous, and to give a precise interpretation for Equation (1-3).

1.1. ***An ODE analogy.*** To get some further intuition as to this reflection phenomenon, let us first study (nonrigorously) the simpler ODE problem, in which we look at solutions $\phi = \phi^{(p)} : \mathbb{R} \to \mathbb{R}$ to the ODE

$$-\phi_{tt} = |\phi|^{p-1}\phi \tag{1-4}$$

with fixed initial data $\phi(0) = \phi_0$, $\phi_t(0) = \phi_1$ with $|\phi_0| \leqslant 1$, and with $p \to \infty$. From the conserved energy $\frac{1}{2}\phi_t^2 + \frac{1}{p+1}|\phi|^{p+1}$ (and recalling that $(p+1)^{1/(p+1)} = 1 + (\log p)/p + O\left(\frac{1}{p}\right)$) we quickly obtain the uniform bounds

$$|\phi_t(t)| = O(1); \quad |\phi(t)| \leqslant 1 + \frac{\log p}{p} + O\left(\frac{1}{p}\right) \tag{1-5}$$

for all p and all times t, where the implied constants in the $O()$ notation depend on ϕ_0, ϕ_1. Thus we already see a barrier effect preventing ϕ from going too far outside of the interval $[-1, 1]$. To investigate what happens near a time t_0 in which $\phi(t_0)$ is close to (say) $+1$, let us make the ansatz

$$\phi(t) = p^{1/(p-1)}\left(1 + \frac{1}{p}\psi\big(p(t-t_0)\big)\right).$$

Observe from (1-5) that $\phi(t)$ is positive for $|t - t_0| \leqslant c$ and some constant $c > 0$ depending only on ϕ_0, ϕ_1. Write $s := p(t - t_0)$. Some brief computation then shows that ψ solves the equation

$$\psi_{ss} = -\left(1 + \frac{1}{p}\psi\right)^p$$

for all $s \in [-cp, cp]$; also, by (1-5) we obtain an upper bound $\psi \leqslant O(1)$ (but no comparable lower bound), as well as the Lipschitz bound $|\psi_s| = O(1)$. In the asymptotic limit $p \to \infty$, we thus expect the rescaled solution $\psi = \psi^{(p)}$ to converge to a solution $\psi = \psi^{(\infty)}$ of the ODE

$$\psi_{ss} = -e^\psi.$$

[2]For instance, if one takes the solution $\phi = \phi^{(p)}$ to the linear wave equation $-\phi_{tt} + \phi_{xx} = p 1_{\mathbb{R}\setminus[-1,1]}(x)\phi$ with initial data smooth and supported on $[-1, 1]$, a simple compactness argument (or explicit computation) shows that ϕ converges (in, say, the uniform topology) to the solution to the free wave equation $-\phi_{tt} + \phi_{xx} = 0$ on $\mathbb{R} \times [-1, 1]$ with the reflective (Dirichlet) boundary conditions $\phi(t, \pm 1) = 0$.

[3]Since each of the equations (1-1) are Hamiltonian, it is reasonable to expect that (1-3) should also be "Hamiltonian" in some sense (although substituting $p = +\infty$ in (1-2) does not directly make sense), and so energy should be reflected rather than absorbed by the barrier.

It turns out that this ODE can be solved explicitly[4], and it is easy to verify that the general solution is

$$\psi(s) = \log \frac{2a^2}{\cosh^2(a(s-s_0))} \tag{1-6}$$

for any $s_0 \in \mathbb{R}$ and $a > 0$. These solutions asymptotically approach $-a|s-s_0| + \log 8a^2$ as $s \to \pm\infty$. Thus we see that if ψ is large and negative but with positive velocity, then the solution to this ODE will be approximately linear until ψ approaches the origin, where it will dwell for a bounded amount of time before reflecting back into the negative axis with the opposite velocity to its initial velocity. Undoing the rescaling, we thus expect the limit $\phi = \phi^{(\infty)}$ of the original ODE solutions $\phi^{(p)}$ to also behave linearly until reaching $\phi = +1$ or $\phi = -1$, at which point they should reflect with equal and opposite velocity, so that $\phi^{(\infty)}$ will eventually be a sawtooth function with range $[-1, 1]$ (except of course in the degenerate case $\phi_1 = 0$, $|\phi_0| < 1$, in which case $\phi^{(\infty)}$ should be constant). Because the ODE can be solved more or less explicitly using the conserved Hamiltonian, it is not difficult to formalise these heuristics rigorously; we leave this as an exercise to the interested reader. Note that the above analysis also suggests a more precise asymptotic for how reflections of $\phi^{(p)}$ should behave for large p, namely (assuming $s_0 = 0$ for simplicity)

$$\phi^{(p)}(t) \approx p^{1/(p-1)} \left(1 + \frac{1}{p} \log \frac{2a^2}{\cosh^2(ap(t-t_0))}\right)$$

or (after Taylor expansion)

$$\phi^{(p)}(t) \approx 1 + \frac{\log p}{p} + \frac{1}{p} \log \frac{2a^2}{\cosh^2(ap(t-t_0))}, \tag{1-7}$$

where a measures the speed of the reflection, and t_0 the time at which reflection occurs, and we are deliberately being vague as to what the symbol $\approx$ means.

Adapting the above ODE analysis to the PDE setting, we can now study the reflection behaviour of $\phi^{(p)}$ near the nonlinear barrier $\phi^{(p)} = 1$ at some point (t_0, x_0) in spacetime by introducing the ansatz

$$\phi(t,x) = p^{1/(p-1)} \Big(1 + \frac{1}{p} \psi(p(t-t_0), p(x-x_0))\Big).$$

where ψ can be computed to solve the equation

$$-\psi_{tt} + \psi_{xx} = -\Big(1 + \frac{1}{p}\psi\Big)^p$$

in the region where ϕ is near 1 (and is in particular nonnegative). In the limit $p \to \infty$, this formally converges to *Liouville's equation*

$$-\psi_{tt} + \psi_{xx} = e^{\psi}. \tag{1-8}$$

Remarkably, this nonlinear wave equation can also be solved explicitly [Liouville 1853], with explicit solution

$$\psi = \log \frac{-8f'(t+x)g'(t-x)}{(f(t+x)+g(t-x))^2} \tag{1-9}$$

[4] Alternatively, one can reach the desired qualitative conclusions by tracking the ODE along the energy surfaces $\frac{1}{2}\psi_s^2 + e^{\psi} =$ const in phase space.

for arbitrary smooth functions f, g for which the right side is well-defined[5]. Somewhat less "magically", one can approach the explicit solvability of this equation by introducing the null coordinates

$$u := t + x; \quad v := t - x \tag{1-10}$$

and their associated derivatives

$$\partial_u := \tfrac{1}{2}(\partial_t + \partial_x); \quad \partial_v := \tfrac{1}{2}(\partial_t - \partial_x) \tag{1-11}$$

and rewriting (1-8) as

$$\psi_{uv} = -\tfrac{1}{4}e^{\psi}, \tag{1-12}$$

and then noting the *pointwise* conservation laws

$$\partial_v(\tfrac{1}{2}\psi_u^2 - \psi_{uu}) = \partial_u(\tfrac{1}{2}\psi_v^2 - \psi_{vv}) = 0 \tag{1-13}$$

which can ultimately (with a certain amount of algebraic computation) be used to arrive at the solution (1-9); see [Tao $\geq$ 2009] for details. Using this explicit solution, one can eventually be led to the (heuristic) conclusion that the reflection profile $\psi^{(\infty)}$ should resemble a Lorentz-transformed version of (1-6), that is,

$$\psi(t, x) = \log \frac{2a^2}{\cosh^2(a[(t - t_0) - v(x - x_0)]/\sqrt{1 - v^2})} \tag{1-14}$$

for some $t_0, x_0 \in \mathbb{R}$, $a > 0$, and $-1 < v < 1$. Thus we expect ϕ to reflect along spacelike curves such as $(t - t_0) - v(x - x_0) = 0$ in order to stay confined to the interval $[-1, 1]$.

1.2. *Main result.* We now state the main result of our paper, which aims to make the above intuition precise.

Theorem 1.3 (Convergence as $p \to \infty$). *Let $\phi_0, \phi_1 : \mathbb{R} \to \mathbb{R}$ be functions obeying the following properties*:

(a) (*Regularity*) *ϕ_0, ϕ_1 are smooth.*

(b) (*Strict barrier condition*) *For all $x \in \mathbb{R}$, $|\phi_0(x)| < 1$.*

(c) (*Nondegeneracy*) *The sets $\{x : \frac{1}{2}(\phi_1 + \partial_x\phi_0)(x) = 0\}$ and $\{x : \frac{1}{2}(\phi_1 - \partial_x\phi_0)(x) = 0\}$ have only finitely many connected components in any compact interval*[6].

For each $p > 1$, let $\phi^{(p)} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be the unique global solution to (1-1) with initial data $\phi^{(p)}(0) = \phi_0$, $\phi_t^{(p)}(0) = \phi_1$. Then, as $p \to \infty$, $\phi^{(p)}$ converges uniformly on compact subsets of $\mathbb{R} \times \mathbb{R}$ to the unique function $\phi = \phi^{(\infty)} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ that obeys the following properties:

(i) (*Regularity, I*) *ϕ is locally Lipschitz continuous* (*and in particular is differentiable almost everywhere, by Radamacher's theorem*).

(ii) (*Regularity, II*) *For each $u \in \mathbb{R}$ and $v \in \mathbb{R}$, the functions $t \mapsto \phi(t, u - t)$ and $t \mapsto \phi(t, t - v)$ are piecewise smooth* (*with finitely many pieces on each compact interval*).

[5] For instance, the ODE solutions (1-6) can be recovered by setting $f(u) := e^{a(u-t_0)}$ and $g(v) := e^{-a(v-t_0)}$.

[6] This condition is automatic if ϕ_0, ϕ_1 are real analytic, since the zeroes of nontrivial real analytic functions cannot accumulate.

(iii) (*Initial data*) *On a neighbourhood of the initial surface* $\{(0, x) : x \in \mathbb{R}\}$, ϕ *agrees with the linear solution*

$$\phi^{(\mathrm{lin})}(t, x) := \frac{1}{2}(\phi_0(x+t) + \phi_0(x-t)) + \frac{1}{2}\int_{x-t}^{x+t} \phi_1(y)\, dy, \tag{1-15}$$

the free wave equation with initial data ϕ_0, ϕ_1.

(iv) (*Barrier condition*) $|\phi(t, x)| \leqslant 1$ *for all* t, x.

(v) (*Defect measure*) *We have*

$$-\phi_{tt} + \phi_{xx} = \mu_+ - \mu_- \tag{1-16}$$

in the sense of distributions, where μ_+, μ_- *are locally finite nonnegative measures supported on the sets* $\{(t, x) : \phi(t, x) = +1\}$, $\{(t, x) : \phi(t, x) = -1\}$, *respectively.*

(vi) (*Null energy reflection*) *For almost every* (t, x), *we have*[7]

$$|\phi_u(t, x)| = |\phi_u^{(\mathrm{lin})}(t, x)| \tag{1-17}$$

and

$$|\phi_v(t, x)| = |\phi_v^{(\mathrm{lin})}(t, x)|. \tag{1-18}$$

In particular, $|\phi_u|$ *is almost everywhere equal to a function of* u *only, and similarly for* $|\phi_v|$.

Remark 1.4. The existence and uniqueness of ϕ obeying the above properties is not obvious, but is part of the theorem. The conditions (i)–(vi) are thus the rigorous substitute for the nonrigorous Equation (1-3); they superficially resemble a *viscosity solution* or *kinetic formulation* of (1-3) (see, for example, [Perthame 2002]), and it would be interesting to see if there is any rigorous connection here to the kinetic theory of conservation laws.

Remark 1.5. The hypotheses (a), (b), (c) on the initial data ϕ_0, ϕ_1 are somewhat stronger than what is likely to be needed for the theorem to hold; in particular, one should be able to relax the strict barrier condition (b) to $|\phi_0(x)| \leqslant 1$, and also omit the nondegeneracy condition (c), although the conclusions (ii), (iii) the limit ϕ would have to be modified in this case; one also expects to be able to relax the smoothness assumption (a), perhaps all the way to the energy class or possibly even the bounded variation class. We will not pursue these matters here.

1.6. *An example.* To illustrate the reflection in action, let us restrict attention to the triangular region

$$\Delta := \big\{(t, x) : t \geqslant 0;\ |t-x|, |t+x| \leqslant 10\big\}$$

and consider the initial data ϕ_0, ϕ_1 associated to the linear solution

$$\phi^{(\mathrm{lin})}(t, x) = 1 - \delta((t-2)^2 + x^2 - 1),$$

where $\delta > 0$ is a small constant (for example, $\delta = 10^{-3}$ is safe). Observe that $\phi^{(\mathrm{lin})}$ lies between -1 and 1 for most of Δ, but exceeds 1 in the disk $\{(t, x) : (t-2)^2 + x^2 < 1\}$. Thus we expect ϕ to follow $\phi^{(\mathrm{lin})}$ until it encounters this disk, at which point it should reflect.

[7]Of course, we can compute the derivatives of $\phi^{(\mathrm{lin})}$ explicitly from (1-15) in terms of the initial data as $\phi_u^{(\mathrm{lin})}(t, x) = \frac{1}{2}(\phi_1(0, x+t) + \partial_x\phi_0(0, x+t))$ and $\phi_v^{(\mathrm{lin})}(t, x) = \frac{1}{2}(\phi_1(0, x-t) - \partial_x\phi_0(0, x-t))$.

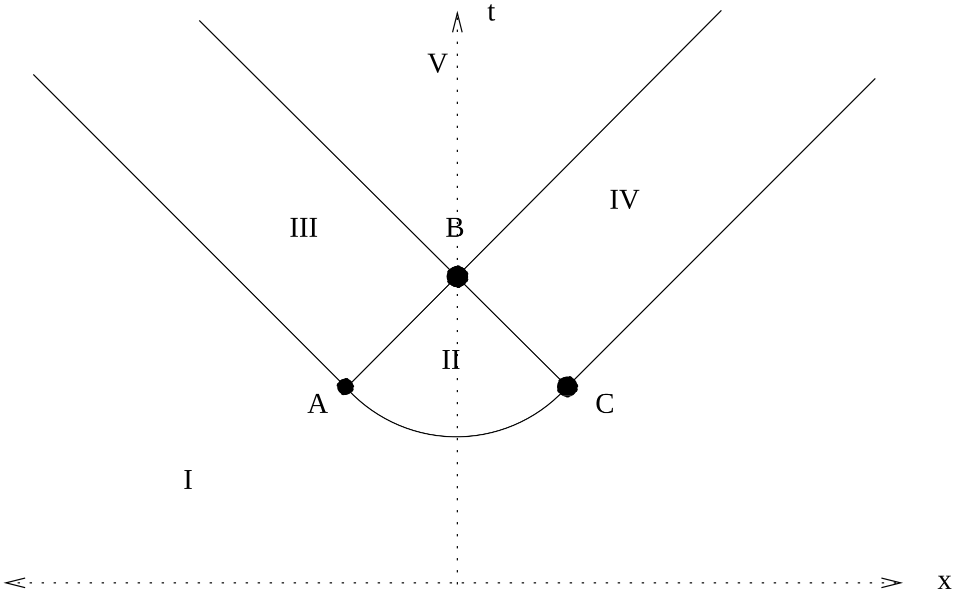

Figure 1. A subdivision of the triangular region Δ (the diagonal boundaries of Δ are beyond the scale of the figure). We have $A = \left(2 - \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$, $B = (2, 0)$, and $C = \left(2 - \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$. The circular arc from A to C is part of the circle $\{(t, x) : (t-2)^2 + x^2 = 1\}$. The rays bounding regions III and IV are all null rays.

The actual solution ϕ can be described using Figure 1. In region I, ϕ is equal to the linear solution $\phi^{(\mathrm{lin})}$. But $\phi^{(\mathrm{lin})}$ exceeds 1 once one passes the circular arc joining A and C, and so a reflection must occur in region II; indeed, one has

$$\phi(t, x) = 2 - \phi^{(\mathrm{lin})}(t, x) = 1 + \delta((t-2)^2 + x^2 - 1)$$

in this region.

Once the solution passes A and C, though, it turns out that the downward velocity of the reflected wave is now sufficient to drag ϕ off of the singular set[8] $\{\phi = +1\}$ (which, in this example, is the circular arc connecting A and C). Indeed, in region III, we have

$$\phi(t, x) = 1 + \delta(2(t-2)x - 1)$$

(note that this is the unique solution to the free wave equation that matches up with ϕ on regions I, II) and similarly on region IV we have

$$\phi(t, x) = 1 + \delta(-2(t-2)x - 1).$$

[8] An inspection of (1-14) suggests that the singular set must remain spacelike, thus the timelike portions of the set $\{\phi^{(\mathrm{lin})} = 1\}$ (which, in this case, are the left and right arcs of the circle $\{(t, x) : (t-2)^2 + x^2 = 1\}$) are not used as a reflective set for ϕ.

Finally, on region V we have another solution to the free wave equation, which is now on a downward trajectory away from $\phi = +1$:

$$\phi(t,x) = 1 - \delta((t-2)^2 + x^2 + 1).$$

If one were to continue the evolution of ϕ forward in time beyond Δ (extending the initial data ϕ_0, ϕ_1 suitably), the solution would eventually hit the $\phi = -1$ barrier and reflect again, picking up further singularities propagating in null directions similar to those pictured here. Thus, while the solution remains piecewise smooth for all time, we expect the number of singularities to increase as time progresses, due to the increasing number of reflections taking place.

It is a routine matter to verify that the solution presented here verifies the properties (i)–(vi) on Δ (if δ is sufficiently small), and so is necessarily the limiting solution ϕ, thanks to Theorem 1.3 (and the uniqueness theory in Section 3 below). We omit the details.

Remark 1.7. The circular arc between A and C supports a component of the defect measure μ_+, which can be computed explicitly from the above formulae. The defect measure can also be computed by integrating (1-16) and observing that

$$\phi\langle u,v\rangle - \phi\langle u-a,v\rangle - \phi\langle u,v-b\rangle + \phi\langle u-a,v-b\rangle = -\mu_+\big(\{(u',v') : u-a \leqslant u' \leqslant u;\, v-b \leqslant v' \leqslant v\}\big)$$

whenever $\langle u,v\rangle$, $\langle u-a,v\rangle$, $\langle u,v-b\rangle$, $\langle u-a,v-b\rangle$ are the corners of a small parallelogram intersecting this arc. Sending $b \to 0$, say, we observe that the left side is asymptotic to $b(\phi_v\langle u,v\rangle - \phi_v\langle u-a,v\rangle)$. Since ϕ_v reflects in sign across the arc, we can simplify this as $b|\phi_v^{(\mathrm{lin})}\langle u,v\rangle|$. This allows us to describe μ_+ explicitly in terms of the conserved quantity $|\phi_v^{(\mathrm{lin})}|$ and the slope of the arc; we omit the details.

Remark 1.8. The above example shows that the barrier set $|\phi| = 1$ has some overlap initially with the set $|\phi^{(\mathrm{lin})}| = 1$, but the situation becomes more complicated after multiple "reflections" off of the two barriers $\phi = +1$ and $\phi = -1$, and the author does not know of a clean way to describe this set for large times t, although as the above example suggests, these sets should be computable for any given choice of t and any given initial data.

1.9. ***Proof strategy.*** We shall shortly discuss the proof of Theorem 1.3, but let us first pause to discuss two techniques that initially look promising for solving this problem, but end up being problematic for a number of reasons.

Each of the nonlinear wave equations (1-1) enjoy a conserved stress-energy tensor $T^{(p)}_{\alpha\beta}$, and it is tempting to try to show that this stress-energy tensor converges to a limit $T^{(\infty)}_{\alpha\beta}$. However, the author found it difficult to relate this limit tensor to the limit solution $\phi^{(\infty)}$. The key technical difficulty was that while it was not difficult to ensure that derivatives $\phi^{(p)}_u, \phi^{(p)}_v$ of $\phi^{(p)}$ converged in a weak sense to the derivatives $\phi^{(\infty)}_u, \phi^{(\infty)}_v$ of a limit $\phi^{(\infty)}$, this did not imply that the *magnitudes* $|\phi^{(p)}_u|, |\phi^{(p)}_v|$ of the derivatives converged (weakly) to the expected limit of $|\phi^{(\infty)}_u|, |\phi^{(\infty)}_v|$, due to the possibility of increasing oscillations in the sign of $\phi^{(p)}_u$ or $\phi^{(p)}_v$ in the limit $p \to \infty$ which could cause some loss of mass in the limit. Because of this, much of the argument is instead focused on controlling this oscillation, and the stress-energy tensor conservation appears to be of limited use for such an objective. Instead, the argument relies much more heavily on pointwise conservation (or almost-conservation) laws such as (1-13), and on the method of characteristics.

Another possible approach would be to try to construct an approximate solution (or parametrix) to $\phi^{(p)}$, along the lines of (1-7), and show that $\phi^{(p)}$ is close enough to the approximate solution that the convergence can be read off directly (much as it can be from (1-7)). While it does seem possible to construct the approximate solution more or less explicitly, the author was unable to find a sufficiently strong stability theory to then close the argument by comparing the exact solution to the approximate solution. The difficulty is that the standard stability theory for (1-1) (for example, by applying energy estimates to the difference equation) exhibits losses which grow exponentially in time with rate proportional to p, thus requiring the accuracy of the approximate solution to be exponentially small in p before there is hope of connecting the approximate solution to the exact one. Because of this, the proof below avoids all use of perturbation theory[9], and instead estimates the nonlinear solutions $\phi^{(p)}$ directly. It may however be of interest to develop a stability theory for (1-1) which is more uniform in p (perhaps using bounded variation type norms rather than energy space norms?). One starting point may be the perturbation theory for (1-8), explored recently in [Kalyakin 2001].

Our arguments are instead based on a compactness method. It is not difficult to use energy conservation to demonstrate equicontinuity and uniform boundedness in the $\phi^{(p)}$, so we know (from the Arzelá–Ascoli theorem) that the $\phi^{(p)}$ have at least one limit point. It thus suffices to show that all such limit points obey the properties (i)–(vi), and that the properties (i)–(vi) uniquely determine ϕ. The uniqueness is established in Section 3, and is based on many applications of the method of characteristics. To establish that all limit points obey (i)–(vi), we first establish in Section 4 a number of a priori estimates on the solutions $\phi^{(p)}$, in particular obtaining some crucial boundedness and oscillation control on ϕ and its first derivatives, uniformly in p. In Section 5 we then take limits along some subsequence of p going to infinity to recover the desired properties (i)–(vi).

Remark 1.10. It seems of interest to obtain more robust methods for proving results for infinite nonlinear barriers; the arguments here rely heavily on the method of characteristics and so do not seem to easily extend to, say, the $p \to \infty$ limit of the one-dimensional nonlinear Schrödinger equation $iu_t + u_{xx} = |u|^{p-1}u$, or to higher-dimensional nonlinear wave equations. In higher dimensions there is also a serious additional problem, namely that the nonlinearity becomes energy-critical in the limit $p \to \infty$ in two dimensions, and (even worse) becomes energy-supercritical for large p in three and higher dimensions. However, while global existence for defocusing supercritical nonlinear wave equations from large data is a notoriously difficult open problem, there is the remote possibility that the asymptotic case $p \to \infty$ is actually better behaved than that of a fixed p. At the very least, one should be able to *conjecture* what the correct limit of the solution should be. Interestingly, another $p \to \infty$ type limit for an evolution equation, namely that of the p-Laplacian diffusion equation to a ∞-Laplacian diffusion equation, has recently been studied in [Andreu et al. 2009].

2. Notation

We use the asymptotic notation $X \ll Y$ to denote the bound $X \leqslant CY$ for some constant C depending on fixed quantities (for example, the initial data); note that X may be negative, so $X \ll Y$ only provides an upper bound. We also use $O(X)$ to denote any quantity bounded in magnitude by CY (thus we have

[9] Except, of course, for the fact that perturbation theory is used to establish global existence of the $\phi^{(p)}$.

both an upper and a lower bound in this case), and $X \sim Y$ for $X \ll Y \ll X$. If the constant C needs to depend on additional parameters, we will denote this by subscripts, for example, $X \ll_\varepsilon Y$ or $X = O_\varepsilon(Y)$.

It is convenient to use both Cartesian coordinates (t, x) and null coordinates $\langle u, v\rangle$ to parameterise spacetime. To reduce confusion we shall use angled brackets to denote the latter, thus

$$(t, x) = \langle t + x, t - x\rangle$$

and

$$\langle u, v\rangle = \left(\frac{u+v}{2}, \frac{u-v}{2}\right).$$

Thus for instance we might write $\phi\langle u, v\rangle$ for $\phi(\frac{u+v}{2}, \frac{u-v}{2})$.

We will frequently rely on three reflection symmetries of (1-1) to normalise various signs: the *time reversal symmetry*

$$\phi(t, x) \mapsto \phi(-t, x) \tag{2-1}$$

(which also swaps u with $-v$), the *space reflection symmetry*

$$\phi(t, x) \mapsto \phi(t, -x) \tag{2-2}$$

(which also swaps u with v), and the *sign reversal symmetry*

$$\phi(t, x) \mapsto -\phi(t, x). \tag{2-3}$$

We will frequently be dealing with (closed) *diamonds* in spacetime, which we define to be regions of the form

$$\big\{\langle u, v\rangle : u_0 - r \leqslant u \leqslant u_0;\ v_0 - r \leqslant v \leqslant v_0\big\}$$

for some $u_0, v_0 \in \mathbb{R}$ and $r > 0$. One can of course define open diamonds similarly. We will also be dealing with *triangles*

$$\big\{\langle u, v\rangle : u_0 - r \leqslant u \leqslant u_0;\ v_0 - r \leqslant v \leqslant v_0; \quad u + v \geqslant u_0 + v_0 - r\big\}$$

which are the upper half of diamonds.

3. Uniqueness

In this section we show that there is at most one function $\phi : \mathbb{R} \to \mathbb{R}$ obeying the properties (i)–(vi) listed in Theorem 1.3. It suffices to prove uniqueness on a diamond region $\{\langle u, v\rangle : |u|, |v| \leqslant T\}$ for any fixed $T > 0$; by the time reversal symmetry (2-1) it in fact suffices to prove uniqueness on a triangular region

$$\Delta := \{\langle u, v\rangle = (t, x) : |u|, |v| \leqslant T;\ t \geqslant 0\}.$$

Suppose for contradiction that uniqueness failed on Δ, then there would exist two functions $\phi, \phi' : \Delta \to \mathbb{R}$ obeying the properties (i)–(vi) in Theorem 1.3 which did not agree identically on Δ.

By property (iii), ϕ and ϕ' already agree on some neighbourhood of the time axis. Since ϕ, ϕ' are continuous by (i), and Δ is compact, we may therefore find some $(t_0, x_0) \in \Delta$ with $0 < t_0 < T$ such that $\phi(t, x) = \phi'(t, x)$ for all $(t, x) \in \Delta$ with $t \leqslant t_0$, but such that ϕ and ϕ' do not agree in any neighbourhood of (t_0, x_0) in Δ. We will show that ϕ and ϕ' must in fact agree in some neighbourhood of (t_0, x_0), achieving the desired contradiction.

First suppose that $|\phi(t_0, x_0)| < 1$, then of course $|\phi'(t_0, x_0)| < 1$. As ϕ, ϕ' are both continuous, we thus have ϕ, ϕ' bounded away from -1 and $+1$ on a neighbourhood of (t_0, x_0) in Δ. By (v), ϕ, ϕ' both solve the free wave equation in the sense of distributions on this region, and since they agree below (t_0, x_0), they must therefore agree on a neighbourhood of (t_0, x_0) by uniqueness of the free wave equation, obtaining the desired contradiction. Thus we may assume that $\phi(t_0, x_0) = \phi'(t_0, x_0)$ has magnitude 1; by the sign reversal symmetry (2-3) we may take

$$\phi(t_0, x_0) = \phi'(t_0, x_0) = +1.$$

By continuity, we thus see that ϕ, ϕ' is positive in a neighbourhood of (t_0, x_0); from (v) we conclude that $-\phi_{tt} + \phi_{xx}$ is a nonnegative measure in this neighbourhood. Integrating this, we conclude that

$$\phi\langle u, v\rangle - \phi\langle u-a, v\rangle - \phi\langle u, v-b\rangle + \phi\langle u-a, v-b\rangle \leqslant 0 \tag{3-1}$$

whenever $a, b > 0$ and $\langle u, v\rangle$, $\langle u-a, v\rangle$, $\langle u, v-b\rangle$, $\langle u-a, v-b\rangle \in \Delta$ lie sufficiently close to (t_0, x_0); this implies in particular that ϕ_u is nonincreasing in v (and ϕ_v nonincreasing in u) in this region whenever the derivatives are defined. Similarly for ϕ'.

3.1. *Extension to the right.* Write $\langle u_0, v_0\rangle := (t_0, x_0)$. We already know that $\phi\langle u, v\rangle = \phi'\langle u, v\rangle$ when $\langle u, v\rangle$ is sufficiently close to $\langle u_0, v_0\rangle$ and $u + v \leqslant u_0 + v_0$. We now make extend this equivalence to the right of u_0, v_0:

Lemma 3.2. *ϕ and ϕ' agree for all $\langle u, v\rangle$ sufficiently close to $\langle u_0, v_0\rangle$ in the region $u \geqslant u_0$, $v \leqslant v_0$.*

Proof. We may of course assume $u_0 < T$ otherwise this extension is vacuous.

Suppose first that $\phi\langle u_0, v_0 - b\rangle = 1$ for a sequence of positive b approaching zero. Since $a \mapsto \phi\langle u_0, v_0 - a\rangle$ is piecewise smooth by (ii), we see from the mean value theorem that $\phi_v\langle u_0, v_0 - b'\rangle = 0$ for a sequence of positive b' approaching zero; by (vi) the same is true for $\phi^{(\mathrm{lin})}$, which by the nondegeneracy assumption (c) implies that $\phi_v^{(\mathrm{lin})}$ must in fact vanish on some left-neighbourhood of v_0, which by (vi) implies that $\phi_v\langle u, v\rangle$ and $\phi'_v\langle u, v\rangle$ vanish almost everywhere whenever v is less than v_0 and sufficiently close to v_0. Since ϕ, ϕ' already agree for $u + v \leqslant u_0 + v_0$, and are Lipschitz continuous, an integration in the v direction then gives the lemma.

Now consider the opposite case, where $\phi\langle u_0, v_0 - b\rangle < 1$ for all sufficiently small positive b. As $b \mapsto \phi\langle u_0, v_0 - b\rangle$ is piecewise smooth, and its stationary points have finitely many connected components in Δ by (vi), (c), we conclude that $\phi_v\langle u_0, v - b\rangle > 0$ for all $0 < b < b_0$ for some sufficiently small b_0.

By continuity, we can find a small $a_0 \in (0, b_0)$ such that $\phi\langle u_0 + a, v - b_0\rangle < 1$ for all $0 < a < a_0$. Since $a \mapsto \phi\langle u_0 + a, v - b_0\rangle$ is piecewise smooth and the stationary points have finitely many connected components in Δ, we see (after shrinking a_0 if necessary) that $\phi_u\langle u_0 + a, v - b_0\rangle$ is either always positive, always negative, or always zero for $0 < a < a_0$.

If $\phi_u\langle u_0 + a, v - b_0\rangle$ is always zero, then by (vi) we see that $\phi_u\langle u, v\rangle = \phi'_u\langle u, v\rangle = 0$ almost everywhere for u sufficiently close to and larger than u_0, and any v, and the lemma then follows from the fundamental theorem of calculus.

If $\phi_u\langle u_0 + a, v - b_0\rangle$ is always negative, then as ϕ_u is nonincreasing in v, we see that $\phi_u\langle u_0 + a, v - b\rangle$ is negative for almost every $0 < a < a_0$ and $-b_0 < b < b_0$; combining this with (vi) we conclude that $\phi_u\langle u_0 + a, v - b\rangle = \phi_u\langle u_0 + a, v - b_0\rangle$ almost everywhere in this region, and similarly for ϕ'. In particular, ϕ_u and ϕ'_u agree in this region, and the lemma again follows from the fundamental theorem of calculus.

Finally, we handle the most difficult case,[10] when $\phi_u\langle u_0+a, v-b_0\rangle$ is always positive. Let Φ denote the unique Lipschitz-continuous solution to the free wave equation on the parallelogram $P := \{\langle u, v\rangle : u_0 \leqslant u \leqslant u_0+a_0;\ v_0-b_0 \leqslant v \leqslant v_0\}$ that agrees with ϕ (and hence ϕ') on the lower two edges of this parallelogram (that is, when $u = u_0$ or $v = v_0 - b_0$). We will show that ϕ and ϕ' agree on P.

Call a point $\langle u, v\rangle \in P$ *good* if we have $\Phi(u', v') \leqslant 1$ for all $\langle u', v'\rangle \in P$ with $u' \leqslant u, v' \leqslant v$; the set of good points is then a closed subset of P.

We first observe that if $\langle u, v\rangle$ is good, then $\phi = \phi' = \Phi$. Indeed, since ϕ_v is positive on the lower left edge of P, and ϕ_u is positive on the lower right edge, we see that $\Phi\langle u', v'\rangle < 1$ for all $\langle u', v'\rangle \in P$ with $u' \leqslant u, v' \leqslant v$, and $\langle u', v'\rangle \neq \langle u, v\rangle$. Also, from (v), ϕ and ϕ' solve the free wave equation in the neighbourhood of any region near $\langle u_0, v_0\rangle$ where they are strictly less than 1. A continuity argument based on the uniqueness of the free wave equation then shows that $\phi\langle u', v'\rangle = \phi'\langle u', v'\rangle = \Phi\langle u', v'\rangle$ for all $\phi_u\langle u_0+a, v-b_0\rangle$, and the claim follows.

Now suppose that $\langle u, v\rangle$ is not good, thus $\Phi\langle u, v\rangle > 1$. Excluding a set of measure zero, we may assume that $v_0 - b_0 < v < v_0$ and that $|\phi_v\langle u, v\rangle|$ exists and is equal to $|\phi_v\langle u_0, v\rangle|$, which is nonzero. We claim that $\phi_v\langle u, v\rangle$ cannot be positive. For if it is, then we can find v' less than v and arbitrarily close to v such that

$$\phi\langle u, v'\rangle < \phi\langle u, v\rangle.$$

See Figure 2. Applying (3-1), we conclude that

$$\phi\langle u', v'\rangle < \phi\langle u', v\rangle$$

for all $u_0 \leqslant u' \leqslant u$. In particular, $\phi\langle u', v'\rangle$ must be bounded away from 1. Also, as Φ is continuous, we may also assume that

$$\Phi\langle u, v'\rangle > 1.$$

In particular, ϕ and Φ disagree at u, v'. This implies that $\phi\langle \tilde{u}, \tilde{v}\rangle = 1$ for at least one $u_0 \leqslant \tilde{u} \leqslant u$, $v_0 - b_0 \leqslant \tilde{v} \leqslant v'$, since otherwise by (v) ϕ would solve the free wave equation in this region and thus be necessarily equal to Φ by uniqueness of that equation. Among all such $\langle \tilde{u}, \tilde{v}\rangle$, we can (by continuity and compactness) pick a pair that maximises $\tilde{v}$. Since $\phi\langle u', v'\rangle$ must be bounded away from 1, we have $\tilde{v} < v'$. From (v), ϕ solves the linear wave equation in the parallelogram $\{\langle u'', v''\rangle : u_0 \leqslant u'' \leqslant \tilde{u};\ \tilde{v} \leqslant v'' \leqslant v'\}$, which implies that

$$\phi\langle \tilde{u}, v'\rangle - \phi\langle \tilde{u}, \tilde{v}\rangle - \phi\langle u_0, v'\rangle + \phi\langle u_0, \tilde{v}\rangle = 0.$$

But by the fundamental theorem of calculus, $\phi\langle u_0, v'\rangle > \phi\langle u_0, \tilde{v}\rangle$, and $\phi\langle \tilde{u}, \tilde{v}\rangle = 1$, hence $\phi\langle \tilde{u}, v'\rangle > 1$, contradicting (iv). Hence $\phi_v\langle u, v\rangle$ cannot be positive, and hence by hypothesis must be equal to $-|\phi_v\langle u_0, v\rangle|$. The same considerations apply to ϕ', and so $\phi_v = \phi'_v$ at almost every point in P that is not good.

Since $\phi = \phi'$ at good points in P, and $\phi_v = \phi'_v$ on all other points of P, we obtain the lemma from the fundamental theorem of calculus. □

[10] Indeed, this is the one case where ϕ will reflect itself on some spacelike curve containing (t_0, x_0), which is essentially the only interesting nonlinear phenomenon that ϕ can exhibit.

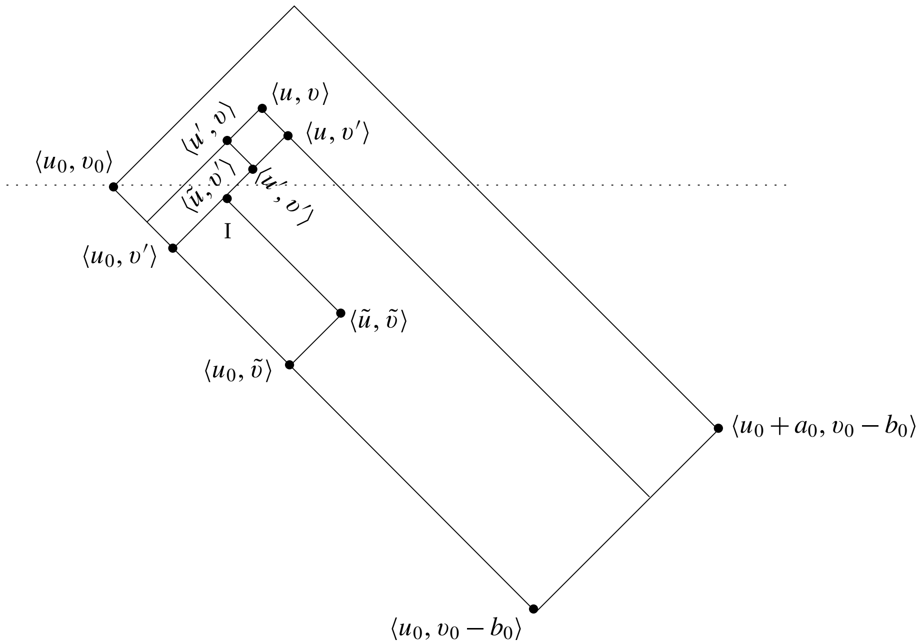

Figure 2. The proof of Lemma 3.2 in the most difficult case. ϕ and ϕ' are already known to agree below the dotted line; the task is to extend this agreement to the parallelogram P with A, B, C as three of its corners. ϕ is also known in this case to be strictly decreasing from A to B and strictly increasing from B to C.

3.3. ***Extension to the left.*** Using space reflection symmetry (2-2), we can reflect the previous lemma and conclude

Lemma 3.4. *ϕ and ϕ' agree for all $\langle u, v\rangle$ sufficiently close to $\langle u_0, v_0\rangle$ in the region $u \leqslant u_0$, $v \geqslant v_0$.*

3.5. ***Extension to the future.*** By the previous lemmas, we know that there exist $a_0, b_0 > 0$ such that $\phi\langle u_0 + a, v_0\rangle = \phi'\langle u_0 + a, v_0\rangle$ for all $0 \leqslant a \leqslant a_0$ and $\phi\langle u_0, v_0 + b\rangle = \phi'\langle u_0, v_0 + b\rangle$ for all $0 \leqslant a \leqslant a_0$. To finish the uniqueness claim, we need to show that ϕ, ϕ' also agree in the future parallelogram $F := \{\langle u, v\rangle : u_0 \leqslant u \leqslant u_0 + a_0;\ v_0 \leqslant v \leqslant v_0 + b_0\}$.

As before, by shrinking a_0 if necessary, we know that $\phi_u\langle u_0 + a, v_0\rangle$ is either always positive, always zero, or always negative for $0 < a < a_0$. The former option is not possible from the barrier condition (iv) since $\phi\langle u_0, v_0\rangle = 1$, so $\phi_u\langle u_0 + a, v_0\rangle$ is always nonpositive. Using the monotonicity of ϕ_u in v and (vi) we thus conclude that $\phi_u\langle u_0 + a, v_0 + b\rangle = \phi_u\langle u_0 + a, v_0\rangle$ for almost every $0 < a < a_0$, $0 < b < b_0$, and similarly for ϕ', such that ϕ_u and ϕ'_u agree almost everywhere on F; similarly ϕ_v and ϕ'_v agree. The claim now follows from the fundamental theorem of calculus, and the uniqueness claim is complete.

Remark 3.6. One could convert the above uniqueness results, with additional effort, into an existence result, but existence of a solution to (i)–(vi) will be automatic for us from the Arzelá–Ascoli theorem, as we will show that any uniformly convergent sequence of $\phi^{(p)}$ will converge to a solution to (i)–(vi).

4. A priori estimates

The next step in the proof of Theorem 1.3 is to establish various a priori estimates on the solution to (1-1) on the diamond

$$\diamond_{T_0} := \{\langle u, v\rangle : |u|, |v| \leqslant T_0\}$$

for some large parameter T_0. Accordingly, let us fix ϕ_0, ϕ_1 obeying the hypotheses (a),(b),(c) of Theorem 1.3, and let $T_0 > 0$; we allow all implied constants to depend on the initial data ϕ_0, ϕ_1 and T_0, but will carefully track the dependence of constants on p. We will assume that p is sufficiently large depending on the initial data and on T_0; in particular, we may take $p \geqslant 100$, say. Standard energy methods (see, for example, [Sogge 1995]) then show that ϕ exists globally and is C^{10} in $\diamond_{T_0}$. This is sufficient to justify all the formal computations below.

We begin with a preliminary (and rather crude) Hölder continuity estimate, which we need to establish some spatial separation (uniformly in p) between the region where ϕ approaches $+1$, and the region where ϕ approaches -1.

Lemma 4.1 (Hölder continuity). *For any $(x_1, t_1), (x_2, t_2) \in \diamond_{T_0}$ we have*

$$\left|\phi(t_1, x_1) - \phi(t_2, x_2)\right| \ll |x_1 - x_2|^{1/2} + |t_1 - t_2|^{1/2}.$$

Proof. We use the monotonicity of local energy

$$E(t) := \int_{x:(t,x)\in\diamond_{T_0}} \frac{1}{2}\phi_t^2 + \frac{1}{2}\phi_x^2 + \frac{1}{p+1}|\phi|^{p+1}\, dx.$$

The standard energy flux identity shows that $E(t) \leqslant E(0)$ for all t. From the hypotheses (a), (b) we have

$$|E(0)| \ll 1$$

(note in particular the uniformity in p), and thus by energy monotonicity

$$|E(t)| \ll 1 \tag{4-1}$$

for all t. By Cauchy–Schwarz, we see in particular that we have some Hölder continuity in space, or more precisely that

$$\left|\phi(t, x) - \phi(t, x')\right| \ll |x - x'|^{1/2} \tag{4-2}$$

whenever $(t, x), (t, x') \in \diamond_{T_0}$. We can also get some Hölder continuity in time by a variety of methods. For instance, from Cauchy–Schwarz again we have

$$\left|\phi(t_1, x) - \phi(t_2, x)\right|^2 \leqslant |t_1 - t_2| \left(\int_{t_1}^{t_2} \phi_t^2\, dt\right)$$

for any $t_1 < t_2$ and any x; integrating this in x on some interval $[x_0 - r, x_0 + r]$ and using (4-1) we obtain

$$\int_{x_0-r}^{x_0+r} \left|\phi(t_1, x) - \phi(t_2, x)\right|^2 dx \ll |t_1 - t_2|^2$$

when $(t_1, x_0 - r), (t_1, x_0 + r), (t_2, x_0 - r), (t_2, x_0 + r) \in \diamond_{T_0}$. On the other hand, from (4-2) we have

$$\left|\phi(t_1, x_0) - \phi(t_2, x_0)\right|^2 \ll |\phi(t_1, x) - \phi(t_2, x)|^2 + r$$

and thus

$$2r\left|\phi(t_1,x_0)-\phi(t_2,x_0)\right|^2 \ll |t_1-t_2|^2+r^2.$$

If we optimise $r := |t_1-t_2|$, we obtain

$$\left|\phi(t_1,x_0)-\phi(t_2,x_0)\right| \ll |t_1-t_2|^{1/2};$$

combining this with (4-2) we obtain the claim (possibly after replacing T_0 with a slightly larger quantity in the above argument). □

Next, we express the Equation (1-1) in terms of the null derivatives (1-11) as

$$\phi_{uv} = -\tfrac{1}{4}|\phi|^{p-1}\phi. \tag{4-3}$$

We can use this to give some important pointwise bounds on ϕ and its derivatives. For any time $-T_0 \leqslant t_0 \leqslant T_0$, let $K(t_0)$ be the best constant such that

$$|\phi(t_0,x)| \leqslant 1+\frac{\log p}{p}+\frac{K}{p} \tag{4-4}$$

and

$$|\phi_u(t_0,x)|,\ |\phi_v(t_0,x)| \leqslant K \tag{4-5}$$

and

$$|\phi_{uu}(t_0,x)|,\ |\phi_{vv}(t_0,x)| \leqslant Kp \tag{4-6}$$

for all x with $(t_0,x)\in\diamondsuit_{T_0}$ (compare with (1-5)). Thus for instance $K(0)\ll 1$. We now show that the (4-4) component of $K(t_0)$, at least, is stable on short time intervals.

Lemma 4.2 (Pointwise bound). *There exists a time increment $\tau>0$ (depending only on the initial data and T_0) such that*

$$|\phi(t_1,x_1)| \leqslant 1+\frac{\log p}{p}+O_{K(t_0)}\Big(\frac{1}{p}\Big)$$

for any $-T_0\leqslant t_0\leqslant T_0$ and $(t_1,x_1)\in\diamondsuit_{T_0}$ with $|t_1-t_0|\leqslant\tau$, and either $t_1\geqslant t_0\geqslant 0$ or $t_1\leqslant t_0\leqslant 0$.

Proof. We shall use the method of characteristics. We take $\tau>0$ to be a small quantity depending on the initial data to be chosen later. Fix t_0 and write $K := K(t_0)$. Let $\epsilon_K>0$ be a small quantity depending on K and the initial data to be chosen later, and then let C_K be a large quantity depending on $K,\epsilon_K>0$ to be chosen later. By time reversal symmetry (2-1) we may take $t_1\geqslant t_0\geq 0$; by sign reversal symmetry (2-3) it suffices to establish the upper bound

$$\phi(t_1,x_1)\leqslant 1+\frac{\log p}{p}+\frac{C_K}{p}.$$

Assume this bound fails, then (by continuity and compact support in space) there exists $t_0\leqslant t_1\leqslant t_0+\tau$ and $x_1\in\mathbb{R}$ such that

$$\phi(t_1,x_1)=1+\frac{\log p}{p}+\frac{C_K}{p} \tag{4-7}$$

and

$$\phi(t,x)\leqslant 1+\frac{\log p}{p}+\frac{C_K}{p} \tag{4-8}$$

for all $t_0 \leqslant t < t_1$ and $x \in \mathbb{R}$. From (4-4), we have $t_1 > t_0$ if C_K is large enough.

From Lemma 4.1 we see (if τ is small enough) that $\phi(t,x)$ is positive in the triangular region

$$\Delta := \{(t,x) : t_0 \leqslant t \leqslant t_1;\ |x - x_1| \leqslant |t - t_1|\}$$

(which is contained in $\diamond_{T_0}$). From (4-3) we conclude that ϕ_u is decreasing in the v direction in Δ, and thus has an upper bound $\phi_u \leqslant K$ on this region thanks to (4-5). Similarly we have $\phi_v \leqslant K$ on Δ. Applying the fundamental theorem of calculus and (4-4) we conclude that

$$\phi(t,x) \leqslant 1 + \frac{\log p}{p} + \frac{K}{p} + O(K(t_1 - t_0)),$$

for all $(t,x) \in \Delta$, which when compared with (4-7) shows (if ϵ_K is small enough and C_K is large enough) that

$$t_1 \geqslant t_0 + 2r$$

where $r := \frac{\epsilon_K C_K}{p}$.

Now we consider the diamond region

$$\diamond := \{(t,x) : t_1 + x_1 - r \leqslant t + x \leqslant t_1 + x_1; \quad t_1 - x_1 - r \leqslant t - x \leqslant t_1 - x_1\}.$$

Since $t_1 > 2r$, this diamond is contained in the triangle Δ (indeed, it is nestled in the upper tip of that triangle). As before, we have the upper bounds

$$\phi_u, \phi_v \leqslant K$$

on this diamond. From this, (4-7), and the fundamental theorem of calculus, we have

$$\phi(t,x) \geqslant 1 + \frac{\log p}{p} \tag{4-9}$$

on this diamond (if ϵ_K is small enough). Applying (4-3) we conclude that

$$\phi_{uv} \leqslant -\left(1 + \frac{\log p}{p}\right)^p \leqslant -cp$$

for some absolute constant $c > 0$. Integrating this on the diamond we conclude that

$$\phi(t_1, x_1) - \phi(t_1 - r, x_1 - r) - \phi(t_1 - r, x_1 + r) + \phi(t_1 - 2r, x_0) \leqslant -cpr^2.$$

But from (4-8), (4-9), the left side is bounded below by $-O(C_K/p)$. We conclude that

$$pr^2 \ll \frac{C_K}{p}.$$

But from the definition of r, we obtain a contradiction if C_K is large enough depending on ϵ_K, and the claim follows. □

Now we establish a similar stability for the (4-5) component of $K(t_0)$.

Lemma 4.3 (Pointwise bound for derivatives). *There exists a time increment $\tau > 0$ (depending only on the initial data and T_0) such that*

$$|\phi_u(t_1, x_1)|, |\phi_v(t_1, x_1)| \ll_{K(t_0)} 1$$

and

$$|\phi_{uu}(t_1,x_1)|,\ |\phi_{vv}(t_1,x_1)| \ll_{K(t_0)} p$$

for any $-T_0 \leqslant t_0 \leqslant T_0$ *and* $(t_1,x_1) \in \diamond_{T_0}$ *with* $|t_1-t_0| \leqslant \tau$, *and either* $t_1 \geqslant t_0 \geqslant 0$ *or* $t_1 \leqslant t_0 \leqslant 0$, *if* p *is large enough depending on* $K(t_0)$.

Proof. This will be a more advanced application of the method of characteristics. We again let $\tau > 0$ be a sufficiently small quantity (depending on the initial data) to be chosen later. Fix t_0 and write $K = K(t_0)$, and let $C_K > 0$ be a large quantity depending on K and the initial data to be chosen later.

It will suffice to show that

$$\big|\phi_u(t_1,x_1)\big|^2 + \frac{1}{p}\big|\phi_{uu}(t_1,x_1)\big|,\ \big|\phi_v(t_1,x_1)\big|^2 + \frac{1}{p}\big|\phi_{vv}(t_1,x_1)\big| \leqslant C_K$$

whenever $|t_1-t_0| \leqslant \tau$ and $x_1 \in \mathbb{R}$.

Suppose for contradiction that this claim failed. As before (using the symmetries (2-1), (2-2)) we may assume that $0 \leqslant t_0 \leqslant t_1 \leqslant t_0+\tau$ and x_1 are such that

$$\big|\phi_u(t_1,x_1)\big|^2 + \frac{1}{p}\big|\phi_{uu}(t_1,x_1)\big| = C_K, \tag{4-10}$$

say, and that

$$\big|\phi_u(t,x)\big|^2 + \frac{1}{p}\big|\phi_{uu}(t,x)\big|,\ \big|\phi_v(t,x)\big|^2 + \frac{1}{p}\big|\phi_{vv}(t,x)\big| \leqslant C_K \tag{4-11}$$

for all $t_0 \leqslant t \leqslant t_1$ and x with $(t,x) \in \diamond_{T_0}$.

We first dispose of an easy case when $\phi(t_1,x_1)$ is small, say $|\phi(t_1,x_1)| \leqslant 1/2$. Then by Lemma 4.1 we conclude (if τ is small enough) that $|\phi| \leqslant 1$ on the triangular region $\{(t,x) : t_0 \leqslant t \leqslant t_1 : |x_1-x| \leqslant |t_1-t|\}$, and the claim then easily follows from (4-3), the fundamental theorem of calculus, and (4-5). Thus we may assume $|\phi(t_1,x_1)| > 1/2$; replacing ϕ with $-\phi$ if necessary we may assume $\phi(t_1,x_1) > 1/2$.

By Lemma 4.1, we see that ϕ is positive whenever $|t-t_1|, |x-x_1| \leqslant 100\tau$, say, if τ is small enough.

It will be convenient to make the change of variables

$$\phi(t,x) = p^{1/(p-1)}\Big(1 + \frac{1}{p}\psi(p(t-t_0), p(x-x_1))\Big);$$

then from (4-3), ψ solves the equation

$$\psi_{uv} = -\frac{1}{4}\Big(1+\frac{1}{p}\psi\Big)^p \tag{4-12}$$

on the region $|t|, |x| \leqslant 100\tau p$. From (4-10), (4-11) we have

$$\big|\psi_u(p(t_1-t_0),0)\big|^2 + \big|\psi_{uu}(p(t_1-t_0),0)\big| \sim C_K \tag{4-13}$$

and

$$|\psi_u(t,x)|^2,\ |\psi_v(t,x)|^2,\ |\psi_{uu}(t,x)|,\ |\psi_{vv}(t,x)| \ll C_K. \tag{4-14}$$

for $0 \leqslant t \leqslant p(t_1-t_0)$ and $|x| \leqslant 100\tau p$. Meanwhile, while from Lemma 4.2 (and shrinking τ as necessary) we have the upper bound

$$\psi(t,x) \ll_K 1 \tag{4-15}$$

for $|t|, |x| \leqslant 100\tau p$. Note though that we do not expect ψ to enjoy a comparable lower bound, but since ϕ is positive in the region of interest, we have

$$\psi(t,x) \geqslant -p \tag{4-16}$$

for $|t|, |x| \leqslant 100\tau p$. Finally, from (4-5), (4-4) we have

$$|\psi_u(0,x)|,\ |\psi_v(0,x)|,\ |\psi_{uu}(0,x)|,\ |\psi_{vv}(0,x)| \ll_K 1 \tag{4-17}$$

whenever $|x| \leqslant 100\tau p$.

Motivated by the pointwise conservation laws (1-13) of the Equation (1-12), which (4-12) formally converges to, we consider the quantity

$$\partial_v\Big(\frac{1}{2}\psi_u^2 - \psi_{uu}\Big).$$

Using (4-12), we compute

$$\partial_v\Big(\frac{1}{2}\psi_u^2 - \psi_{uu}\Big) = -\frac{1}{4p}F_p(\psi)\psi_v \tag{4-18}$$

where $F_p(s) := s\big(1+\frac{1}{p}s\big)^{p-1}$. From (4-15) we see that $F_p(\psi) = O_K(1)$, and thus by (4-14) we have

$$\partial_v\Big(\frac{1}{2}\psi_u^2 - \psi_{uu}\Big) = O_K(C_K^{1/2}/p).$$

From this, (4-17), (4-14) and the fundamental theorem of calculus we see that

$$\Big|\frac{1}{2}\psi_u^2 - \psi_{uu}\Big|(t,x) \leqslant A \tag{4-19}$$

for all $0 \leqslant t \leqslant p(t_1 - t_0)$ and $|x| \leqslant 50\tau p$, and some $A \sim_K C_K^{1/2}$.

From (4-12) we know that ψ_u is decreasing in the v direction, so from (4-17) we also have the upper bound

$$\psi_u(t,x) \leqslant O_K(1)$$

for all $0 \leqslant t \leqslant p(t_1 - t_0)$ and $|x| \leqslant 50\tau p$. To get a lower bound, suppose that $\psi_u(t,x) \leqslant -A^{1/2}$ for some $0 \leqslant t \leqslant p(t_1 - t_0)$ and $|x| \leqslant 40\tau p$. Then from (4-19) we have $\psi_{uu}(t,x) \geqslant 0$. If we move backwards in the u direction, we thus see that ψ_u decreases; continuing this (by the usual continuity argument) until we hit the initial surface $t = 0$ and applying (4-17) we conclude that $\psi_u(t,x) \geqslant -O_K(1)$, a contradiction if C_K is large enough. We thus conclude that $\psi_u \geqslant -A^{1/2}$ for $0 \leqslant t \leqslant p(t_1 - t_0)$ and $|x| \leqslant 40\tau p$. Combining this with the upper bound and with (4-19) we contradict (4-13) if C_K is large enough, and the claim follows. □

Combining Lemmas 4.2, 4.3 and the definition of $K(t)$ we conclude that

$$K(t) \ll_{K_{t_0}} 1$$

whenever $|t - t_0| \leqslant \tau$ and $T_0 \geqslant t \geqslant t_0 \geqslant 0$ or $-T_0 \leqslant t \leqslant t_0 \leqslant 0$. Since $K(0) \ll 1$, we thus conclude on iteration that

$$K(t) \ll 1$$

for all $t \in [-T_0, T_0]$. Thus we have

$$|\phi(t,x)| \leqslant 1 + \frac{\log p}{p} + O\Big(\frac{1}{p}\Big), \tag{4-20}$$

$$|\phi_u(t,x)|,\ |\phi_v(t,x)| \ll 1, \tag{4-21}$$

$$|\phi_{uu}(t,x)|,\ |\phi_{vv}(t,x)| \ll p, \tag{4-22}$$

for all $(t,x) \in \diamondsuit_{T_0}$.

Now we revisit the conservation laws (1-13) which were implicitly touched upon in the proof of Lemma 4.3.

Lemma 4.4 (Approximate pointwise conservation law). *There exists $\tau > 0$ (depending on T_0 and the initial data) such that the following claim holds: whenever $\langle u_0, v_0\rangle \in \diamondsuit_{T_0}$ is such that $\phi\langle u_0, v_0\rangle \geqslant -1/2$, then*

$$\Big(\frac{1}{2}\phi_u^2 - \frac{1}{p}\phi_{uu}\Big)\langle u_0, v_0 + r\rangle = \Big(\frac{1}{2}\phi_u^2 - \frac{1}{p}\phi_{uu}\Big)\langle u_0, v_0\rangle + O\Big(\frac{\log p}{p}\Big) \tag{4-23}$$

and

$$\Big(\frac{1}{2}\phi_v^2 - \frac{1}{p}\phi_{vv}\Big)\langle u_0 + r, v_0\rangle = \Big(\frac{1}{2}\phi_v^2 - \frac{1}{p}\phi_{vv}\Big)\langle u_0, v_0\rangle + O\Big(\frac{\log p}{p}\Big)$$

for all $-\tau \leqslant r \leqslant \tau$.

If instead $\phi\langle u_0, v_0\rangle \leqslant +1/2$, then we have

$$\Big(\frac{1}{2}\phi_u^2 + \frac{1}{p}\phi_{uu}\Big)\langle u_0, v_0 + r\rangle = \Big(\frac{1}{2}\phi_u^2 + \frac{1}{p}\phi_{uu}\Big)\langle u_0, v_0\rangle + O\Big(\frac{\log p}{p}\Big)$$

and

$$\Big(\frac{1}{2}\phi_v^2 + \frac{1}{p}\phi_{vv}\Big)\langle u_0 + r, v_0\rangle = \Big(\frac{1}{2}\phi_v^2 + \frac{1}{p}\phi_{vv}\Big)\langle u_0, v_0\rangle + O\Big(\frac{\log p}{p}\Big)$$

for all $-\tau \leqslant r \leqslant \tau$.

Proof. Let $\tau > 0$ be sufficiently small to be chosen later. By sign reversal symmetry (2-3) we may assume that $\phi\langle u_0, v_0\rangle \geqslant -1/2$. By spatial reflection symmetry (2-2) it suffices to prove (4-23).

Suppose first that $-1/2 \leqslant \phi\langle u_0, v_0\rangle \leqslant 1/2$, then by Lemma 4.1 we have $|\phi\langle u, v\rangle| \leqslant 0.9$i, say, whenever $|u - u_0|, |v - v_0| \leqslant 100\tau$. Applying (4-3) and the fundamental theorem of calculus, we see that

$$\phi_u\langle u_0 + r, v_0\rangle = \phi_u\langle u_0, v_0\rangle + O\Big(\frac{\log p}{p}\Big)$$

for all $-\tau \leqslant r \leqslant \tau$; similarly, if one differentiates (4-3) in the u direction and applies the bound $|\phi\langle u, v\rangle| \leqslant 0.9$ as well as (4-21), we obtain

$$\phi_{uu}\langle u_0 + r, v_0\rangle = \phi_{uu}\langle u_0, v_0\rangle + O(\log p)$$

and the claim (4-23) follows.

Henceforth we assume $\phi\langle u_0, v_0\rangle > 1/2$. By Lemma 4.1 (or (4-21)) we see that $\phi\langle u, v\rangle$ is positive when $|u - u_0|, |v - v_0| \leqslant 100\tau$, so by making the ansatz

$$\phi\langle u, v\rangle = p^{1/(p-1)}\Big(1 + \frac{1}{p}\psi\big\langle p(u - u_0), p(v - v_0)\big\rangle\Big)$$

as before, we see that ψ obeys (4-12) for $|u|, |v| \leqslant 100\tau p$. Also, from (4-20)–(4-22) (and the positivity of ϕ) we see that

$$|\psi_u\langle u, v\rangle|,\ |\psi_v\langle u, v\rangle|,\ |\psi_{uu}\langle u, v\rangle|,\ |\psi_{vv}\langle u, v\rangle| \ll 1 \tag{4-24}$$

and

$$-p \leqslant \psi\langle u, v\rangle \leqslant O(1) \tag{4-25}$$

for all $|u|, |v| \leqslant 100\tau p$. Our objective is to show that

$$\Big(\frac{1}{2}\psi_u^2 - \psi_{uu}\Big)\langle 0, r\rangle = \Big(\frac{1}{2}\psi_u^2 - \psi_{uu}\Big)\langle 0, 0\rangle + O\Big(\frac{\log p}{p}\Big)$$

for all $-\tau p \leqslant r \leqslant \tau p$. By (4-18) and the fundamental theorem of calculus it suffices to show that

$$\int_{-\tau p}^{\tau p} \big|F_p(\psi\langle 0, r\rangle)\big|\big|\psi_v\langle 0, r\rangle\big|\, dr \ll_T \frac{\log p}{p}. \tag{4-26}$$

Applying (4-24) we can discard the $|\psi_v\langle 0, r\rangle|$ factor. Meanwhile, from (1-12), (4-24), and the fundamental theorem of calculus we have

$$\int_{-\tau p}^{\tau p} e^{\psi\langle 0, r\rangle}\, dr \ll_T 1.$$

Observe that $F_p(x) \ll_T (\log p)e^x$ whenever $-100 \log p \leqslant x \leqslant O_T(1)$, and that $F_p(x) \ll p^{-50}$ when $x \leqslant -100 \log p$, and the claim (4-26) follows. □

We now use this law to show a more precise bound on ϕ_u and ϕ_v than is provided by (4-21). We first handle the case when ϕ has large derivative.

Lemma 4.5 (Piecewise convergence, nondegenerate case). *Let $\varepsilon > 0$, and let $I \subset [-T_0, T_0]$ be an interval such that $|\phi_u^{(\mathrm{lin})}\langle u, v\rangle| \geqslant \varepsilon$ for all $u \in I$ (note that v is irrelevant here). Then for each $v \in [-T_0, T_0]$, we have*

$$|\phi_u\langle u, v\rangle|^2 = |\phi_u^{(\mathrm{lin})}\langle u, v\rangle|^2 + O\Big(\frac{\log p}{p}\Big) \tag{4-27}$$

and

$$\phi_{uu}\langle u, v\rangle = O(\log p) \tag{4-28}$$

for all u in I, excluding at most $O(1)$ intervals in I of length $O_\varepsilon(\frac{\log p}{p})$.

Similarly with the roles of u and v reversed.

Proof. Let $\tau > 0$ be a small number (depending on the initial data and T_0) to be chosen later. We may assume that p is sufficiently large depending on ε, since the claim is trivial otherwise. By space reflection symmetry (2-2) it will suffice to prove (4-27), (4-28); by time reversal symmetry (2-1) we may assume that $t = \frac{u+v}{2}$ is nonnegative.

We introduce an auxiliary parameter $0 \leqslant T \leqslant T_0$, and only prove the claim for $t = \frac{u+v}{2}$ between 0 and T; setting $T = T_0$ will then yield the claim. We establish this claim by induction on T, incrementing T by steps of τ at a time.

Let us first handle the base case when $0 \leqslant T \leqslant 2\tau$. Fix v. By sign reversal symmetry (2-3) and Lemma 4.1 we may assume that $\phi\langle u, v\rangle \geqslant -1/2$ whenever $t = \frac{u+v}{2}$ has magnitude at most 100τ. Applying Lemma 4.4, we conclude that

$$\Big(\frac{1}{2}\phi_u^2 - \frac{1}{p}\phi_{uu}\Big)\langle u, v\rangle = \Big(\frac{1}{2}\phi_u^2 - \frac{1}{p}\phi_{uu}\Big)\langle u, -u\rangle + O\Big(\frac{\log p}{p}\Big)$$

for all $-v - 10\tau \leqslant u \leqslant -v + 10\tau$. From the hypotheses (a), (b) and (1-1) we see that $\phi_{uu}\langle u, -u\rangle = O(1)$, and so we have

$$\phi_{uu}\langle u, v\rangle = \frac{p}{2}\bigg(\phi_u^2\langle u, v\rangle - \phi_u^2\langle u, -u\rangle + O\Big(\frac{\log p}{p}\Big)\bigg). \tag{4-29}$$

If we write $f(t) := \phi_u\langle -v + 2t, v\rangle$, $g(t) := \phi_u\langle -v + 2t, v\rangle$ we thus have

$$f'(t) = p\bigg(f(t)^2 - g(t)^2 + O\Big(\frac{\log p}{p}\Big)\bigg) \tag{4-30}$$

for all $-5\tau \leqslant t \leqslant 5\tau$.

Let $J := \{0 \leqslant t \leqslant T : -v + 2t \in I\}$, thus J is a (possibly empty) interval such that $\varepsilon \leqslant |g(t)| \ll 1$ for all $t \in J$. Also observe from the smoothness of the initial data and (1-1) that $g'(t) = O(1)$ for all $t \in J$. Also from (4-21) we have $f(t) = O(1)$ for all $t \in J$.

Suppose first that $f(t_0) > 0$ and $f(t_0)^2 \geqslant g(t_0)^2 + C\frac{\log p}{p}$ for some $t_0 \in J$ and some sufficiently large C. Then from the bounds on g and (4-30), we have

$$\partial_t(f(t_0)^2 - g(t_0)^2) \gg_\varepsilon p(f(t_0)^2 - g(t_0)^2).$$

A continuity argument (using Gronwall's inequality) then shows that $f(t)^2 - g(t)^2$ increases exponentially fast (with rate $\gg_\varepsilon p$) as t increases. Since $f(t)^2 - g(t)^2$ is $O(1)$ and was $\gg \frac{\log p}{p}$ at t_0, we arrive at a contradiction unless t_0 lies within $O_\varepsilon(\frac{\log p}{p})$ of the boundary of J. Similarly if $f(t_0) < 0$ and $f(t_0)^2 \geqslant g(t_0)^2 + C\frac{\log p}{p}$ for some $t_0 \in J$. We conclude that

$$f(t)^2 \leqslant g(t)^2 + O\left(\frac{\log p}{p}\right)$$

for all $t \in J$ except for those t which are within $O_\varepsilon(\frac{\log p}{p})$ of the boundary of J.

Now suppose that $f(t)^2 \leqslant \varepsilon^2/2$, then we see from (4-30) and the bounds on f, g that $-f'(t) \gg_{T,\varepsilon} p$; thus the set of $t \in J$ for which this occurs must be contained in a single interval of length $O_\varepsilon(\frac{1}{p})$.

Next, if $\varepsilon^2/2 \leqslant f(t)^2 \leqslant g(t)^2 - C\frac{\log p}{p}$, then from (4-30) we obtain a bound of the form

$$\pm\partial_t(g(t)^2 - f(t)^2) \gg_\varepsilon p(g(t)^2 - f(t)^2),$$

where $\pm$ is the sign of $f(t)$. Applying the continuity and Gronwall argument again, either forwards or backwards in time as appropriate, we see that this event can only occur either within $O_\varepsilon(\frac{\log p}{p})$ of the boundary of J, or on an interval of length $O_\varepsilon(\frac{\log p}{p})$ adjacent to the interval where

$$f(t)^2 \leqslant \varepsilon^2/2.$$

Putting all of this together, we see that

$$f(t)^2 = g(t)^2 + O\Big(\frac{\log p}{p}\Big)$$

for all $t \in J$ outside of at most $O(1)$ intervals of length $O_\varepsilon(\frac{\log p}{p})$. This gives the desired bound (4-27). The bound (4-28) then follows from (4-29).

Now suppose inductively that $T > 2\tau$, and that the claim has already been shown for $T - \tau$. By inductive hypothesis we only need to establish the claim for $t \in [T-\tau, T]$. Fix v. By sign reversal symmetry (2-3) and Lemma 4.1 we may assume that $\phi\langle u, v\rangle \geqslant -1/2$ whenever $t = \frac{u+v}{2}$ lies within 100τ of T.

We can now repeat the previous arguments, except that the interval J must first be subdivided by removing the $O(1)$ subintervals of J of length $O_\varepsilon(\frac{\log p}{p})$ for which (4-27) and (4-28) (with v replaced by $v-\tau$) already failed (and which are provided by the inductive hypothesis), and then working on each remaining subinterval of J separately. Note that on each such interval we still have the ODE (4-30) (using the inductive hypothesis (4-27), (4-28) as a substitute for control of the initial data). We omit the details. □

Now we handle the opposing case when ϕ has small derivative.

Lemma 4.6 (Piecewise convergence, degenerate case). *Let $\varepsilon > 0$, and let $I \subset [-T_0, T_0]$ be an interval such that $|\phi_u^{(\mathrm{lin})}\langle u, v\rangle| \leqslant \varepsilon$ for all $u \in I$ (again, v is irrelevant). Then one has*

$$|\phi_u\langle u, v\rangle| \ll \varepsilon \tag{4-31}$$

and

$$|\phi_{uu}\langle u, v\rangle| \ll \varepsilon^2 p \tag{4-32}$$

whenever $\langle u, v\rangle \in \diamond_{T_0}$ and $u \in I$. Similarly with the roles of u and v reversed.

Proof. This is very similar to Lemma 4.5, in that we first establish the base case $0 \leqslant T \leqslant 4\tau$ and then induct by steps of τ, where $\tau > 0$ is a fixed timestep independent of T and p. Whereas in the proof of Lemma 4.5 we did the base case in detail and left the inductive step to the reader, here we shall leave the base case to the reader and do the inductive step in detail. Thus, assume $T > 4\tau$ and that the claim has already been proven for $T - \tau$ and $T - 2\tau$. We may assume $\varepsilon < \tau$ since the claim follows from (4-21), (4-22) otherwise.

By inductive hypothesis and time and space reversal symmetry ((2-1) and (2-2)) we only need to establish the claims (4-31) and (4-32) for $t = \frac{u+v}{2} \in [T-\tau, T]$. By the sign reversal symmetry (2-3) and Lemma 4.1 we may assume that $\phi\langle u, v\rangle > -1/2$ whenever $t \in [T - 100\tau, T + 100\tau]$. We can assume that p is large compared to T_0, ε, and the initial data since the claim is vacuous otherwise.

Let $J := \{u \in I : t = \frac{u+v}{2} \in [T-\tau, T]\}$. Observe (from the smoothness of the initial data) that $|\phi_u\langle u, u\rangle| \ll \varepsilon$ whenever u lies within ε of J. By inductive hypothesis (replacing ε by $O(\varepsilon)$), we conclude that

$$|\phi_u\langle u, v-2\tau\rangle| \ll \varepsilon,$$
$$|\phi_{uu}\langle u, v-2\tau\rangle| \ll \varepsilon^2 p,$$

for all in the ε-neighbourhood of J. Applying Lemma 4.4, we conclude that

$$\Big(\frac{1}{2}\phi_u^2 - \frac{1}{p}\phi_{uu}\Big)\langle u, v\rangle = O(\varepsilon^2) \tag{4-33}$$

for all t in the ε-neighbourhood of J. Thus, if $f(u) := \phi_u\langle u, v\rangle$, then we have

$$f'(u) = p(f^2(u) + O(\varepsilon^2))$$

for all u in the ε-neighbourhood of J.

The ODE $f'(u) = \frac{p}{2} f^2(u)$ blows up either forward or backward in time within a duration of $O_\varepsilon(1/p)$ as soon as $|f(u)|$ exceeds ε. From this and a continuity and comparison argument, we see that $|f(u)|$ cannot exceed $C\varepsilon$ for $u \in J$ for some constant C depending only on u, thus $f(u) = O(\varepsilon)$ for all $u \in J$. Applying (4-33) we close the induction as required; the base case is similar. □

Remark 4.7. The a priori estimates here did not use the full force of the hypotheses (a)–(c); the condition (c) was not used at all, and the strict barrier condition (b) could be replaced by the nonstrict condition $|\phi_0(x)| \leqslant 1$. Also, a careful examination of the dependence of the implied constants on the initial data, combined with a standard limiting argument using the usual local-wellposedness theory reveals that (a) can be replaced with a $C^2 \times C^1$ condition on the initial data (ϕ_0, ϕ_1). However, we use the hypotheses (a)–(c) more fully in the uniqueness theory of the previous section, and the compactness arguments in the next section.

5. Compactness

Now we can prove Theorem 1.3. Fix $T_0 > 0$. From Lemma 4.1, the solutions $\phi = \phi^{(p)}$ are equicontinuous and uniformly bounded on the region $\diamond_{T_0} := \{(t, x) : |t| + |x| \leqslant T_0\}$, and hence (by the Arzelá–Ascoli theorem) precompact in the uniform topology on this region. In view of the uniqueness theory in Section 3, we see that to show Theorem 1.3, it suffices to show that any limit point of this sequence obeys the properties (i)–(vi) on $\diamond_{T_0}$. Accordingly, let $p_n \to \infty$ be a sequence such that $\phi^{(p_n)}$ converges uniformly to a limit ϕ.

We can now quickly verify several of the required properties (i)-(vi). From (4-20) we obtain the property barrier condition (iv), while from Lemma 4.1 we have the Lipschitz condition (i). From the strict barrier hypothesis (b) and Lemma 4.1, we know that the $|\phi^{(p_n)}|$ stay bounded away from 1 in a neighbourhood of the initial interval $\{(0, x) : -T_0 \leqslant x \leqslant T_0\}$, and so the nonlinearity $|\phi^{(p_n)}|^{p_n-1}\phi^{(p_n)}$ converges uniformly to zero in this neighbourhood. Because of this and (1-1), $\phi^{(p_n)}$ converges uniformly to $\phi^{(\mathrm{lin})}$ in this neighbourhood, yielding the initial condition (iii).

5.1. ***The defect measure condition.*** Now we verify (v). Suppose (t_0, x_0) is a point in $\diamond$ such that $|\phi(t_0, x_0)| < 1$. Then by Lemma 4.1 and uniform convergence, we can find a neighbourhood B of (t_0, x_0) in $\diamond$ and a constant $c < 1$ such that $|\phi^{(p)}(t, x)| \leqslant c$ for all $(t, x) \in B$ and all sufficiently large p. In particular, the nonlinearity in (1-1) converges uniformly to zero on B as $p \to \infty$. Taking limits, we see that $-\phi_{tt} + \phi_{xx} = 0$ on B in the sense of distributions. Taking unions over all such B, and using null coordinates we conclude that the distribution $-\phi_{tt}+\phi_{xx}$ is supported on the set $\{(t, x) \in \diamond : |\phi(t, x)| = 1\}$.

Next, we consider the neighbourhood of a point (t_0, x_0) where $\phi(t_0, x_0) = +1$, say. Then by Lemma 4.1 and uniform convergence, we can find a diamond D centred at (t_0, x_0) (with length bounded below uniformly in (t_0, x_0)) such that $\phi^{(p)}(t,x)$ is nonnegative for all $(t,x) \in D$ and all sufficiently large p. In particular, the nonlinearity in (4-3) is nonnegative on this diamond, which implies by the fundamental theorem of calculus that

$$\phi^{(p)}(t,x) - \phi^{(p)}(t-r, x-r) - \phi^{(p)}(t-s, x+s) + \phi^{(p)}(t-r-s, x-r+s) \geqslant 0$$

whenever (t,x), $(t-r, x-r)$, $(t-s, x+s)$, $(t-r-s, x-r+s)$ lie in D. Taking uniform limits, we conclude that the same statement is true for ϕ. By the usual Lebesgue–Stieltjes measure construction (adapted to two dimensions) we thus see that

$$\begin{aligned}\phi(t,x) - \phi(t-r, x-r) - \phi(t-s, x+s) + \phi(t-r-s, x-r+s) \\ = \mu_+\big(\{(t-r'-s', x-r'+s') : 0 \leqslant r' \leqslant r; 0 \leqslant s' \leqslant s\}\big)\end{aligned}$$

for some positive finite measure μ_+ on D, which implies that $-\phi_{tt} + \phi_{xx} = \mu_+$ in the sense of distributions on D. Similarly when $\phi(t_0, x_0) = -1$ (now replacing μ_+ by $-\mu_-$). Piecing together these diamonds D and neighbourhoods B we obtain the claim.

5.2. ***The reflection condition.*** Now we verify (vi). By space reflection symmetry (2-2) it suffices to show (1-17).

Let us first consider the region where $\phi_u^{(\mathrm{lin})}\langle u, v\rangle$ vanishes. Applying Lemma 4.6 and taking weak limits, we see that $\phi_u\langle u, v\rangle$ vanishes almost everywhere when $\phi_u^{(\mathrm{lin})}\langle u, v\rangle$ vanishes, which of course gives (1-17) in this region. As the countable union of null sets is still null, it thus suffice to verify (1-17) for almost every (t,x) in the parallelogram $P := \{\langle u, v\rangle \in \Diamond_{T_0} : u \in I\}$, whenever I is an interval such that

$$|\phi_u^{(\mathrm{lin})}\langle u, v\rangle| \geqslant \varepsilon$$

for all $u \in I$ (v is irrelevant) and some $\varepsilon > 0$. Applying Lemma 4.5 and taking square roots, we know that for all p, we have

$$|\phi_u^{(p)}\langle u, v\rangle| = |\phi_u^{(\mathrm{lin})}\langle u, v\rangle| + O_T\left(\frac{\log^{1/2} p}{p^{1/2}}\right) \tag{5-1}$$

for all $\langle u, v\rangle \in \Diamond_{T_0}$ with $u \in I$, where we exclude for each fixed choice of v, a union $I_v \subset I$ of $O(1)$ intervals of length $O_\varepsilon(\frac{\log p}{p})$ from I.

We would like to take limits as $p = p_n \to \infty$, but we encounter a technical difficulty: while we know that $\phi_u^{(p)}$ converges weakly to ϕ_u, this does not imply that $|\phi_u^{(p)}|$ converges weakly to $|\phi_u|$, due to the possibility of increasing oscillation of sign[11] in $\phi_u^{(p)}$. The fact that (5-1) only fails on a bounded number of short intervals for each v rules out oscillation in the u direction, but one must also address the issue of oscillation in the v direction. Fortunately, from (4-3) we have some monotonicity of $\phi_u^{(p)}$ in v that allows us to control this possibility.

We turn to the details. As ϕ is Lipschitz, we can cover the parallelogram P by a bounded number of open diamonds D in P, on which each ϕ varies by at most 0.1, say. If ϕ takes any value between $-1/2$ and $1/2$ on a diamond D, then by (4-3) ϕ solves the free wave equation on D, so in particular ϕ_u is constant in v (and agrees with $\frac{1}{2}(\phi_1 + \partial_x \phi_0)$ whenever the diamond intersects the initial surface $\{t = 0\}$). Thus it suffices to establish the claim on those diamonds D on which ϕ avoids the interval $[-1/2, 1/2]$;

[11] If one does not address this oscillation issue, one can only get the lower bound in (1-17) rather than equality.

by the symmetry $\phi \to -\phi$ we may assume that $\phi \geqslant 1/2$ on D, and hence (for n large enough) $\phi^{(p_n)}$ is also positive. By (4-3), we conclude that $\phi_u^{(p_n)}$ is decreasing in the v direction.

Let $\delta > 0$ be a small number. We can partition the diamond D into $O_T(\delta^{-2})$ subdiamonds of length δ in a regular grid pattern. Fix n sufficiently large depending on δ, ε, and call a subdiamond *totally positive* (with respect to n) if $\phi_u^{(p_n)} > 0$ at every point on this subdiamond; similarly define the notion of a subdiamond being *totally negative*. Call a subdiamond *degenerate* if it is neither totally positive nor totally negative (that is, it attains a zero somewhere in the diamond). We claim that at most $O_\varepsilon(\delta^{-1})$ degenerate subdiamonds. To see this, let d be a degenerate subdiamond. Since $\phi_u^{(p_n)}$ is decreasing in the v direction, we know that $\phi_u^{(p_n)}$ must be negative in at least one point on the northwest edge of d, and positive in at least one point on the southeast edge. Suppose that $\phi_u^{(p_n)}$ is negative at every point on the northwest edge. Then from the monotonicity of $\phi_u^{(p_n)}$ in the v direction, we see that there can be at most one degenerate subdiamond of this type on each northwest-southeast column of subdiamonds; thus there are only $O(\delta^{-1})$ subdiamonds of this type. Thus we may assume that $\phi_u^{(p_n)}$ changes sign on the northwest edge. But on the line ℓ that that edge lies on, we have (5-1) holding in D outside of $O(1)$ intervals of length $O_\varepsilon(\frac{\log p_n}{p_n})$; also, by hypothesis, we have

$$|\phi_u^{(\mathrm{lin})}\langle u, v\rangle| \geqslant \varepsilon$$

for $\langle u, v\rangle$ in D. We conclude (for n large enough) that there are at most $O(1)$ subdiamonds with northwest edge lying on this line ℓ for which $\phi_u^{(p_n)}$ changes sign on this edge. Summing over all $O(\delta^{-1})$ possible edges, we obtain the claim.

Fix δ, and let $n \to \infty$. The set of subdiamonds on which $\phi^{(p_n)}$ is totally positive or totally negative can change with n; however there are only a finite number of possible values for this set for fixed δ. Hence, by the infinite pigeonhole principle, we may refine the sequence p_n and assume that these sets are in fact independent of n. For any totally positive or totally negative diamond, $\phi_u^{(p_n)}$ has a definite sign; since $\phi_u^{(p_n)}$ converges weakly to ϕ_u, we conclude that $|\phi_u^{(p_n)}|$ converges weakly to $|\phi_u|$. Since there are no sign changes on this diamond, (5-1) must hold throughout the subdiamond (by the intermediate value theorem); we thus conclude that (1-17) holds on any such subdiamond. Since the measure of all the degenerate subdiamonds is $O_\varepsilon(\delta)$, we conclude that (1-17) holds on D outside of a set of measure $O_\varepsilon(\delta)$. Letting $\delta \to 0$ we obtain the claim.

5.3. ***Piecewise smoothness.*** The only remaining property we need to verify is (ii). By spatial reflection symmetry (2-2), it suffices to show that for each $v \in [-T_0, T_0]$, the map $u \mapsto \phi\langle u, v\rangle$ is piecewise smooth on $[-T_0, T_0]$, with only finitely many pieces.

From (c), we know that $\phi_u^{(\mathrm{lin})}\langle u, v\rangle$ vanishes for u in a finite union of intervals and points in $[-T_0, T_0]$. On any one of these intervals, we know from (vi) that $\phi_u\langle u, v\rangle$ also vanishes almost everywhere, which by the Lipschitz nature of ϕ and the fundamental theorem of calculus ensures that $\phi\langle u, v\rangle$ is constant in u on each of these intervals, for any fixed v. So it will suffice to verify the piecewise smoothness of $u \mapsto \phi\langle u, v\rangle$ for any v and on any compact interval I of u for which $\phi_u^{(\mathrm{lin})}\langle u, v\rangle$ is bounded away from zero, so long as the number of pieces is bounded uniformly in I and v.

Fix I and v. By hypothesis, we can find $\varepsilon > 0$ such that $|\phi_u^{(\mathrm{lin})}\langle u, v\rangle| \geqslant \varepsilon$ for all $u \in I$; in particular, $\phi_u^{(\mathrm{lin})}$ does not change sign on this interval. By Lemma 4.5, we conclude for each n that

$$\left|\phi_u^{(p_n)}\langle u, v\rangle\right| = \left|\phi_u^{(\mathrm{lin})}\langle u, v\rangle\right| + O\left(\frac{\log^{1/2} p_n}{p_n^{1/2}}\right)$$

for $u \in I$ outside of $O(1)$ intervals of length $O_\varepsilon(\frac{\log p_n}{p_n})$ intersecting I.

By pigeonholing, we may assume that the number $k = O(1)$ of such intervals is constant; denoting the midpoints of these intervals by $u_1^{(p_n)} < \cdots < u_k^{(p_n)}$; without loss of generality we may take $u_1^{(p_n)}$ and $u_k^{(p_n)}$ to be the endpoints of I. We may assume from the Bolzano–Weierstrass theorem and passing to a further subsequence that each of the $u_j^{(p_n)}$ converge to some limit u_j.

Between $u_j^{(p_n)}$ and $u_{j+1}^{(p_n)}$, excluding those u lying within $O_\varepsilon(\frac{\log p_n}{p_n})$ of either endpoint, we may write

$$\phi_u^{(p_n)}\langle u, v\rangle = \epsilon_j^{(p_n)}\phi_u^{(\mathrm{lin})}\langle u, v\rangle + O\left(\frac{\log^{1/2} p_n}{p_n^{1/2}}\right),$$

where $\epsilon_j^{(p_n)} \in \{-1, +1\}$. By a further pigeonholing we may take $\epsilon_j^{(p_n)} = \epsilon_j$ independent of n. Using the fundamental theorem of calculus and then taking limits, we conclude that $\phi\langle u, v\rangle$ is piecewise smooth for $u \in I$, with possible discontinuities at $u_1, \ldots, u_k$, and with $\phi\langle u, v\rangle$ equal to $\epsilon_j\phi_u^{(\mathrm{lin})}\langle u, v\rangle$ on the interval (u_j, u_{j+1}) for any $1 \leqslant j < k$. The claim follows, and the proof of Theorem 1.3 is complete.

Acknowledgments

We are indebted to Tristan Roy for posing this question, and Rowan Killip for many useful discussions, Ut V. Le for pointing out a misprint and for the reference [Andreu et al. 2009], to Patrick Dorey for pointing out the link to Liouville's equation, and to the anonymous referee for many useful comments and corrections.

References

[Andreu et al. 2009] F. Andreu, J. M. Mazón, J. D. Rossi, and J. Toledo, "The limit as $p \to \infty$ in a nonlocal p-Laplacian evolution equation: a nonlocal approximation for sandpiles", *Calc. Var.* **35** (2009), 279–316.

[Kalyakin 2001] L. A. Kalyakin, "Liouville equation under perturbation", *Inverse Problems* **17**:4 (2001), 879–883. MR2002h:35197 Zbl 0982.35089

[Liouville 1853] J. Liouville, "Sur l'equation aux differences partielles $\partial^2 \ln \lambda/\partial u \partial v \pm 2\lambda q^2 = 0$", *J. Math. Pure Appl.* **18** (1853), 71–74.

[Perthame 2002] B. Perthame, *Kinetic formulation of conservation laws*, Oxford Lecture Series in Mathematics and its Applications **21**, Oxford University Press, Oxford, 2002. MR 2005d:35005 Zbl 1030.35002

[Sogge 1995] C. D. Sogge, *Lectures on nonlinear wave equations*, Monographs in Analysis, II, International Press, Boston, 1995. MR 2000g:35153 Zbl 1089.35500

[Tao $\geq$ 2009] T. Tao, "An explicitly solvable wave equation", preprint, Available at http://tinyurl.com/dd9o7r.

Received 22 Jan 2009. Revised 13 Feb 2009. Accepted 5 Apr 2009.

TERENCE TAO: tao@math.ucla.edu
Department of Mathematics, University of California, Los Angeles, Mathematics Department, Los Angeles, CA 90095-1555, United States
http://ftp.math.ucla.edu/~tao/

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at pjm.math.berkeley.edu/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@mathscipub.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 2 No. 2 2009