

# ANALYSIS & PDE

Volume 11      No. 8      2018

SCOTT ARMSTRONG, ALEXANDRE BORDAS  
AND JEAN-CHRISTOPHE MOURRAT

**QUANTITATIVE STOCHASTIC HOMOGENIZATION AND  
REGULARITY THEORY OF PARABOLIC EQUATIONS**





# QUANTITATIVE STOCHASTIC HOMOGENIZATION AND REGULARITY THEORY OF PARABOLIC EQUATIONS

SCOTT ARMSTRONG, ALEXANDRE BORDAS AND JEAN-CHRISTOPHE MOURRAT

We develop a quantitative theory of stochastic homogenization for linear, uniformly parabolic equations with coefficients depending on space and time. Inspired by recent works in the elliptic setting, our analysis is focused on certain subadditive quantities derived from a variational interpretation of parabolic equations. These subadditive quantities are intimately connected to spatial averages of the fluxes and gradients of solutions. We implement a renormalization-type scheme to obtain an algebraic rate for their convergence, which is essentially a quantification of the weak convergence of the gradients and fluxes of solutions to their homogenized limits. As a consequence, we obtain estimates of the homogenization error for the Cauchy–Dirichlet problem which are optimal in stochastic integrability. We also develop a higher regularity theory for solutions of the heterogeneous equation, including a uniform  $C^{0,1}$ -type estimate and a Liouville theorem of every finite order.

1. Introduction	1945
2. Variational structure and subadditive quantities	1954
3. Functional inequalities	1962
4. Convergence of subadditive quantities	1975
5. Quantitative homogenization of the Cauchy–Dirichlet problem	1988
6. Regularity theory	1999
Appendix A. Variational structure of uniformly parabolic equations	2000
Appendix B. Meyers-type estimates	2007
Acknowledgments	2013
References	2013

## 1. Introduction

**1A. Motivation and informal summary of results.** In this paper, we develop a quantitative theory of stochastic homogenization for linear, uniformly parabolic equations with coefficients depending on both the space and time variables. We consider equations of the form

$$\partial_t u^\varepsilon - \nabla \cdot \left( \mathbf{a} \left( \frac{t}{\varepsilon^2}, \frac{x}{\varepsilon} \right) \nabla u^\varepsilon \right) = 0 \quad \text{in } I \times U, \quad (1-1)$$

*MSC2010:* primary 35B27, 35B45; secondary 60K37, 60F05.

*Keywords:* stochastic homogenization, parabolic equation, large-scale regularity, variational methods.

where  $I \subseteq \mathbb{R}$  is an open interval,  $U$  is a bounded Lipschitz domain of  $\mathbb{R}^d$ , and  $(t, x) \mapsto \mathbf{a}(t, x)$  is a stationary random field taking values in the set of real  $d$ -by- $d$  matrices satisfying, for a fixed constant  $\Lambda \in [1, \infty)$ ,

$$\text{for all } \xi \in \mathbb{R}^d, \quad \xi \cdot \mathbf{a}(t, x)\xi \geq \Lambda^{-1}|\xi|^2 \quad \text{and} \quad |\mathbf{a}(t, x)\xi| \leq \Lambda|\xi|. \quad (1-2)$$

Here the symbol  $\nabla$  denotes the gradient in the space variables only; that is,  $\nabla w = (\partial_{x_1} w, \dots, \partial_{x_d} w)$ . We let  $\mathbb{P}$  be the law of the random field  $\mathbf{a}(t, x)$ , which we assume to be invariant under translations by elements of  $\mathbb{Z} \times \mathbb{Z}^d$  and to have a finite range of dependence. (See the following subsection for the precise assumptions.)

We are interested in the behavior of the solutions  $u^\varepsilon(t, x)$  for  $0 < \varepsilon \ll 1$ . It is well known that, under very general qualitative assumptions on the coefficients (stationarity and ergodicity), (1-1) homogenizes to an effective limiting equation of the form

$$\partial_t u - \nabla \cdot (\bar{\mathbf{a}} \nabla u) = 0 \quad \text{in } I \times U, \quad (1-3)$$

where  $\bar{\mathbf{a}}$  is a deterministic  $d$ -by- $d$  matrix. This principle can be formulated in various ways, but it means for example that the solutions  $u^\varepsilon$  of (1-1), subject to appropriate initial-boundary conditions, converge as  $\varepsilon \rightarrow 0$ ,  $\mathbb{P}$ -almost surely and in some appropriate function space, to solutions of the homogenized equation (1-3).<sup>1</sup> Such a result is usually proved by soft arguments, using an abstract version of the ergodic theorem, which unfortunately does not give *quantitative* information concerning the convergence.

There has been a lot of recent interest in quantitative stochastic homogenization for elliptic equations, particularly in the case of linear, uniformly elliptic equations. This essentially began with the work of Gloria and Otto [2011; 2012], who proved the first quantitative results which are optimal in the scaling of the parameter  $\varepsilon$ ; see also [Gloria et al. 2015]. Their work motivated a great number of subsequent works, and we refer to the recently completed [Armstrong et al. 2017b] for more background, references and historical information.

In this paper, motivated by the desire to obtain quantitative homogenization results — in particular, explicit estimates of the homogenization error — we develop an analytic approach for parabolic equations with random coefficients based on the ideas recently introduced in [Armstrong and Smart 2016; Armstrong and Mourrat 2016; Armstrong et al. 2016; 2017a], which are perhaps best presented in [Armstrong et al. 2017b]. Those papers developed a rather complete quantitative theory of elliptic homogenization starting from the observation that certain energy quantities — which are very natural from a variational perspective — are also rather convenient for studying the homogenization process. This is because: (i) they efficiently encode information about the weak convergence of the fluxes, gradients, and energy densities of solutions; and (ii) they are amenable to renormalization arguments in the sense that we can obtain rates of convergence for the quantities by iterating the length scale. This variational approach allows one to circumvent the need for nonlinear concentration inequalities, because it reveals a “linear” structure of the randomness: while the solutions are very nonlinear functions of the coefficients, the energy quantities

<sup>1</sup>We remark that we are unaware of a reference which proves this specific result in the parabolic setting. Nevertheless, we maintain that it is essentially well known, since the classical qualitative proof given in the elliptic case, see for instance [Papanicolaou and Varadhan 1981; Bensoussan et al. 1978; Jikov et al. 1994], can be straightforwardly generalized to the parabolic setting.

turn out to be essentially linear. This observation greatly simplifies the theory and allows one to derive estimates which are optimal both in the scaling of  $\varepsilon$  as well as in stochastic integrability. A related approach inspired by [Armstrong and Smart 2016; Armstrong and Mourrat 2016; Armstrong et al. 2016; 2017a] has also recently been developed in [Gloria and Otto 2015; 2016].

The two main results of this paper are (i) a quantitative estimate on the homogenization error for Cauchy–Dirichlet problems (Theorem 1.1) and (ii) a complete large-scale regularity theory (Theorem 1.2). It has already been observed in the elliptic case, see [Armstrong et al. 2017b], that results of this type are the first step towards optimal quantitative estimates and scaling limits for first-order correctors as well as optimal error estimates for boundary-value problems. At the same time, the results in this paper are the first quantitative stochastic homogenization results, to our knowledge, for parabolic equations with coefficients with space-time dependence.

The starting point for adapting the techniques of [Armstrong et al. 2017b] to the parabolic case is a variational characterization of divergence-form parabolic equations that was first discovered by Brezis and Ekeland [1976a; 1976b]. We give a self-contained presentation of this characterization in Appendix A, where we also give a convex analytic proof of the well-posedness of general Cauchy–Dirichlet problems inspired by [Ghoussoub and Tzou 2004]. Based on this variational principle, we introduce subadditive quantities for the homogenization problem in Section 2 and adapt the methods of [Armstrong et al. 2017b], using an iteration of scales and a renormalization-type argument, to obtain an algebraic rate of convergence in Section 4. Compared to the elliptic case, the main sources of additional difficulty in the iteration argument have to do with the need to control certain weak Sobolev norms of the time derivatives of the solutions. We accomplish this with the help of some functional inequalities we prove in Section 3. In Section 5, we show that the convergence of the subadditive quantities gives us approximate first-order correctors with good quantitative bounds, which allows us to prove Theorem 1.1. In the last section, we obtain the regularity result of Theorem 1.2. In the rest of this introduction, we state the assumptions, notation and main results.

**1B. Assumptions.** We fix a spatial dimension  $d \geq 2$  and a parameter  $\Lambda \in [1, \infty)$ . We let  $\Omega$  denote the set of all possible coefficient fields  $\mathbf{a}(t, x)$ , which are assumed to be measurable maps from  $\mathbb{R} \times \mathbb{R}^d$  into the set  $\Omega_0$  of matrices  $\mathbf{a}$  satisfying

$$\text{for all } \xi \in \mathbb{R}^d, \quad \xi \cdot \mathbf{a}\xi \geq \Lambda^{-1}|\xi|^2 \quad \text{and} \quad |\mathbf{a}\xi| \leq \Lambda|\xi|. \tag{1-4}$$

That is, we define

$$\Omega_0 := \{\mathbf{a} \in \mathbb{R}^{d \times d} : \mathbf{a} \text{ satisfies (1-4)}\}$$

and then set

$$\Omega := \{\mathbf{a} : \mathbb{R} \times \mathbb{R}^d \rightarrow \Omega_0 \text{ is Lebesgue-measurable}\}. \tag{1-5}$$

For every Borel subset  $V \subseteq \mathbb{R} \times \mathbb{R}^d$ , we define  $\mathcal{F}_V$  to be the  $\sigma$ -algebra representing the information obtaining by observing the coefficient field in  $V$ . Formally,

$$\begin{aligned} \mathcal{F}_V &:= \text{the } \sigma\text{-algebra generated by the random elements} \\ \mathbf{a} &\mapsto \int_V \varphi(t, x) \mathbf{a}(t, x) dt dx, \quad \varphi \in C_c^\infty(\mathbb{R} \times \mathbb{R}^d). \end{aligned} \tag{1-6}$$

The largest of the  $\sigma$ -algebras in this family is  $\mathcal{F} := \mathcal{F}_{\mathbb{R} \times \mathbb{R}^d}$ . We assume that  $\mathbb{P}$  is a given probability measure on the measurable space  $(\Omega, \mathcal{F})$  which satisfies the following two assumptions:

(P1)  $\mathbb{P}$  is *stationary* with respect to  $\mathbb{Z} \times \mathbb{Z}^d$ -translations. For every  $z \in \mathbb{Z} \times \mathbb{Z}^d$  and event  $A \in \mathcal{F}$ ,

$$\mathbb{P}[A] = \mathbb{P}[T_z A].$$

(P2)  $\mathbb{P}$  has a *unit range of dependence*. For every pair of Borel subsets  $U, V \subseteq \mathbb{R} \times \mathbb{R}^d$ ,

$$\text{dist}(U, V) \geq 1 \implies \mathcal{F}_U \text{ and } \mathcal{F}_V \text{ are } \mathbb{P}\text{-independent.}$$

Here “dist” is defined with respect to the usual Euclidean distance on  $\mathbb{R} \times \mathbb{R}^d$ . We denote by  $\mathbb{E}[X]$  the expectation of an  $\mathcal{F}$ -measurable random variable  $X$  with respect to  $\mathbb{P}$ . While we assume that the coefficient field has a finite range of dependence for simplicity, we point out that this hypothesis can be weakened using arguments similar to those exposed in [Armstrong and Mourrat 2016].

**1C. Notation.** We unfortunately must introduce quite a bit of notation, particularly since we are working with parabolic equations which require us to define various function spaces. We collect the notation needed in this subsection, which the reader is encouraged to skim and consult as a reference.

*General notation.* We denote the set of natural numbers by  $\mathbb{N} := \{0, 1, 2, \dots\}$ . We use the symbols  $\wedge$  and  $\vee$  to denote minimum and maximum, respectively; for example  $r \wedge s = \min\{r, s\}$  for  $r, s \in \mathbb{R}$ . For every  $r \in \mathbb{R}$ , we also define  $r_+ := r \vee 0$  and  $r_- := r \wedge 0$ . For any  $m \in \mathbb{N}$  and measurable subset  $E \subseteq \mathbb{R}^m$ , the Lebesgue measure of  $E$  is denoted by  $|E|$ , unless  $E$  is a finite set, in which case  $|E|$  is the cardinality of  $E$ . This is often used for  $m \in \{1, d, 1+d\}$ . A slash through the integral denotes normalization by the Lebesgue measure:  $f_E := (1/|E|) \int_E$ . The mean of a function  $f \in L^1(E)$  is also denoted by  $(f)_E := f_E f$ .

A *parabolic cylinder* is any set of the form  $I \times U$  where  $I = (I_-, I_+) \subseteq \mathbb{R}$  is a bounded open interval and  $U \subseteq \mathbb{R}^d$  is a bounded Lipschitz domain. We denote the parabolic boundary of  $I \times U$  by

$$\partial_{\sqcup}(I \times U) := (I \times \partial U) \cup (\{I_-\} \times U).$$

We denote the Euclidean ball of  $\mathbb{R}^d$  of radius  $r \in (0, \infty]$  centered at  $x \in \mathbb{R}^d$  by  $B_r(x)$ , and put  $B_r := B_r(0)$ . Throughout, we work with the triadic cubes defined for every  $n \in (0, \infty)$  by

$$I_n := \left(-\frac{3^{2n}}{2}, \frac{3^{2n}}{2}\right), \quad \square_n := \left(-\frac{3^n}{2}, \frac{3^n}{2}\right)^d, \quad \square_n := I_n \times \square_n.$$

Note that the parabolic cylinder  $\square_n$  is evidently not a cube per se since its sides have a scaling which matches the parabolic scaling. However, we note that for each  $m, n \in \mathbb{N}$  with  $m < n$ , we can write  $\square_n$  as the disjoint union (up to a set of Lebesgue measure zero) of exactly  $3^{(2+d)(m-n)}$  cubes of the form  $z + \square_m$  with  $z \in 3^{2m}\mathbb{Z} \times 3^m\mathbb{Z}$ .

We also use the following notation for parabolic cylinders: for each  $r \in (0, \infty]$  and  $(t, x) \in \mathbb{R} \times \mathbb{R}^d$ , we define

$$\tilde{I}_r := (-r^2, 0], \quad Q_r(t, x) := (t, x) + \tilde{I}_r \times B_r, \quad \text{and} \quad Q_r := Q_r(0, 0). \tag{1-7}$$

*Function spaces.* For every bounded Lipschitz domain  $U \subseteq \mathbb{R}^d$  with  $|U| < \infty$  and  $p \in [1, \infty)$ , we denote the normalized  $L^p(U)$  norm of a function  $f \in L^p(U)$  by

$$\|f\|_{\underline{L}^p(U)} := \left( \int_U |f|^p \right)^{\frac{1}{p}} = |U|^{-\frac{1}{p}} \|f\|_{L^p(U)}. \quad (1-8)$$

For  $p = \infty$ , we define  $\|f\|_{\underline{L}^\infty(U)} := \|f\|_{L^\infty(U)}$ . We use similar notation to denote normalized (scale-invariant) Sobolev norms: for every  $p \in [1, \infty)$  and  $f \in W^{1,p}(U)$ ,

$$\|f\|_{\underline{W}^{1,p}(U)} := |U|^{-\frac{1}{d}} \|f\|_{\underline{L}^p(U)} + \|\nabla f\|_{\underline{L}^p(U)}.$$

In the case  $p = 2$  we use the notation  $\|f\|_{\underline{H}^1(U)} := \|f\|_{\underline{W}^{1,2}(U)}$ . As usual,  $H_0^1(U)$  and  $W_0^{1,p}(U)$  respectively denote the closure in  $H^1(U)$  and  $W^{1,p}(U)$ , respectively, of the compactly supported smooth functions in  $U$ . The dual spaces to  $W^{1,p}(U)$  and  $W_0^{1,p}(U)$  are denoted by  $\widehat{W}^{-1,p'}(U)$  and  $W^{-1,p'}(U)$ , respectively, where  $p' := p/(p-1)$  is the Hölder conjugate exponent of  $p$ . The normalized, scale-invariant dual norms are respectively defined by

$$\begin{aligned} \|v\|_{\widehat{W}^{-1,p'}(U)} &:= \sup \left\{ \int_U uv : u \in W^{1,p}(U), \|u\|_{\underline{W}^{1,p}(U)} \leq 1 \right\}, \\ \|v\|_{W^{-1,p'}(U)} &:= \sup \left\{ \int_U uv : u \in W_0^{1,p}(U), \|u\|_{\underline{W}^{1,p}(U)} \leq 1 \right\}. \end{aligned}$$

Here we are abusing notation by denoting the natural pairing  $\langle u, w \rangle$  between the two dual spaces (up to a constant) by the normalized integral. This is done to emphasize the normalization that we wish to enforce, which extends the action of an element  $w \in C_c^\infty(U)$  on  $W^{1,p}(U)$  by  $u \mapsto \int_U uw$ . For  $p = 2$ , we also write  $\|\cdot\|_{\widehat{H}^{-1}(U)} := \|v\|_{\widehat{W}^{-1,2}(U)}$  and  $\|\cdot\|_{\underline{H}^{-1}(U)} := \|v\|_{\underline{W}^{-1,2}(U)}$ .

We next introduce function spaces designed for parabolic equations. For each  $n \in \mathbb{N}$ , bounded Lipschitz domain  $U \subseteq \mathbb{R}^n$ , Banach space  $X$  and  $p \in [1, \infty)$ , we denote by  $L^p(U; X)$  the space of Lebesgue-measurable mappings  $u : U \rightarrow X$  such that

$$\|u\|_{\underline{L}^p(U; X)} := \left( \int_U \|u(x)\|_X^p dx \right)^{\frac{1}{p}} < \infty.$$

For every interval  $I = (I_-, I_+) \subseteq \mathbb{R}$  and bounded Lipschitz domain  $U \subseteq \mathbb{R}^d$ , we define the function space

$$H_{\text{par}}^1(I \times U) := \{u \in L^2(I; H^1(U)) : \partial_t u \in L^2(I; H^{-1}(U))\}, \quad (1-9)$$

which is the closure of bounded smooth functions on  $I \times U$  with respect to the norm

$$\|u\|_{\underline{H}_{\text{par}}^1(I \times U)} := \|u\|_{\underline{L}^2(I; \underline{H}^1(U))} + \|\partial_t u\|_{\underline{L}^2(I; \underline{H}^{-1}(U))}. \quad (1-10)$$

We denote by  $H_{\text{par}, \sqcup}^1(I \times U)$  the closure in  $H_{\text{par}}^1(I \times U)$  of the set of smooth functions with compact support in  $(I_-, I_+) \times U$ . In other words, a function in  $H_{\text{par}, \sqcup}^1(I \times U)$  has zero trace on the lateral boundary  $I \times \partial U$  and the initial time  $\{I_-\} \times U$  but does not necessarily vanish at the final time.

We let  $H_{\text{par},0}^1(I \times U)$  denote the completion of the set of smooth functions with compact support in  $I \times U$  with respect to the norm

$$\|v\|_{\underline{H}_{\text{par},0}^1(I \times U)} := \|v\|_{\underline{L}^2(I; \underline{H}^1(U))} + \|\partial_t v\|_{\underline{L}^2(I; \widehat{H}^{-1}(U))}. \tag{1-11}$$

Note that compared with (1-10), here we require the time derivative  $\partial_t v(t, \cdot)$  to be an element of  $\widehat{H}^{-1}(U)$  instead of  $H^{-1}(U)$ . In particular, for  $v \in H_{\text{par},0}^1(I \times U)$ , the spatial average of  $\partial_t v$  over  $I \times U$  is well-defined, since constant functions belong to  $\underline{L}^2(I; H^1(U))$ , while they do not belong to  $L^2(I; H_0^1(U))$ . Moreover, the boundary condition imposes that for every  $v \in H_{\text{par},0}^1(I \times U)$ ,

$$\int_{I \times U} \partial_t v = 0. \tag{1-12}$$

This identity is indeed clear if  $v$  is smooth and compactly supported in  $I \times U$ , and we can then obtain the general case by density.

In certain situations, it is useful to work with variations of  $H_{\text{par}}^1(I \times U)$  in which the exponent of integrability is  $p \in (1, \infty)$  rather than 2. So we also define the function space

$$W_{\text{par}}^{1,p}(I \times U) := \{u \in L^p(I; W^{1,p}(U)) : \partial_t u \in L^p(I; W^{-1,p}(U))\}, \tag{1-13}$$

which is the closure of bounded smooth functions on  $I \times U$  with respect to the norm

$$\|u\|_{\underline{W}_{\text{par}}^{1,p}(I \times U)} := \|u\|_{\underline{L}^p(I; W^{1,p}(U))} + \|\partial_t u\|_{\underline{L}^p(I; W^{-1,p}(U))}. \tag{1-14}$$

Similarly to  $H_{\text{par},\sqcup}^1(I \times U)$ , we denote by  $W_{\text{par},\sqcup}^{1,p}(I \times U)$  the closure in  $W_{\text{par}}^{1,p}(I \times U)$  of the set of smooth functions with compact support in  $(I_-, I_+) \times U$ . Finally, for every parabolic cylinder  $V$ , we denote by  $W_{\text{par},\text{loc}}^{1,p}(V)$ ,  $H_{\text{par},\text{loc}}^1(V)$ , and so forth, the functions on  $V$  which are, respectively, elements of  $W_{\text{par}}^{1,p}(W)$  and  $H_{\text{par}}^1(W)$ , etc., for every subcylinder  $W \subseteq V$  with  $\overline{W} \subseteq V$ .

We next turn to the definitions of the negative parabolic Sobolev spaces. We denote by  $\widehat{H}_{\text{par}}^{-1}(V)$  and  $H_{\text{par}}^{-1}(V)$  the dual spaces to  $H_{\text{par}}^1(V)$  and  $H_{\text{par},\sqcup}^1(V)$ , respectively, with (normalized, scale-invariant) dual norms given by

$$\begin{aligned} \|f\|_{\widehat{H}_{\text{par}}^{-1}(V)} &:= \sup \left\{ \int_V f w : w \in H_{\text{par}}^1(V), \|w\|_{\underline{H}_{\text{par}}^1(V)} \leq 1 \right\}, \\ \|f\|_{\underline{H}_{\text{par}}^{-1}(V)} &:= \sup \left\{ \int_V f w : w \in H_{\text{par},\sqcup}^1(V), \|w\|_{\underline{H}_{\text{par}}^1(V)} \leq 1 \right\}. \end{aligned} \tag{1-15}$$

As explained above, the notation  $\int_V f w$  should be interpreted as the canonical pairing between  $f \in \widehat{H}_{\text{par}}^{-1}(V)$  or  $f \in H_{\text{par}}^{-1}(V)$ , respectively, and  $w \in H_{\text{par}}^1(V)$  or  $w \in H_{\text{par},\sqcup}^1(V)$ , which extends the action of bounded smooth functions on  $H_{\text{par}}^1(V)$  or  $H_{\text{par},\sqcup}^1(V)$ . We similarly define the space  $W_{\text{par}}^{-1,p}(V)$  to be the dual space of the Banach space  $W_{\text{par},\sqcup}^{1,p'}(V)$ , where  $p' := p/(p - 1)$ , and endow it with the (normalized, scale-invariant) norm

$$\|f\|_{\underline{W}_{\text{par}}^{-1,p}(V)} := \sup \left\{ \int_V f w : w \in W_{\text{par},\sqcup}^{1,p'}(V), \|w\|_{\underline{W}_{\text{par}}^{1,p'}(V)} \leq 1 \right\}. \tag{1-16}$$



Recall that negative Sobolev norms arise naturally when one wishes to quantify *weak* convergence in  $L^p$  or positive Sobolev spaces; see [Armstrong et al. 2017b, Section 1.4]. This is indeed their purpose in this paper.

*The  $\mathcal{O}_s$  notation.* Since the random variables we encounter in this paper are very often the sum of a deterministic quantity and a “small” random part, it is useful to work with the notation introduced in [Armstrong et al. 2017a] for expressing the sizes of random variables (essentially, an alternative notation for certain Orlicz norms). It is intended to remind us of “big- $O$ ” notation and is convenient because it compresses some of our computations and makes our inequalities easier to understand at a glance.

If  $X$  is a random variable and  $s, k \in (0, \infty)$ , then we write

$$X \leq \mathcal{O}_s(k)$$

as a shorthand for the statement

$$\mathbb{E} \left[ \exp \left( \left( \frac{X_+}{k} \right)^s \right) \right] \leq 2. \tag{1-17}$$

Roughly, this means that “ $X$  is of order  $k$  with stretched exponential tails with exponent  $s$ .” More precisely, we can use Chebyshev’s inequality to see that

$$X \leq \mathcal{O}_s(k) \implies \text{for all } \lambda > 0, \mathbb{P}[X > \lambda k] \leq 2 \exp(-\lambda^s). \tag{1-18}$$

The converse of this statement is almost true: for every  $k \geq 0$ ,

$$\text{for all } \lambda \geq 0, \mathbb{P}[X \geq \lambda k] \leq \exp(-\lambda^s) \implies X \leq \mathcal{O}_s(2^{\frac{1}{s}} \theta). \tag{1-19}$$

This can be obtained by integration. We also use the notation

$$X = \mathcal{O}_s(k) \iff X \leq \mathcal{O}_s(k) \text{ and } -X \leq \mathcal{O}_s(k).$$

Similarly, we write  $X \leq Y + \mathcal{O}_s(k)$  to mean that  $X - Y \leq \mathcal{O}_s(k)$  and  $X = Y + \mathcal{O}_s(k)$  to mean that  $X - Y = \mathcal{O}_s(k)$ . If  $s \in [1, \infty)$ , then Jensen’s inequality gives us a triangle inequality for  $\mathcal{O}_s(\cdot)$  in the following sense: for any measure space  $(E, \mathcal{S}, \mu)$ , measurable function  $K : E \rightarrow (0, \infty)$  and jointly measurable family  $\{X(z)\}_{z \in E}$  of nonnegative random variables, we have

$$\text{for all } z \in E, X(z) \leq \mathcal{O}_s(K(z)) \implies \int_E X \, d\mu \leq \mathcal{O}_s \left( \int_E K \, d\mu \right). \tag{1-20}$$

If  $s \in (0, 1]$ , then the statement is true after adding a prefactor constant  $C_s > 1$  to the right side. For a proof of (1-18)–(1-20), see [Armstrong et al. 2017b, Appendix A].

**1D. Statement of the main results.** We present two main results. The first provides an algebraic convergence rate for the homogenization limit of the Cauchy–Dirichlet initial-value problem in a parabolic cylinder  $I \times U$ , where  $U \subseteq \mathbb{R}^d$  is a bounded Lipschitz domain. This is a parabolic counterpart of a theorem proved in the elliptic setting in [Armstrong and Smart 2016]; see also [Armstrong et al. 2017b, Theorem 2.16].

**Theorem 1.1.** Fix  $s \in (0, 2 + d)$ , a bounded Lipschitz domain  $U \subseteq B_1$ , an interval  $I := (I_-, 0) \subseteq (-\frac{1}{4}, 0)$  and an exponent  $\delta > 0$ . Put  $V := I \times U$ . There exist an exponent  $\beta(\delta, V, d, \Lambda) > 0$ , a constant  $C(s, V, \delta, d, \Lambda) < \infty$  and a random variable  $\mathcal{X}$  satisfying

$$\mathcal{X} = \mathcal{O}_1(C)$$

such that the following convergence result holds: for each  $\varepsilon \in (0, \frac{1}{2}]$  and initial-boundary condition  $f \in W_{\text{par}}^{1,2+\delta}(V)$ , defining

$$\mathbf{a}^\varepsilon(t, x) := \mathbf{a}\left(\frac{t}{\varepsilon^2}, \frac{x}{\varepsilon}\right)$$

and taking  $u^\varepsilon, u \in f + H_{\text{par},\square}^1(V)$  to be the solutions of the Cauchy–Dirichlet problems

$$\begin{cases} \partial_t u^\varepsilon - \nabla \cdot (\mathbf{a}^\varepsilon \nabla u^\varepsilon) = 0 & \text{in } V, \\ u^\varepsilon = f & \text{on } \partial_\square V, \end{cases} \quad \text{and} \quad \begin{cases} \partial_t u - \nabla \cdot (\bar{\mathbf{a}} \nabla u) = 0 & \text{in } V, \\ u = f & \text{on } \partial_\square V, \end{cases} \quad (1-21)$$

we have the estimate

$$\|\nabla u^\varepsilon - \nabla u\|_{\widehat{H}_{\text{par}}^{-1}(V)} + \|\mathbf{a}^\varepsilon \nabla u^\varepsilon - \bar{\mathbf{a}} \nabla u\|_{\widehat{H}_{\text{par}}^{-1}(V)} + \|u^\varepsilon - u\|_{L^2(V)} \leq C \|\nabla f\|_{W_{\text{par}}^{1,2+\delta}(V)} (\varepsilon^{\beta(2+d-s)} + \mathcal{X} \varepsilon^s). \quad (1-22)$$

As well as estimating the homogenization error, notice that the estimate (1-22) quantifies the weak convergence in  $L^2(V)$  of the gradients and fluxes of  $u^\varepsilon$  to those of  $u$ . The random part of the error, namely  $\mathcal{X} \varepsilon^s$  for an  $s$  arbitrarily close to  $2 + d$ , is very small compared to the deterministic part,  $\varepsilon^{\beta(2+d-s)}$ . It is also important for applications to observe that  $\mathcal{X}$  is independent of the initial-boundary condition  $f$ .

On the right side of (1-22), we have split the error into a possibly rather large deterministic part (large, since we do not control the smallness of  $\beta > 0$ ) plus a random error. While the typical size of the error is estimated suboptimally, since  $\beta > 0$  is small, the *tail behavior* of this random part is sharply estimated. In particular, we see that the probability for the term  $(\varepsilon^{\beta(2+d-s)} + \mathcal{X} \varepsilon^s)$  to be  $O(1)$  is smaller than  $\exp(-\varepsilon^{-s}/C)$ , for arbitrary  $s < 2 + d$ . This estimate is sharp, in the sense that it would be false for any  $s > 2 + d$ . We refer to [Armstrong et al. 2017b, Remark 2.5 and Section 3.5] for similar considerations in the elliptic setting.

The second theorem we present here is a large-scale regularity result, a parabolic counterpart to [Armstrong et al. 2017b, Theorem 3.6]. In particular, we seek to classify all ancient solutions of the parabolic equation which exhibit at most polynomial growth at infinity and backwards in time. This requires us to introduce some additional notation.

We denote polynomials in the variables  $t, x_1, \dots, x_d$  by  $\mathcal{P}(\mathbb{R} \times \mathbb{R}^d)$ . The *parabolic degree*  $\text{deg}_p(w)$  of an element  $w \in \mathcal{P}(\mathbb{R} \times \mathbb{R}^d)$  is the degree of the polynomial  $(t, x) \mapsto w(t^2, x)$ . For each  $k \in \mathbb{N}$  we let  $\mathcal{P}_k(\mathbb{R} \times \mathbb{R}^d)$  be the subset of  $\mathcal{P}(\mathbb{R} \times \mathbb{R}^d)$  of polynomials with parabolic degree at most  $k$ . For  $\alpha > 0$ , we say that a function  $\phi : \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}$  or  $\phi : (-\infty, 0] \times \mathbb{R}^d \rightarrow \mathbb{R}$  is *parabolically  $\alpha$ -homogeneous* if

$$\text{for all } \lambda \in \mathbb{R}, \quad \phi(\lambda^2 t, \lambda x) = \lambda^\alpha \phi(t, x).$$

Any element of  $\mathcal{P}_k(\mathbb{R} \times \mathbb{R}^d)$  can be written as a sum of at most  $C(d, k) < \infty$  many parabolically homogeneous polynomials.

We denote by  $\bar{\mathcal{A}}_k(Q_\infty)$  the set of  $\bar{\mathbf{a}}$ -caloric functions on  $Q_\infty$  with growth which is strictly less than a polynomial of parabolic degree  $k + 1$ :

$$\bar{\mathcal{A}}_k(Q_\infty) := \left\{ w \in H_{\text{par}}^1(Q_\infty) : \limsup_{r \rightarrow \infty} r^{-(k+1)} \|w\|_{\underline{L}^2(Q_r)} = 0, \partial_t w - \nabla \cdot (\bar{\mathbf{a}} \nabla w) = 0 \text{ in } Q_\infty \right\}.$$

It turns out that  $\bar{\mathcal{A}}_k(Q_\infty)$  coincides with the set of  $\bar{\mathbf{a}}$ -caloric polynomials<sup>2</sup> of parabolic degree at most  $k$ . That is,

$$\bar{\mathcal{A}}_k(Q_\infty) = \left\{ w|_{Q_\infty} : w \in \mathcal{P}_k(\mathbb{R} \times \mathbb{R}^d), \partial_t w - \nabla \cdot (\bar{\mathbf{a}} \nabla w) = 0 \text{ in } \mathbb{R} \times \mathbb{R}^d \right\}. \quad (1-23)$$

The vector space of  $n$ -homogeneous  $\bar{\mathbf{a}}$ -caloric polynomials is isomorphic to that of  $n$ -homogeneous polynomials of  $\mathbb{R}^d$ . This can be shown by backwards uniqueness and the fact that this vector space is spanned by products of homogeneous  $\bar{\mathbf{a}}$ -caloric polynomials depending only on  $t$  and one of the space variables; see for instance [Widder 1961] or [Nualart 2006, Proposition 1.1.1]. In any case, we have that  $\dim(\bar{\mathcal{A}}_k(Q_\infty)) = \binom{d+k}{d} < \infty$ .

In the next result, we generalize the parabolic Liouville theorem implicit in (1-23) to  $\mathbf{a}(x)$ -caloric functions. At the same time we provide a *quantitative* version of this Liouville principle, in other words, a  $C^{k,1}$ -type regularity estimate. Define, for every parabolic cylinder  $I \times U \subseteq \mathbb{R} \times \mathbb{R}^d$ ,

$$\mathcal{A}(I \times U) := \left\{ w \in H_{\text{par, loc}}^1(I \times U) : \partial_t w - \nabla \cdot (\mathbf{a} \nabla w) = 0 \text{ in } I \times U \right\}$$

and, for every  $k \in \mathbb{N}$ ,

$$\mathcal{A}_k(Q_\infty) := \left\{ w \in \mathcal{A}(Q_\infty) : \limsup_{r \rightarrow \infty} r^{-(k+1)} \|w\|_{\underline{L}^2(Q_r)} = 0 \right\}.$$

Note that these vector spaces are *random* since they depend on  $\mathbf{a} \in \Omega$ . The following theorem is a parabolic analogue of [Armstrong et al. 2017b, Theorem 3.6].

**Theorem 1.2** (parabolic higher regularity theory). *Fix  $s \in (0, 2 + d)$ . There exist an exponent  $\delta(s, d, \Lambda) \in (0, \frac{1}{2}]$  and a random variable  $\mathcal{X}_s$  satisfying the estimate*

$$\mathcal{X}_s \leq \mathcal{O}_s(C(s, d, \Lambda)) \quad (1-24)$$

such that the following statements hold, for every  $k \in \mathbb{N}$ :

(i)<sub>k</sub> *There exists  $C(k, d, \Lambda) < \infty$  such that, for every  $u \in \mathcal{A}_k(Q_\infty)$ , there exists  $p \in \bar{\mathcal{A}}_k(Q_\infty)$  such that for every  $R \geq \mathcal{X}_s$ ,*

$$\|u - p\|_{\underline{L}^2(Q_R)} \leq CR^{-\delta} \|p\|_{\underline{L}^2(Q_R)}. \quad (1-25)$$

(ii)<sub>k</sub> *For every  $p \in \bar{\mathcal{A}}_k(Q_\infty)$ , there exists  $u \in \mathcal{A}_k(Q_\infty)$  satisfying (1-25) for every  $R \geq \mathcal{X}_s$ .*

(iii)<sub>k</sub> *There exists  $C(k, d, \Lambda) < \infty$  such that, for every  $R \geq \mathcal{X}_s$  and  $u \in \mathcal{A}(Q_R)$ , there exists  $\phi \in \mathcal{A}_k(Q_\infty)$  such that, for every  $r \in [\mathcal{X}_s, R]$ , we have the estimate*

$$\|u - \phi\|_{\underline{L}^2(Q_r)} \leq C \left( \frac{r}{R} \right)^{k+1} \|u\|_{\underline{L}^2(Q_R)}. \quad (1-26)$$

<sup>2</sup> $\bar{\mathbf{a}}$ -caloric polynomials are often called *heat polynomials* in the literature, in the case  $\bar{\mathbf{a}} = I_d$ .

In particular, we have,  $\mathbb{P}$ -almost surely, for every  $k \in \mathbb{N}$ ,

$$\dim(\mathcal{A}_k(Q_\infty)) = \dim(\bar{\mathcal{A}}_k(Q_\infty)) = \binom{d+k}{d}. \quad (1-27)$$

Observe that, as in the elliptic case, even for  $k = 0$  the third statement of Theorem 1.2 gives us an important gradient estimate on solutions. Indeed, the combination of statement (iii)<sub>0</sub> and the Caccioppoli inequality yields that, for every  $R \geq \mathcal{X}_s$ ,  $u \in \mathcal{A}(Q_R)$  and  $r \in [\mathcal{X}_s, R]$ , we have

$$\|\nabla u\|_{\underline{L}^2(Q_r)} \leq C \|\nabla u\|_{\underline{L}^2(Q_R)}.$$

This should be seen as a  $C^{0,1}$ -type estimate and compared to pointwise gradient bounds for the solutions of the heat equation.

The proof of Theorem 1.2 is obtained as a consequence of Theorem 1.1 and a routine adaptation of the proof of [Armstrong et al. 2017b, Theorem 3.6], which is the statement of the analogous result in the elliptic case. In Section 6, we explain the modifications required in the parabolic setting.

Soon after the first version of this paper was submitted and posted to the arXiv, a new preprint of Bella, Chiarini and Fehrman [Bella et al. 2018] appeared which contains a large-scale regularity result which has some overlap with Theorem 1.2. In particular, under qualitative assumptions, they obtain the statement of Theorem 1.2 in the case  $k = 1$  with the estimate (1-24) on  $\mathcal{X}_s$  replaced by the qualitative bound  $\mathbb{P}[\mathcal{X}_s < \infty] = 1$ .

**1E. Outline of the paper.** In the next section, we introduce the subadditive quantities inherited from the variational structure of the equation and record some of their basic properties. For convenience, the variational formulation of uniformly parabolic equations is recalled in a self-contained presentation in Appendix A. In Section 3, we present several functional inequalities which are needed later in the paper. Of particular interest are inequalities giving us control of certain weak norms of functions in terms of the spatial averages of the functions in cubes as well as Caccioppoli-type inequalities giving us control of strong norms of solutions in terms of weak norms. Section 4 is the heart of the paper, where we prove the convergence of the subadditive quantities by an iteration over the length scales. In Section 5, we demonstrate how to pass from control of the convergence of the subadditive quantities to general homogenization results. Finally, in Section 6 we summarize the passage from the quantitative homogenization results to the higher regularity theory (which is entirely analogous to the elliptic setting). In Appendix B, we give local and global versions of the Meyers higher integrability estimate for gradients of solutions. We remark that the statement of the global Meyers estimate we prove appears to be new and somewhat sharper compared to what has previously appeared in the literature.

## 2. Variational structure and subadditive quantities

**2A. Variational formulation of parabolic equations.** As we now explain, the solution of a parabolic equation can be obtained as the minimizer of a uniformly convex functional. This is an entirely deterministic statement, valid for an arbitrary fixed coefficient field  $\mathbf{a} \in \Omega$ .

The following proposition states the solvability of parabolic equations. It relies on convex analysis and calculus of variations, and is close to the main result of [Ghoussoub and Tzou 2004]; see also [Ghoussoub



2009]. We provide a self-contained proof in Appendix A in the more general setting of maximal monotone operators, and for a larger set of pairs  $(w, w^*)$ ; see Proposition A.1.

**Proposition 2.1** (parabolic variational principle). *Let  $\mathcal{J}$  be defined below in (2-4). For each  $w \in H_{\text{par}}^1(I \times U)$  and  $w^* \in L^2(I; H^{-1}(U))$ , the mapping*

$$w + H_{\text{par}, \sqcup}^1(I \times U) \rightarrow \mathbb{R},$$

$$u \mapsto \mathcal{J}[u, w^*],$$

*is uniformly convex. Moreover, its minimum is zero, and the associated minimizer is the unique  $u \in w + H_{\text{par}, \sqcup}^1(I \times U)$  that is a solution of*

$$(\partial_t - \nabla \cdot \mathbf{a} \nabla)u = w^* \quad \text{in } I \times U. \tag{2-1}$$

Equation (2-1) is interpreted as

$$\text{for all } \phi \in L^2(I; H_0^1(U)), \quad \int_{I \times U} \nabla \phi \cdot \mathbf{a} \nabla u = \int_{I \times U} \phi (w^* - \partial_t u). \tag{2-2}$$

The left side of (2-2) can be more explicitly written as

$$\int_{I \times U} \nabla \phi(t, x) \cdot \mathbf{a}(t, x) \nabla u(t, x) dt dx,$$

while the right side of (2-2) could be more properly written as

$$\int_I \langle \phi(t, \cdot), (w^* - \partial_t u)(t, \cdot) \rangle dt,$$

with  $\langle \cdot, \cdot \rangle$  the duality pairing between  $H_0^1(U)$  and  $H^{-1}(U)$ .

We proceed to define the functional  $\mathcal{J}$  appearing in Proposition 2.1. To start, we decompose the matrix  $\mathbf{a}$  into its symmetric and skew-symmetric parts:

$$s(t, x) := \frac{\mathbf{a}(t, x) + \mathbf{a}^t(t, x)}{2}, \quad \mathbf{m}(x) := \frac{\mathbf{a}(t, x) - \mathbf{a}^t(t, x)}{2},$$

and set

$$A(p, q, t, x) := \frac{1}{2} p \cdot s(t, x) p + \frac{1}{2} (q - \mathbf{m}(t, x) p) \cdot s^{-1}(t, x) (q - \mathbf{m}(t, x) p), \tag{2-3}$$

so that the following lemma holds.

**Lemma 2.2.** *There exists a constant  $C(\Lambda) < \infty$  such that, for every  $(t, x) \in \mathbb{R} \times \mathbb{R}^d$ ,*

$$(p, q) \mapsto A(p, q, t, x) - C^{-1}(|p|^2 + |q|^2) \quad \text{is convex,}$$

and

$$(p, q) \mapsto A(p, q, t, x) - C(|p|^2 + |q|^2) \quad \text{is concave.}$$

Moreover, for every  $(t, x) \in \mathbb{R} \times \mathbb{R}^d$  and  $p, q \in \mathbb{R}^d$ ,

$$A(p, q, t, x) \geq p \cdot q,$$

with equality if and only if  $q = \mathbf{a}(t, x) p$ .

*Proof.* We briefly recall the proof; see also [Armstrong and Mourrat 2016, (2.6)]. The fact that  $(p, q) \mapsto A(p, q, t, x)$  is uniformly convex and  $C^{1,1}$  follows from the definition of  $\Omega$  in (1-5). The second part of the lemma is a consequence of the identity

$$A(p, q, t, x) - p \cdot q = \frac{1}{2}(\mathbf{a}(t, x)p - q) s^{-1}(t, x)(\mathbf{a}(t, x)p - q). \quad \square$$

The functional  $\mathcal{J}$  appearing in Proposition 2.1 is defined, for every  $u \in H_{\text{par}}^1(I \times U)$  and  $u^* \in L^2(I; H^{-1}(U))$ , by

$$\mathcal{J}[u, u^*] := \inf \left\{ \int_{I \times U} (A(\nabla u, \mathbf{g}, \cdot) - \nabla u \cdot \mathbf{g}) : -\nabla \cdot \mathbf{g} = u^* - \partial_t u \right\}. \quad (2-4)$$

In the infimum above, we understand that  $\mathbf{g} \in L^2(I \times U; \mathbb{R}^d)$ , and the last condition is interpreted as

$$\text{for all } \phi \in L^2(I; H_0^1(U)), \quad \int_{I \times U} \nabla \phi \cdot \mathbf{g} = \int_{I \times U} \phi(u^* - \partial_t u).$$

In the integral on the right side of (2-4), the dot in the expression  $A(\nabla u, \mathbf{g}, \cdot)$  stands for the time-space variable; that is,

$$\int_{I \times U} (A(\nabla u, \mathbf{g}, \cdot) - \nabla u \cdot \mathbf{g}) = \int_I \int_U (A(\nabla u(t, x), \mathbf{g}(t, x), t, x) - \nabla u(t, x) \cdot \mathbf{g}(t, x)) dx dt.$$

**2B. Subadditive quantities and basic properties.** In this subsection, we define the subadditive quantities and collect their basic properties. Although their definitions are actually very natural and intuitive, many readers will not find them to be on first reading. In order to understand the motivation for studying them, it is best to first have some familiarity with the elliptic case with symmetric coefficients, which is described in [Armstrong et al. 2017b]. Indeed, much of what appears below can be compared to Chapter 2 of that book, and in fact this paper can be seen as a generalization of Chapters 1–3 of that book to the parabolic setting. Now, since the subadditive quantities are endowed from the variational structure of the equation, it is natural that the parabolic versions should be somewhat more complicated than the elliptic ones. A similar issue was encountered in [Armstrong and Mourrat 2016], where subadditive quantities were defined and analyzed for “nonvariational” elliptic equations.

In any case, the most convincing demonstration that these are the “right” quantities will have to wait until Section 5, where we prove that quantitative information about the convergence of the subadditive quantities can be translated directly into control of the first-order correctors and therefore into estimates on the rate of homogenization.

Without further ado, we give the definitions of the subadditive quantities. For every Lipschitz domain  $U \subseteq \mathbb{R}^d$ , bounded interval  $I \subseteq \mathbb{R}$  and  $p, q \in \mathbb{R}^d$ , we define

$$\mu(I \times U, p, q) := \inf_{(\nabla v, \mathbf{h}) \in \mathcal{C}_0(I \times U)} \int_{I \times U} A(p + \nabla v, q + \mathbf{h}, \cdot), \quad (2-5)$$

where the infimum is taken over  $(\nabla v, \mathbf{h})$  ranging in the space

$$\mathcal{C}_0(I \times U) := \left\{ (\nabla v, \mathbf{h}) \in L^2(I \times U, \mathbb{R}^d)^2 : v \in H_{\text{par},0}^1(I \times U) \text{ and} \right. \\ \left. \text{for all } \phi \in L^2(I; H^1(U)), \int_{I \times U} \nabla \phi \cdot \mathbf{h} = - \int_{I \times U} \phi \partial_t v \right\}. \quad (2-6)$$

Since  $\phi \in L^2(I; H^1(U))$  and  $\partial_t v \in L^2(I; \widehat{H}^{-1}(U))$ , the last integral is well-defined, in the usual interpretation as

$$\int_I \langle \phi(t, \cdot), \partial_t v(t, \cdot) \rangle dt,$$

where here  $\langle \cdot, \cdot \rangle$  denotes the duality pairing between  $H^1(U)$  and  $\widehat{H}^{-1}(U)$ . Testing the condition in (2-6) against the function  $\phi(t, x) := p \cdot x$  and integrating by parts in time, we see that any candidate  $\mathbf{h}$  must satisfy

$$\int_{I \times U} \mathbf{h} = 0. \quad (2-7)$$

The dual subadditive quantity  $\mu^*$  is defined, for every  $p^*, q^* \in \mathbb{R}^d$ , by

$$\mu^*(I \times U, q^*, p^*) := \sup_{(\nabla u, \mathbf{g}) \in \mathcal{C}(I \times U)} \int_{I \times U} (-A(\nabla u, \mathbf{g}, \cdot) + q^* \cdot \nabla u + p^* \cdot \mathbf{g}), \quad (2-8)$$

where the supremum is taken over  $(\nabla u, \mathbf{g})$  ranging in the space

$$\mathcal{C}(I \times U) := \left\{ (\nabla u, \mathbf{g}) \in L^2(I \times U; \mathbb{R}^d)^2 : u \in H_{\text{par}}^1(I \times U), \right. \\ \left. \text{for all } \phi \in L^2(I; H_0^1(U)), \int_{I \times U} \nabla \phi \cdot \mathbf{g} = - \int_{I \times U} \phi \partial_t u \right\}. \quad (2-9)$$

Note that for each  $p \in \mathbb{R}^d$  and  $(\nabla v, \mathbf{h}) \in \mathcal{C}_0(I \times U)$ , we have  $(p + \nabla v, q + \mathbf{h}) \in \mathcal{C}(I \times U)$ . Using also (1-12) and (2-7), we thus deduce that for every  $p, q, p^*, q^* \in \mathbb{R}^d$ ,

$$\mu^*(I \times U, q^*, p^*) \geq q^* \cdot p + p^* \cdot q - \mu(I \times U, p, q). \quad (2-10)$$

That is, the function  $(q^*, p^*) \mapsto \mu^*(I \times U, q^*, p^*)$  is bounded below by the convex dual of the function  $(p, q) \mapsto \mu(I \times U, p, q)$ . As in the elliptic case, see [Armstrong et al. 2016, Lemma 3.1; 2017a], we will combine  $\mu$  and  $\mu^*$  into a master quantity denoted by  $J$  which monitors the defect in this convex duality pairing. For concision, we set

$$V := I \times U, \quad (2-11)$$

and define a  $2d$ -by- $2d$  matrix field  $\mathbf{A} : \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}^{2d \times 2d}$  by

$$\mathbf{A}(t, x) := \begin{pmatrix} \mathbf{s}(t, x) - \mathbf{m}(t, x) \mathbf{s}^{-1}(t, x) \mathbf{m}(t, x) & \mathbf{m}(t, x) \mathbf{s}^{-1}(t, x) \\ -\mathbf{s}^{-1}(t, x) \mathbf{m}(t, x) & \mathbf{s}^{-1}(t, x) \end{pmatrix},$$

so that

$$A(p, q, t, x) = \frac{1}{2} \begin{pmatrix} p \\ q \end{pmatrix} \cdot \mathbf{A} \begin{pmatrix} p \\ q \end{pmatrix}.$$

This notation allows us to rewrite the definitions of  $\mu$  and  $\mu^*$  in (2-5) and (2-8) in more compact notation: for every  $X, X^* \in \mathbb{R}^{2d}$ , we have

$$\mu(V, X) = \inf_{S \in X + \mathcal{C}_0(V)} \int_V \frac{1}{2} S \cdot AS, \quad (2-12)$$

$$\mu^*(V, X^*) = \sup_{S \in \mathcal{C}(V)} \int_V \left( -\frac{1}{2} S \cdot AS + X^* \cdot S \right), \quad (2-13)$$

and the inequality (2-10) can be rewritten as

$$\mu^*(V, X^*) \geq X \cdot X^* - \mu(V, X). \quad (2-14)$$

We now set

$$\mathcal{S}(V) := \left\{ (\nabla v, \mathbf{h}) \in \mathcal{C}(V) : \text{for all } (\nabla \phi, \mathbf{f}) \in \mathcal{C}_0(V), \int_V \begin{pmatrix} \nabla \phi \\ \mathbf{f} \end{pmatrix} \cdot A \begin{pmatrix} \nabla v \\ \mathbf{h} \end{pmatrix} = 0 \right\}, \quad (2-15)$$

and for every  $X, X^* \in \mathbb{R}^{2d}$ ,

$$J(V, X, X^*) := \sup_{S \in \mathcal{S}(V)} \int_V \left( -\frac{1}{2} S \cdot AS - X \cdot AS + X^* \cdot S \right). \quad (2-16)$$

The master quantity  $J$  can be rewritten in the following more explicit notation:

$$J\left(V, \begin{pmatrix} p \\ q \end{pmatrix}, \begin{pmatrix} q^* \\ p^* \end{pmatrix}\right) := \sup_{(\nabla v, \mathbf{g}) \in \mathcal{S}(V)} \int_V \left( -\frac{1}{2} \begin{pmatrix} \nabla v \\ \mathbf{g} \end{pmatrix} \cdot A \begin{pmatrix} \nabla v \\ \mathbf{g} \end{pmatrix} - \begin{pmatrix} p \\ q \end{pmatrix} \cdot A \begin{pmatrix} \nabla v \\ \mathbf{g} \end{pmatrix} + \begin{pmatrix} q^* \\ p^* \end{pmatrix} \cdot \begin{pmatrix} \nabla v \\ \mathbf{g} \end{pmatrix} \right). \quad (2-17)$$

The next lemma shows that  $J$  indeed monitors the defect in convex duality between  $\mu$  and  $\mu^*$ .

**Lemma 2.3.** *For every  $X, X^* \in \mathbb{R}^{2d}$ ,*

$$J(V, X, X^*) = \mu(V, X) + \mu^*(V, X^*) - X \cdot X^*. \quad (2-18)$$

*Moreover, the maximizer  $S(\cdot, V, X, X^*)$  in (2-16) is the difference between the maximizer of  $\mu^*(V, X^*)$  in (2-13) and the minimizer of  $\mu(V, X)$  in (2-12).*

*Proof.* We first argue that, for every  $X^* \in \mathbb{R}^{2d}$ ,

$$\mu^*(V, X^*) = \sup_{S \in \mathcal{S}(V)} \int_V \left( -\frac{1}{2} S \cdot AS + X^* \cdot S \right). \quad (2-19)$$

Let  $S^* \in \mathcal{C}(V)$  denote the maximizer in the definition of  $\mu^*(V, X^*)$ . Note that, for every  $X^* \in \mathbb{R}^{2d}$  and  $S \in \mathcal{C}_0(V)$ ,

$$\int_U X^* \cdot S = 0. \quad (2-20)$$

By the first variation for  $\mu^*$ , we deduce that

$$\text{for all } S' \in \mathcal{C}_0(V), \quad \int_V S' \cdot AS^* = 0.$$

That is,  $S^* \in \mathcal{S}(U)$ , and thus (2-19) holds.



Let  $X = (p, q) \in \mathbb{R}^{2d}$  and

$$S_0 = (\nabla v, \mathbf{h}) \in X + \mathcal{C}_0(V) \quad (2-21)$$

denote the minimizer in the definition of  $\mu(V, p, q) = \mu(V, X)$ . For every  $S \in \mathcal{S}(V)$ ,

$$\mu(V, X) + \int_V (X^* \cdot S - \frac{1}{2} S \cdot AS) - X \cdot X^* = \int_V (\frac{1}{2} S_0 \cdot AS_0 - \frac{1}{2} S \cdot AS + X^* \cdot S) - X \cdot X^*. \quad (2-22)$$

By (2-21), we have  $X = \int_V S_0$ . For each  $S \in \mathcal{S}(V)$ ,

$$\int_V S \cdot AS_0 = \int_V S \cdot AX,$$

and this last identity holds true in particular for  $S = S_0$ . We obtain that the left side of (2-22) is equal to

$$\int_V (-\frac{1}{2}(S - S_0) \cdot A(S - S_0) - X \cdot A(S - S_0) + X^* \cdot (S - S_0)).$$

We compare this result to the identities (2-19) and (2-16) to obtain the lemma.  $\square$

The next lemma collects elementary properties of  $J$  and its minimizer. It can be compared with [Armstrong et al. 2017b, Lemma 2.2].

**Lemma 2.4** (basic properties of  $J$ ). *The quantity  $J(V, X, X^*)$  and its maximizer  $S(\cdot, V, X, X^*)$  satisfy the following properties:*

- *The mapping  $(X, X^*) \mapsto J(V, X, X^*)$  is quadratic.*
- *Uniformly convex and  $C^{1,1}$  in  $X$  and  $X^*$  separately. There exists a constant  $C(d, \Lambda) < \infty$  such that, for every  $X_1, X_2, X^* \in \mathbb{R}^{2d}$ ,*

$$\frac{1}{C}|X_1 - X_2|^2 \leq \frac{1}{2}J(V, X_1, X^*) + \frac{1}{2}J(V, X_2, X^*) - J(V, \frac{1}{2}X_1 + \frac{1}{2}X_2, X^*) \leq C|X_1 - X_2|^2 \quad (2-23)$$

and, for every  $X_1^*, X_2^*, X \in \mathbb{R}^{2d}$ ,

$$\frac{1}{C}|X_1^* - X_2^*|^2 \leq \frac{1}{2}J(V, X, X_1^*) + \frac{1}{2}J(V, X, X_2^*) - J(V, X, \frac{1}{2}X_1^* + \frac{1}{2}X_2^*) \leq C|X_1^* - X_2^*|^2. \quad (2-24)$$

- **Subadditivity.** *Let  $V_1, \dots, V_N \subseteq V$  be parabolic cylinders that partition  $V$ , in the sense that  $V_i \cap V_j = \emptyset$  if  $i \neq j$  and*

$$\left| V \setminus \bigcup_{i=1}^N V_i \right| = 0.$$

For every  $X, X^* \in \mathbb{R}^{2d}$ , we have

$$J(V, X, X^*) \leq \sum_{i=1}^N \frac{|V_i|}{|V|} J(V_i, X, X^*). \quad (2-25)$$

- **First variation for  $J$ .** *For  $X, X^* \in \mathbb{R}^{2d}$ , the function  $S(\cdot, V, X, X^*)$  is the unique element of  $\mathcal{S}(V)$  such that*

$$\text{for all } T \in \mathcal{S}(V), \quad \int_V T \cdot AS(\cdot, V, X, X^*) = \int_V (-X \cdot AT + X^* \cdot T). \quad (2-26)$$

- **Quadratic response.** For every  $X, X^* \in \mathbb{R}^{2d}$  and  $T \in \mathcal{S}(V)$ ,

$$\begin{aligned} \frac{1}{C} \int_V |T - S(\cdot, V, X, X^*)|^2 &\leq J(V, X, X^*) - \int_V \left(-\frac{1}{2}T \cdot AT - X \cdot AT + X^* \cdot T\right) \\ &\leq C \int_V |T - S(\cdot, V, X, X^*)|^2. \end{aligned} \quad (2-27)$$

- **Formulas for derivatives of  $J$ .** For every  $X, X^* \in \mathbb{R}^{2d}$ ,

$$\nabla_X J(V, X, X^*) = - \int_V AS(\cdot, V, X, X^*), \quad (2-28)$$

$$\nabla_{X^*} J(V, X, X^*) = \int_V S(\cdot, V, X, X^*). \quad (2-29)$$

*Proof.* Since these properties are easy to check and their proofs are almost the same of those of [Armstrong et al. 2017b, Lemma 2.2], we omit the details.  $\square$

**Remark 2.5.** Since  $X^* \mapsto \mu^*(V, X^*)$  is a quadratic form, we obtain from (2-28) and Lemma 2.3 that

$$\int_V S(\cdot, V, X, 0) = \nabla_{X^*} J(V, X, 0) = -X, \quad (2-30)$$

a property which also follows directly from the definition of  $\mu$  in (2-12) and the identification of  $-S(\cdot, V, X, 0)$  as the minimizer in this definition. From (2-29) and Lemma 2.3, we also obtain the dual identity

$$\int_V AS(\cdot, V, 0, X^*) = X^*.$$

In the next lemma, we relate the space  $\mathcal{S}(I \times U)$  with the space of solutions of the parabolic equation and of its dual. Define the vector space  $\mathcal{A}(I \times U)$  to be the set of weak solutions  $u \in H_{\text{par}}^1(I \times U)$  of the equation

$$\partial_t u - \nabla \cdot (\mathbf{a} \nabla u) = 0 \quad \text{in } I \times U,$$

and the vector space  $\mathcal{A}^*(I \times U)$  to be the set of weak solutions  $u^* \in H_{\text{par}}^1(I \times U)$  of the dual equation

$$\partial_t u^* + \nabla \cdot (\mathbf{a}^\dagger \nabla u^*) = 0 \quad \text{in } I \times U.$$

Note that the direction of time is reversed in the dual equation. Precisely,

$$\begin{aligned} \mathcal{A}(I \times U) &:= \left\{ u \in H_{\text{par}}^1(I \times U) : \text{for all } w \in L^2(I; H_0^1(U)), \int_{I \times U} w \partial_t u = - \int_{I \times U} \nabla w \cdot \mathbf{a} \nabla u \right\}, \\ \mathcal{A}^*(I \times U) &:= \left\{ u^* \in H_{\text{par}}^1(I \times U) : \text{for all } w \in L^2(I; H_0^1(U)), \int_{I \times U} w \partial_t u^* = \int_{I \times U} \nabla w \cdot \mathbf{a}^\dagger \nabla u^* \right\}. \end{aligned}$$

**Lemma 2.6.** *We have*

$$\mathcal{S}(V) := \{(\nabla u + \nabla u^*, \mathbf{a} \nabla u - \mathbf{a}^\dagger \nabla u^*) : u \in \mathcal{A}(V), u^* \in \mathcal{A}^*(V)\}. \quad (2-31)$$

*Proof.* Recall that  $V = I \times U$ , and denote by  $\mathcal{S}'(I \times U)$  the set on the right side of (2-31). The condition

$$\int_{I \times U} (\nabla \phi, \mathbf{f}) \cdot \mathbf{A}(\nabla v, \mathbf{h}) = 0$$

appearing in (2-15) can be rewritten more explicitly as

$$\int_{I \times U} (\nabla \phi \cdot s \nabla v + (\mathbf{f} - \mathbf{m} \nabla \phi) \cdot s^{-1}(\mathbf{h} - \mathbf{m} \nabla v)) = 0. \quad (2-32)$$

We first verify that  $\mathcal{S}(I \times U) \subseteq \mathcal{S}'(I \times U)$ . The space  $\{0\} \times L^2(I, L^2_{\text{sol},0}(U))$  is a subspace of  $\mathcal{C}_0(I \times U)$ . Hence, if (2-32) holds for every  $(\nabla \phi, \mathbf{h}) \in \mathcal{C}_0(I \times U)$ , then in particular

$$\text{for all } \mathbf{f} \in L^2(I; L^2_{\text{sol},0}(U)), \quad \int_{I \times U} \mathbf{f} \cdot s^{-1}(\mathbf{h} - \mathbf{m} \nabla v) = 0.$$

In other words,  $s^{-1}(\mathbf{h} - \mathbf{m} \nabla v)$  belongs to the space orthogonal to  $L^2(I; L^2_{\text{sol},0}(U))$  in  $L^2(I \times U)$ . That is, there exists  $w \in L^2(I; H^1(U))$  such that

$$s^{-1}(\mathbf{h} - \mathbf{m} \nabla v) = \nabla w,$$

and we deduce that for every  $(\nabla \phi, \mathbf{f}) \in \mathcal{C}_0(I \times U)$ ,

$$\int_{I \times U} (\nabla \phi \cdot (s \nabla v + \mathbf{m} \nabla w) - w \partial_t \phi) = 0.$$

Denoting by  $\Delta_N^{-1}$  the solution operator for the Laplace equation on  $U$  with null Neumann boundary condition, we observe that for each  $\phi \in H^1_{\text{par},0}(I \times U)$ , the pair  $(\nabla \phi, \nabla \Delta_N^{-1}(\partial_t \phi))$  belongs to  $\mathcal{C}_0(I \times U)$ . The identity above therefore holds for arbitrary  $\phi \in H^1_{\text{par},0}(I \times U)$ , and we thus deduce that  $\partial_t w \in L^2(I, H^{-1}(U))$ . We can then integrate by parts in time and obtain that

$$\text{for all } \phi \in H^1_{\text{par},0}(I \times U), \quad \int_{I \times U} (\nabla \phi \cdot (s \nabla v + \mathbf{m} \nabla w) + \phi \partial_t w) = 0. \quad (2-33)$$

This property can be extended to arbitrary  $\phi \in L^2(I; H^1_0(U))$  by density. The additional requirement that  $(\nabla v, \mathbf{h}) \in \mathcal{C}(I \times U)$  gives

$$\text{for all } \psi \in L^2(I; H^1_0(U)), \quad \int_{I \times U} (\nabla \psi \cdot (s \nabla w + \mathbf{m} \nabla v) + \psi \partial_t v) = 0. \quad (2-34)$$

Setting

$$u := \frac{1}{2}(v + w), \quad u^* := \frac{1}{2}(v - w), \quad (2-35)$$

we deduce that  $u \in \mathcal{A}(I \times U)$ ,  $u^* \in \mathcal{A}^*(I \times U)$ , with

$$v = \frac{1}{2}(u + u^*), \quad \mathbf{h} = \mathbf{a} \nabla u - \mathbf{a}^t \nabla u^*,$$

and this completes the proof that  $\mathcal{S}(I \times U) \subseteq \mathcal{S}'(I \times U)$ .

Conversely, given  $u \in \mathcal{A}(I \times U)$  and  $u^* \in \mathcal{A}^*(I \times U)$ , we set

$$v = u + u^*, \quad w := u - u^*, \quad \mathbf{h} := \mathbf{a} \nabla u - \mathbf{a}^t \nabla u^* = s \nabla w + \mathbf{m} \nabla v,$$

and observe that

$$\mathbf{a}\nabla u + \mathbf{a}^\dagger \nabla u^* = s\nabla v + \mathbf{m}\nabla w.$$

The identities (2-33) and (2-34) follow. This implies that the condition (2-32) is satisfied for every  $(\nabla\phi, \mathbf{f}) \in \mathcal{C}_0(I \times U)$ , and hence that  $(\nabla v, \mathbf{h}) \in \mathcal{S}(I \times U)$ . We have thus shown that  $\mathcal{S}'(I \times U) \subseteq \mathcal{S}(I \times U)$ , which completes the proof.  $\square$

**Remark 2.7.** Note that for  $S = (\nabla u + \nabla u^*, \mathbf{a}\nabla u - \mathbf{a}^\dagger \nabla u^*) \in \mathcal{S}(V)$ , we have

$$AS = (\mathbf{a}\nabla u + \mathbf{a}^\dagger \nabla u^*, \nabla u - \nabla u^*). \tag{2-36}$$

Indeed, (2-36) is implicit in the proof of Lemma 2.6 above and can also be checked by a direct computation. In particular,  $\nabla u$  can be written as one half of the sum of the first component of  $S$  and second component of  $AS$ , and  $\mathbf{a}\nabla u$  can be recovered similarly. This observation is needed in Section 5 in the construction of (approximate) correctors.

### 3. Functional inequalities

We collect here some functional inequalities which will be useful in the rest of the paper. The two main results are a ‘‘multiscale’’ version of the Poincaré inequality, and a Caccioppoli-type inequality for elements of  $\mathcal{S}(\square_n)$ . The proof of the latter is based on a parabolic version of the Helmholtz–Hodge decomposition of vector fields, which is of independent interest.

We first recall a useful version of the Poincaré inequality, for functions of the space variable only.

**Lemma 3.1.** *Let  $\psi \in L^2(\square_n)$  satisfy*

$$\int_{\square_n} \psi = 1.$$

*There exists  $C(d) < \infty$  such that, for every  $u \in H^1(\square_n)$ ,*

$$\left\| u - \int_{\square_n} u \psi \right\|_{L^2(\square_n)} \leq C \|\psi\|_{L^2(\square_n)} \|\nabla u\|_{\underline{H}^{-1}(\square_n)}. \tag{3-1}$$

*Proof.* By the usual Poincaré inequality, all we need to show is that

$$\left| \int_{\square_n} u(1 - \psi) \right| \leq C \|\psi\|_{L^2(\square_n)} \|\nabla u\|_{\underline{H}^{-1}(\square_n)}. \tag{3-2}$$

Let  $w$  be the solution of the Neumann problem

$$\begin{cases} -\Delta w = 1 - \psi & \text{in } \square_n, \\ \mathbf{n} \cdot \nabla w = 0 & \text{on } \partial \square_n. \end{cases}$$

Notice that this has a solution because  $\int_{\square_n} (1 - \psi) = 0$ , and we have the  $H^2$  estimate, see for instance [Armstrong et al. 2017b, Lemma B.18],

$$\|\nabla w\|_{\underline{H}^1(\square_n)} \leq C \|1 - \psi\|_{L^2(\square_n)} \leq C(1 + \|\psi\|_{L^2(\square_n)}) \leq C \|\psi\|_{L^2(\square_n)}.$$



Testing the equation for  $w$  by  $u$  thus yields

$$\left| \int_{\square_n} u(1 - \psi) \right| = \left| \int_{\square_n} \nabla u \cdot \nabla w \right| \leq C \|\nabla u\|_{\widehat{H}^{-1}(\square_n)} \|\nabla w\|_{\underline{H}^1(\square_n)} \leq C \|\psi\|_{\underline{L}^2(\square_n)} \|\nabla u\|_{\widehat{H}^{-1}(\square_n)}. \quad \square$$

For every parabolic cylinder  $V$  and  $f \in L^1(V)$ , we recall that we use the following shorthand notation for the spatial average of  $f$  over  $V$ :

$$(f)_V := \int_V f. \tag{3-3}$$

By the standard Poincaré inequality in  $1 + d$  coordinates, we have

$$\|u - (u)_{\square_n}\|_{\underline{L}^2(\square_n)} \leq C3^n \|\nabla u\|_{\underline{L}^2(\square_n)} + C3^{2n} \|\partial_t u\|_{\underline{L}^2(\square_n)}. \tag{3-4}$$

In the context of parabolic equations, it is natural to try to preserve a matching between the number of times a function is differentiated in space and *half* the number of times it is differentiated in time. The estimate (3-4) is not consistent with this scaling. The purpose of the next proposition is to obtain such a bound — see also Corollary 3.4 below.

**Proposition 3.2.** *There exists  $C(d) < \infty$  such that, for every  $u \in H^1_{\text{par}}(\square_n)$  and  $\mathbf{g} \in L^2(\square_n; \mathbb{R}^d)$  satisfying  $\partial_t u = \nabla \cdot \mathbf{g}$ , we have*

$$\|u - (u)_{\square_n}\|_{\underline{L}^2(\square_n)} \leq C(\|\nabla u\|_{\underline{L}^2(I_n; \widehat{H}^{-1}(\square_n))} + \|\mathbf{g}\|_{\underline{L}^1(I_n; \underline{H}^{-1}(\square_n))}).$$

**Remark 3.3.** In the statement of Proposition 3.2 (and similarly for Corollary 3.4 and Proposition 3.7 below), the condition  $\partial_t u = \nabla \cdot \mathbf{g}$  is interpreted as

$$\text{for all } \phi \in L^2(I_n; H^1_0(\square_n)), \quad \int_{\square_n} \nabla \phi \cdot \mathbf{g} = - \int_{\square_n} \phi \partial_t u.$$

Equivalently, this amounts to saying that  $(\nabla u, \mathbf{g}) \in \mathcal{C}(\square_n)$ . As an example, we can always take  $\mathbf{g} = \nabla \Delta_{\square_n}^{-1} \partial_t u$ , where  $\Delta_{\square_n}^{-1}$  is the solution operator for the Laplacian in  $\square_n$  with null Dirichlet boundary condition.

*Proof of Proposition 3.2.* Let  $\psi \in C_c^\infty(\square_n)$  be a smooth function of compact support in  $\square_n$  such that  $\int_{\square_n} \psi(x) dx = 1$ ,  $0 \leq \psi \leq 2$ , and

$$3^{-n} \|\nabla \psi\|_{L^\infty(\square_n)} + \|\nabla^2 \psi\|_{L^\infty(\square_n)} \leq C3^{-2n}. \tag{3-5}$$

We write  $(u)_{\square_n, \psi} := \int_{\square_n} u(x) \psi(x) dx$  and  $(u)_{\square_n} := \int_{\square_n} u(x) dx$ . Using the Poincaré inequality (in the form given by Lemma 3.1) in time slices gives, for every  $t \in I_n$ ,

$$\|u(t, \cdot) - (u(t, \cdot))_{\square_n, \psi}\|_{\underline{L}^2(\square_n)} \leq C \|\nabla u(t, \cdot)\|_{\widehat{H}^{-1}(\square_n)}.$$

Thus

$$\int_{I_n} \int_{\square_n} |u(t, x) - (u(t, \cdot))_{\square_n, \psi}|^2 dx dt \leq C \int_{I_n} \|\nabla u(t, \cdot)\|_{\widehat{H}^{-1}(\square_n)}^2 dt. \tag{3-6}$$

Since  $\psi \in H_0^1(U)$  and  $\nabla\psi \in H_0^1(U; \mathbb{R}^d)$ , we have, for every  $t \in I_n$ ,

$$\begin{aligned} |\partial_t(u(t, \cdot))_{\square_n, \psi}| &= \left| \int_{\square_n} \psi(x) \partial_t u(t, x) dx \right| = \left| \int_{\square_n} \nabla\psi(x) \cdot \mathbf{g}(t, \cdot)(x) dx \right| \\ &\leq C \|\nabla\psi\|_{\underline{H}^1(\square_n)} \|\mathbf{g}(t, \cdot)\|_{\underline{H}^{-1}(\square_n)} \leq C3^{-2n} \|\mathbf{g}(t, \cdot)\|_{\underline{H}^{-1}(\square_n)}. \end{aligned}$$

Thus

$$\begin{aligned} \sup_{t \in I_n} \left| (u(t, \cdot))_{\square_n, \psi} - \int_{I_n \times \square_n} \psi(x) u(t, x) dx dt \right| &\leq \int_{I_n} |\partial_t(u(t, \cdot))_{\square_n, \psi}| dt \\ &\leq C3^{-2n} \int_{I_n} \|\mathbf{g}(t, \cdot)\|_{\underline{H}^{-1}(\square_n)} dt = C \int_{I_n} \|\mathbf{g}(t, \cdot)\|_{\underline{H}^{-1}(\square_n)} dt. \end{aligned}$$

Combining this with (3-6), we obtain

$$\begin{aligned} \int_{I_n} \int_{\square_n} \left| u(t, x) - \int_{I_n \times \square_n} \psi(y) u(s, y) ds dy \right|^2 dx dt \\ \leq C \int_{I_n} \|\nabla u(t, \cdot)\|_{\underline{\hat{H}}^{-1}(\square_n)}^2 dt + \left( \int_{I_n} \|\mathbf{g}(t, \cdot)\|_{\underline{H}^{-1}(\square_n)} dt \right)^2. \end{aligned}$$

Since

$$\|u - (u)_{\square_n}\|_{\underline{L}^2(\square_n)} = \inf_{c \in \mathbb{R}} \|u - c\|_{\underline{L}^2(\square_n)},$$

this yields the announced result. □

**Corollary 3.4.** *There exists  $C(d) < \infty$  such that, for every  $u \in H_{\text{par}}^1(\square_n)$  and  $\mathbf{g} \in L^2(\square_n, \mathbb{R}^d)$  satisfying  $\partial_t u = \nabla \cdot \mathbf{g}$ ,*

$$\|u - (u)_{\square_n}\|_{\underline{L}^2(\square_n)} \leq C3^n (\|\nabla u\|_{\underline{L}^2(\square_n)} + \|\mathbf{g}\|_{\underline{L}^2(\square_n)}). \tag{3-7}$$

*Proof.* This follows from Proposition 3.2 and the inequalities

$$\|v\|_{\underline{L}^2(I_n; \underline{H}^{-1}(\square_n))} \leq \|v\|_{\underline{L}^2(I_n; \underline{\hat{H}}^{-1}(\square_n))} \leq C3^n \|v\|_{\underline{L}^2(\square_n)}. \quad \square$$

**Remark 3.5.** Recall from Remark 3.3 that in the statement of Corollary 3.4, we can take  $\mathbf{g} = \nabla \Delta_{\square_n}^{-1} \partial_t u$ . Moreover, there exists  $C(d) < \infty$  such that, for every  $f \in L^2(I_n; H^{-1}(\square_n))$ ,

$$C^{-1} \|f\|_{\underline{L}^2(I_n; \underline{H}^{-1}(\square_n))} \leq \|\nabla \Delta_{\square_n}^{-1} f\|_{\underline{L}^2(\square_n)} \leq C \|f\|_{\underline{L}^2(I_n; \underline{H}^{-1}(\square_n))}.$$

Indeed, by the standard Poincaré inequality, the norm  $\|\cdot\|_{\underline{H}^1(\square_n)}$  is equivalent to the norm  $u \mapsto \|\nabla u\|_{\underline{L}^2(\square_n)}$  on  $H_0^1(\square_n)$ , and moreover,

$$\begin{aligned} \sup \left\{ \int_{\square_n} f \phi : \phi \in L^2(I_n; H_0^1(\square_n)), \|\nabla \phi\|_{\underline{L}^2(\square_n)} \leq 1 \right\} \\ = \sup \left\{ \int_{\square_n} \nabla \phi \cdot \nabla \Delta_{\square_n}^{-1} f : \phi \in L^2(I_n; H_0^1(\square_n)), \|\nabla \phi\|_{\underline{L}^2(\square_n)} \leq 1 \right\} \\ = \|\nabla \Delta_{\square_n}^{-1} f\|_{\underline{L}^2(\square_n)}. \end{aligned}$$

In particular, on the right side of (3-7), we can replace the term  $\|\mathbf{g}\|_{\underline{L}^2(\square_n)}$  with  $\|\partial_t u\|_{\underline{L}^2(I_n; \underline{H}^{-1}(\square_n))}$ .

The next proposition allows us to obtain a control of the  $H_{\text{par}}^{-1}(V)$  norm of a function from knowledge of its spatial averages over large scales. For each  $m, n \in \mathbb{N}$ ,  $m \leq n$ , we set

$$\mathcal{Z}_m := [(3^{2m}\mathbb{Z}) \times (3^m\mathbb{Z}^d)] \cap \square_n. \quad (3-8)$$

Although  $\mathcal{Z}_m$  depends on  $n$ , we keep this dependence implicit in the notation, since its identity will be clear from the context. This is a parabolic version of the inequality which first appeared in [Armstrong et al. 2016, Proposition 6.1].

**Proposition 3.6** (multiscale Poincaré inequality). *There exists  $C(d, \Lambda) < \infty$  such that, for every  $n \geq 1$  and  $f \in L^2(\square_n)$ ,*

$$\|f\|_{\widehat{H}_{\text{par}}^{-1}(\square_n)} \leq C \|f\|_{L^2(\square_n)} + C \sum_{m=0}^{n-1} 3^m \left( |\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} |(f)_{z+\square_m}|^2 \right)^{\frac{1}{2}}.$$

*Proof of Proposition 3.6.* Recalling (1-15), we fix  $g \in H_{\text{par}}^1(\square_n)$  such that

$$\|g\|_{\widehat{H}_{\text{par}}^1(\square_n)} \leq 1, \quad (3-9)$$

and decompose the proof into two steps.

*Step 1.* In this step, we show that there exists a constant  $C(d) < \infty$  such that, for every  $m \in \{0, \dots, n\}$ ,

$$\sum_{z \in \mathcal{Z}_m} \int_{z+\square_m} |g - (g)_{z+\square_m}|^2 \leq C |\square_n| 3^{2m}. \quad (3-10)$$

By Corollary 3.4 and Remark 3.5, the left side above is bounded by

$$C 3^{2m} |\square_m| \sum_{z \in \mathcal{Z}_m} (\|\nabla g\|_{L^2(z+\square_m)}^2 + \|\partial_t g\|_{L^2(z_0+I_m, \underline{H}^{-1}(z'+\square_m))}^2),$$

where we write  $z = (z_0, z') \in \mathbb{Z} \times \mathbb{Z}^d$ . The contribution of the first term is easily estimated, since by (3-9),

$$|\square_m| \sum_{z \in \mathcal{Z}_m} \|\nabla g\|_{L^2(z+\square_m)}^2 = \int_{\square_n} |\nabla g|^2 \leq |\square_n|.$$

For the second term, we write

$$\begin{aligned} & \sum_{z \in \mathcal{Z}_m} \|\partial_t g\|_{L^2(z_0+I_m, \underline{H}^{-1}(z'+\square_m))}^2 \\ &= \sup \left\{ \sum_{z \in \mathcal{Z}_m} \left( \int_{z+\square_m} \phi_z \partial_t g \right)^2 : \phi_z \in L^2(z_0+I_m; H_0^1(z'+\square_m)), \|\nabla \phi_z\|_{L^2(z+\square_m)} \leq 1 \right\}. \end{aligned}$$

For  $\phi_z$  satisfying the conditions in the supremum above, we have

$$\begin{aligned} |\square_n|^{-1} \sum_{z \in \mathcal{Z}_m} \left( \int_{z+\square_m} \phi_z \partial_t g \right)^2 &= \int_{\square_n} \partial_t g \left( \sum_{z \in \mathcal{Z}_m} \phi_z \int_{z+\square_m} \phi_z \partial_t g \right) \\ &\leq \|\partial_t g\|_{L^2(I_n; \underline{H}^{-1}(\square_n))} \left\| \sum_{z \in \mathcal{Z}_m} \phi_z \int_{z+\square_m} \phi_z \partial_t g \right\|_{L^2(I_n; \underline{H}^1(\square_n))}. \end{aligned}$$

Notice that, by (3-9), the first term on the right side of the previous inequality is bounded by 1. Moreover, by the normalization of the functions  $\phi_z$ ,

$$\left\| \sum_{z \in \mathcal{Z}_m} \phi_z \int_{z+\square_n} \phi_z \partial_t g \right\|_{L^2(I_n; \underline{H}^1(\square_n))}^2 \leq C \left\| \sum_{z \in \mathcal{Z}_m} \nabla \phi_z \int_{z+\square_n} \phi_z \partial_t g \right\|_{L^2(\square_n)}^2 \leq C \frac{|\square_m|}{|\square_n|} \sum_{z \in \mathcal{Z}_m} \left( \int_{z+\square_m} \phi_z \partial_t g \right)^2.$$

Combining the last three displays, we arrive at

$$\sum_{z \in \mathcal{Z}_m} \|\partial_t g\|_{L^2(z_0+I_m, \underline{H}^{-1}(z'+\square_m))}^2 \leq C \frac{|\square_n|}{|\square_m|},$$

and this completes the proof of (3-10).

*Step 2.* We aim to control  $\int_{\square_n} f g$ , which we decompose into

$$\int_{\square_n} f g = \int_{\square_n} f(g - (g)_{\square_n}) + (f)_{\square_n} (g)_{\square_n}. \tag{3-11}$$

By the definition of the  $H_{\text{par}}^1$  norm in (1-10), we have  $(g)_{\square_n} \leq 3^n$ , and therefore, by Jensen’s inequality,

$$(f)_{\square_n} (g)_{\square_n} \leq 3^n |(f)_{\square_n}| \leq C 3^{n-1} \left( |\mathcal{Z}_{n-1}|^{-1} \sum_{z \in \mathcal{Z}_{n-1}} |(f)_{z+\square_{n-1}}|^2 \right)^{\frac{1}{2}}. \tag{3-12}$$

We then proceed to decompose the first integral on the right side of (3-11) recursively. For every  $m \in \{0, \dots, n-1\}$  and  $z \in \mathcal{Z}_{m+1}$ , we have

$$\begin{aligned} \int_{z+\square_{m+1}} f \cdot (g - (g)_{z+\square_{m+1}}) &= \sum_{y \in \mathcal{Z}_m \cap (z+\square_{m+1})} \int_{y+\square_m} f \cdot (g - (g)_{y+\square_m}) \\ &\quad + |\square_m| \sum_{y \in \mathcal{Z}_m \cap (z+\square_{m+1})} ((g)_{y+\square_m} - (g)_{z+\square_{m+1}}) \cdot (f)_{y+\square_m}. \end{aligned} \tag{3-13}$$

Summing over  $z \in \mathcal{Z}_{m+1}$  and using Hölder’s inequality, we get

$$\begin{aligned} \sum_{z \in \mathcal{Z}_{m+1}} \int_{z+\square_{m+1}} f \cdot (g - (g)_{z+\square_{m+1}}) &\leq \sum_{y \in \mathcal{Z}_m} \int_{y+\square_m} f \cdot (g - (g)_{y+\square_m}) \\ &\quad + |\square_m| \left( \sum_{\substack{z \in \mathcal{Z}_{m+1} \\ y \in \mathcal{Z}_m \cap (z+\square_{m+1})}} |(g)_{y+\square_m} - (g)_{z+\square_{m+1}}|^2 \right)^{\frac{1}{2}} \left( \sum_{y \in \mathcal{Z}_m} |(f)_{y+\square_m}|^2 \right)^{\frac{1}{2}}. \end{aligned}$$

By Jensen’s inequality, we have, for each  $z \in \mathcal{Z}_{m+1}$ ,

$$\sum_{y \in \mathcal{Z}_m \cap (z+\square_{m+1})} |(g)_{y+\square_m} - (g)_{z+\square_{m+1}}|^2 \leq \sum_{y \in \mathcal{Z}_m \cap (z+\square_{m+1})} \int_{y+\square_m} |g - (g)_{z+\square_{m+1}}|^2,$$

and thus, by (3-10),

$$\sum_{\substack{z \in \mathcal{Z}_{m+1} \\ y \in \mathcal{Z}_m \cap (z+\square_{m+1})}} |(g)_{y+\square_m} - (g)_{z+\square_{m+1}}|^2 \leq C \frac{|\square_n|}{|\square_m|} 3^{2m}.$$

Using also that  $|\mathcal{Z}_m| = |\square_n|/|\square_m|$  and combining with (3-13), we obtain

$$\sum_{z \in \mathcal{Z}_{m+1}} \int_{z+\square_{m+1}} f \cdot (g - (g)_{z+\square_{m+1}}) \leq \sum_{y \in \mathcal{Z}_m} \int_{y+\square_m} f \cdot (g - (g)_{y+\square_m}) + C |\square_n| 3^m \left( |\mathcal{Z}_m|^{-1} \sum_{y \in \mathcal{Z}_m} |(f)_{y+\square_m}|^2 \right)^{\frac{1}{2}}.$$

Summing over  $m \in \{0, \dots, n-1\}$  yields

$$\int_{\square_n} f(g - (g)_{\square_n}) \leq \sum_{z \in \mathcal{Z}_0} \int_{z+\square_0} f \cdot (g - (g)_{z+\square_0}) + C |\square_n| \sum_{m=0}^{n-1} 3^m \left( |\mathcal{Z}_m|^{-1} \sum_{y \in \mathcal{Z}_m} |(f)_{y+\square_m}|^2 \right)^{\frac{1}{2}}.$$

Hence, by Hölder's inequality and (3-10),

$$\int_{\square_n} f(g - (g)_{\square_n}) \leq C |\square_n|^{\frac{1}{2}} \left( \int_{\square_n} |f|^2 \right)^{\frac{1}{2}} + C |\square_n| \sum_{m=0}^{n-1} 3^m \left( |\mathcal{Z}_m|^{-1} \sum_{y \in \mathcal{Z}_m} |(f)_{y+\square_m}|^2 \right)^{\frac{1}{2}}.$$

Dividing by  $|\square_n|$  and combining with (3-11)–(3-12), we obtain

$$\int_{\square_n} fg \leq C \|f\|_{L^2(\square_n)} + C \sum_{m=0}^n 3^m \left( |\mathcal{Z}_m|^{-1} \sum_{y \in \mathcal{Z}_m} |(f)_{y+\square_m}|^2 \right)^{\frac{1}{2}}.$$

Taking the supremum over all  $g$  satisfying (3-9) yields the result.  $\square$

The name “multiscale Poincaré inequality” for Proposition 3.6 is best understood in conjunction with the following statement.

**Proposition 3.7.** *There exists a constant  $C(d) < \infty$  such that, for every  $n \in \mathbb{N}$ ,  $u \in H_{\text{par}}^1(\square_{n+1})$  and  $\mathbf{g} \in L^2(\square_{n+1}; \mathbb{R}^d)$  satisfying  $\partial_t u = \nabla \cdot \mathbf{g}$ , we have*

$$\|u - (u)_{\square_n}\|_{L^2(\square_n)} \leq C \|\nabla u\|_{\widehat{H}_{\text{par}}^{-1}(\square_{n+1})} + C \|\mathbf{g}\|_{\widehat{H}_{\text{par}}^{-1}(\square_{n+1})}.$$

**Remark 3.8.** It is clear that the proof of Proposition 3.7 can be adapted to show that for every  $r > 0$ , there exists a constant  $C(r, d) < \infty$  such that, for every  $n \in \mathbb{N}$ ,  $u \in H_{\text{par}}^1(\square_n)$  and  $\mathbf{g} \in L^2(\square_n; \mathbb{R}^d)$  satisfying  $\partial_t u = \nabla \cdot \mathbf{g}$ , we have

$$\|u - (u)_{\square_{n-r}}\|_{L^2(\square_{n-r})} \leq C \|\nabla u\|_{\widehat{H}_{\text{par}}^{-1}(\square_n)} + C \|\mathbf{g}\|_{\widehat{H}_{\text{par}}^{-1}(\square_n)}.$$

Although one can expect that the estimate above still holds for  $r = 0$ , we leave it as an open question here, and content ourselves with an interior estimate.

Combining Propositions 3.6 and 3.7 allows to estimate the (interior)  $L^2$  oscillation of  $u$  in terms of spatial averages of  $\nabla u$  and  $\mathbf{g}$ ; see also [Armstrong et al. 2016, Proposition 6.1]. The estimate yields better interior information than the “single-scale” Poincaré inequality provided by Proposition 3.2 as soon as the spatial averages of  $\nabla u$  and  $\mathbf{g}$  display nontrivial cancellations over large scales. This feature will be crucial to our subsequent arguments.

Before turning to the proof of Proposition 3.7, we recall the classical  $H^2$  estimate for solutions of the heat equation. For simplicity, we state it using periodic boundary conditions in the space variable. We denote the corresponding function spaces by  $H_{\#}^1(U)$ ,  $H_{\text{par},\#}^1(I \times U)$ , etc.

**Lemma 3.9** ( $H^2$  estimate for the Cauchy problem). *There exists  $C(d) < \infty$  such that, for every  $n \in \mathbb{N}$ , if  $f \in L^2(\square_n)$  and  $u \in H^1_{\text{par},\#}(\square_n)$  solves the Cauchy problem*

$$\begin{cases} (\partial_t - \Delta)u = f & \text{in } \square_n, \\ u = 0 & \text{on } \{-3^{2n}/2\} \times \square_n, \end{cases} \tag{3-14}$$

then  $\nabla u \in H^1_{\text{par},\#}(\square_n)$  and

$$3^{-n} \|u\|_{H^1_{\text{par},\#}(\square_n)} + \|\nabla u\|_{H^1_{\text{par},\#}(\square_n)} \leq C \|f\|_{L^2(\square_n)}. \tag{3-15}$$

*Proof.* By scaling, it suffices to prove the result for  $n = 0$ . For each  $s \in I_0$ , we test (3-14) against the function  $(t, x) \mapsto \mathbb{1}_{t < s} u(t, x)$  and get

$$\int_{(-\frac{1}{2}, s) \times \square_0} (u \partial_t u + |\nabla u|^2) = \int_{(-\frac{1}{2}, s) \times \square_0} u f,$$

which implies

$$\frac{1}{2} \|u(s, \cdot)\|_{L^2(\square_0)}^2 + \int_{(-\frac{1}{2}, s) \times \square_0} |\nabla u|^2 \leq \|u\|_{L^2(\square_0)} \|f\|_{L^2(\square_0)}.$$

Taking the supremum over  $s \in I_0$ , observing that

$$\sup_{s \in I_0} \|u(s, \cdot)\|_{L^2(\square_0)} \leq \|u\|_{L^2(\square_0)}$$

and using Young's inequality, we obtain

$$\|u\|_{L^2(\square_0)} + \|\nabla u\|_{L^2(\square_0)} \leq C \|f\|_{L^2(\square_0)}.$$

We now turn to the estimation of  $\|\nabla^2 u\|_{L^2(\square_0)}$ . We first observe that by integration by parts, we have  $\|\nabla^2 u\|_{L^2(\square_0)} = \|\Delta u\|_{L^2(\square_0)}$ . Moreover, using (3-14), we get

$$\|\Delta u\|_{L^2(\square_0)}^2 = \int_{\square_0} \Delta u (f + \partial_t u) = \int_{\square_0} f \Delta u - \int_{\square_0} \nabla u \cdot \partial_t \nabla u,$$

with

$$\int_{\square_0} \nabla u \cdot \partial_t \nabla u = \frac{1}{2} \|\nabla u(\frac{1}{2}, \cdot)\|_{L^2(\square_0)}^2 \geq 0,$$

and therefore

$$\|\Delta u\|_{L^2(\square_0)} \leq \|f\|_{L^2(\square_0)}. \tag{3-16}$$

We also need bounds on the time derivatives of  $u$  and  $\nabla u$ . Note that

$$\begin{aligned} \|\partial_t \nabla u\|_{L^2(I_0; H_{\#}^{-1}(\square_0))} &= \sup \left\{ \int_{\square_0} \partial_t \nabla u \cdot F : \|F\|_{L^2(I_0; H_{\#}^1(\square_0))} \leq 1 \right\} \\ &= \sup \left\{ \int_{\square_0} \partial_t u \nabla \cdot F : \|F\|_{L^2(I_0; H_{\#}^1(\square_0))} \leq 1 \right\} \leq \|\partial_t u\|_{L^2(\square_0)}, \end{aligned}$$

and we can estimate the  $L^2$ -norm of  $\partial_t u$  using (3-14) and (3-16):

$$\|\partial_t u\|_{L^2(\square_0)} = \|f + \Delta u\|_{L^2(\square_0)} \leq 2 \|f\|_{L^2(\square_0)}.$$

The obvious bound

$$\|\partial_t u\|_{L^2(I_0; H_{\#}^{-1}(\square_0))} \leq C \|\partial_t u\|_{L^2(\square_0)}$$

completes the proof of (3-15).  $\square$

*Proof of Proposition 3.7.* By homogeneity, it suffices to show the result for  $n = 0$ . Let  $\psi \in C_c^\infty(\mathbb{R}^{1+d})$  be a smooth function with compact support in  $\square_{-2}$  and such that  $\int_{\mathbb{R}^{1+d}} \psi = 1$ . We decompose the proof into three steps.

*Step 1.* Let  $\eta \in C_c^\infty(\square_1)$  be a smooth function with compact support in  $\square_{3/4}$  and such that  $\eta \equiv 1$  on  $\square_{1/2}$ . In this step, we show that there exists a constant  $C(d, \psi, \eta) < \infty$  such that, for every  $u \in L^2(\square_1)$ ,

$$\|\eta(u - u \star \psi)\|_{\widehat{H}_{\text{par}}^{-1}(\square_1)} \leq C(\|\nabla u\|_{\widehat{H}_{\text{par}}^{-1}(\square_1)} + \|\mathbf{g}\|_{\widehat{H}_{\text{par}}^{-1}(\square_1)}). \quad (3-17)$$

For each  $n \in \mathbb{N}$ , we set

$$\psi_n := 3^{n(2+d)} \psi(3^{2n}t, 3^n x).$$

Since the function  $\psi_{-1} - \psi_0$  is compactly supported and of mean zero, we can use, e.g., [Armstrong et al. 2017b, Lemma 5.7] (in  $1 + d$  dimensions) to rewrite it in the form

$$\psi_{-1} - \psi_0 = \partial_t \mathbf{h}^\circ + \nabla \cdot \mathbf{h}',$$

where  $\mathbf{h} = (\mathbf{h}^\circ, \mathbf{h}') \in C_c^\infty(\mathbb{R}^{1+d}; \mathbb{R} \times \mathbb{R}^d)$  is supported in  $\square_{-2}$  (with  $\mathbf{h}^\circ$  taking values in  $\mathbb{R}$  and  $\mathbf{h}'$  in  $\mathbb{R}^d$ ). For each  $n \in \mathbb{N}$ , we define

$$\mathbf{h}_n(t, x) = (\mathbf{h}_n^\circ, \mathbf{h}_n')(t, x) := 3^{n(2+d)} \mathbf{h}(3^{2n}t, 3^n x),$$

so that

$$\psi_{n-1} - \psi_n = 3^{-2n} \partial_t \mathbf{h}_n^\circ + 3^{-n} \nabla \cdot \mathbf{h}_n'.$$

Since  $u \star \psi_n \rightarrow u$  in  $L^2(\square_{3/4})$ , we can use the triangle inequality to bound

$$\|\eta(u - u \star \psi)\|_{\widehat{H}_{\text{par}}^{-1}(\square_1)} \leq \sum_{n=0}^{+\infty} \|\eta(u \star \psi_{n-1} - u \star \psi_n)\|_{\widehat{H}_{\text{par}}^{-1}(\square_1)}. \quad (3-18)$$

We next observe that, for every  $z \in \square_{3/4}$ ,

$$\begin{aligned} (u \star \psi_{n-1} - u \star \psi_n)(z) &= \int_{\square_1} u(y) (3^{-2n} \partial_t \mathbf{h}_n^\circ + 3^{-n} \nabla \cdot \mathbf{h}_n')(z - y) dy \\ &= \int_{\square_1} (3^{-2n} \partial_t u(y) \mathbf{h}_n^\circ(z - y) + 3^{-n} \nabla u(y) \cdot \mathbf{h}_n'(z - y)) dy \\ &= 3^{-n} \int_{\square_1} \nabla u(y) \cdot \mathbf{h}_n'(z - y) dy - 3^{-2n} \int_{\square_1} \mathbf{g}(y) \cdot \nabla \mathbf{h}_n^\circ(z - y) dy. \end{aligned}$$

We fix  $f \in H_{\text{par}}^1(\square_1)$ , set  $\tilde{f} := \eta f$ , and compute

$$\int_{\square_1} \eta(u \star \psi_{n-1} - u \star \psi_n) f = 3^{-n} \int_{\square_1} \nabla u \cdot (\tilde{f} \star \mathbf{h}_n') - 3^{-2n} \int_{\square_1} \mathbf{g} \cdot (\tilde{f} \star \nabla \mathbf{h}_n^\circ).$$

One can check that there exists a constant  $C(d, \psi, \mathbf{h}) < \infty$  such that

$$\|\tilde{f} * \mathbf{h}'_n\|_{H^1_{\text{par}}(\square_1)} + 3^{-n} \|\tilde{f} * \nabla \mathbf{h}_n^\circ\|_{H^1_{\text{par}}(\square_1)} \leq C \|f\|_{H^1_{\text{par}}(\square_1)}.$$

Summarizing, we have thus shown that

$$\|\eta(u \star \psi_{n-1} - u \star \psi_n)\|_{\widehat{H}^{-1}_{\text{par}}(\square_1)} \leq C 3^{-n} (\|\nabla u\|_{\widehat{H}^{-1}_{\text{par}}(\square_1)} + \|\mathbf{g}\|_{\widehat{H}^{-1}_{\text{par}}(\square_1)}).$$

Summing over  $n \in \mathbb{N}$  in (3-18), we obtain (3-17).

*Step 2.* Define

$$c(u) := u \star \psi(0).$$

In this step, we show that there exists a constant  $C(d, \psi, \eta) < \infty$  such that

$$\|\eta(u - c(u))\|_{\widehat{H}^{-1}_{\text{par}}(\square_1)} \leq C (\|\nabla u\|_{\widehat{H}^{-1}_{\text{par}}(\square_1)} + \|\mathbf{g}\|_{\widehat{H}^{-1}_{\text{par}}(\square_1)}). \quad (3-19)$$

This is an immediate consequence of the fact that there exists a constant  $C(d, \psi) < \infty$  such that, for every  $y, z \in \square_{3/4}$ ,

$$|u \star \psi(z) - u \star \psi(y)| \leq C (\|\nabla u\|_{\widehat{H}^{-1}_{\text{par}}(\square_1)} + \|\mathbf{g}\|_{\widehat{H}^{-1}_{\text{par}}(\square_1)}). \quad (3-20)$$

The proof of (3-20) is very similar to the previous step, only simpler: we represent the function  $\psi(\cdot - z) - \psi(\cdot - y)$  in the form

$$\partial_t \tilde{\mathbf{h}}^\circ + \nabla \cdot \tilde{\mathbf{h}}',$$

with  $(\tilde{\mathbf{h}}^\circ, \tilde{\mathbf{h}}') \in C_c^\infty(\square_1; \mathbb{R} \times \mathbb{R}^d)$ , and then obtain (3-20) thanks to an integration by parts.

*Step 3.* For concision, we write

$$\tilde{u} := u - c(u).$$

Let  $\chi \in C_c^\infty(\square_1)$  be a smooth function with compact support in  $\square_{1/2}$  and such that  $\chi \equiv 1$  on  $\square_0$ . In this step, we show that there exists a constant  $C(d, \psi, \eta, \chi) < \infty$  such that

$$\|\chi \tilde{u}\|_{L^2(\square_1)} \leq C (\|\nabla u\|_{\widehat{H}^{-1}_{\text{par}}(\square_1)} + \|\mathbf{g}\|_{\widehat{H}^{-1}_{\text{par}}(\square_1)}). \quad (3-21)$$

Let  $w \in H^1_{\text{par}, \#}(\square_1)$  solve the Cauchy problem

$$\begin{cases} (\partial_t - \Delta)w = \chi \tilde{u} & \text{in } \square_1, \\ w = 0 & \text{on } \{-\frac{9}{2}\} \times \square_1. \end{cases} \quad (3-22)$$

By Lemma 3.9, there exists a constant  $C(d) < \infty$  such that

$$\|w\|_{H^1_{\text{par}, \#}(\square_1)} + \|\nabla w\|_{H^1_{\text{par}, \#}(\square_1)} \leq C \|\chi \tilde{u}\|_{L^2(\square_1)}. \quad (3-23)$$



Testing (3-22) against  $\chi \tilde{u}$  and integrating by parts gives

$$\begin{aligned} \|\chi \tilde{u}\|_{L^2(\square_1)}^2 &= \int_{\square_1} (\nabla w \cdot \nabla(\chi \tilde{u}) - w \partial_t(\chi \tilde{u})) \\ &= \int_{\square_1} (\nabla w \cdot \nabla(\chi \tilde{u}) + \nabla(\chi w) \cdot \mathbf{g} - \partial_t \chi w \tilde{u}) \\ &\leq \|\nabla w\|_{H_{\#, \text{par}}^1(\square_1)} (\|\nabla u\|_{\widehat{H}_{\text{par}}^{-1}(\square_1)} + \|\tilde{u} \nabla \chi\|_{\widehat{H}_{\text{par}}^{-1}(\square_1)} + \|\mathbf{g}\|_{\widehat{H}_{\text{par}}^{-1}(\square_1)}) \\ &\quad + \|w\|_{H_{\#, \text{par}}^1(\square_1)} (\|\mathbf{g}\|_{\widehat{H}_{\text{par}}^{-1}(\square_1)} + \|\tilde{u} \nabla \chi\|_{\widehat{H}_{\text{par}}^{-1}(\square_1)}). \end{aligned}$$

Using the result of the previous step and (3-23), we obtain (3-21). This completes the proof of Proposition 3.7, since

$$\|u - (u)_{\square_0}\|_{L^2(\square_0)} \leq \|u - c(u)\|_{L^2(\square_0)} \leq \|\chi(u - c(u))\|_{L^2(\square_0)}. \quad \square$$

Finally, we prove a Caccioppoli-type inequality for elements of  $\mathcal{S}(\square_n)$ .

**Proposition 3.10.** *There exists a constant  $C(d, \Lambda) < \infty$  such that, for every  $n \in \mathbb{N}$  and  $S \in \mathcal{S}(\square_{n+1})$ ,*

$$\|S\|_{L^2(\square_n)} \leq C3^{-n} \|S\|_{\widehat{H}_{\text{par}}^{-1}(\square_{n+1})}.$$

In order to prove this result, we first describe more explicitly the structure of elements of  $\widehat{H}_{\text{par}}^{-1}(I \times U)$ .

**Lemma 3.11** (identification of  $\widehat{H}_{\text{par}}^{-1}$ ). *There exists a constant  $C(I, U, d) < \infty$  and, for each  $u^* \in L^2(I \times U)$ , a pair  $(w, w^*) \in L^2(I; H_0^1(U)) \times L^2(I; H^{-1}(U))$  such that*

$$u^* = \partial_t w + w^*, \quad (3-24)$$

with

$$\|w\|_{L^2(I; H^1(U))} + \|w^*\|_{L^2(I; H^{-1}(U))} \leq C \|u^*\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)}. \quad (3-25)$$

Let  $V \subseteq V' \subseteq I \times U$  be subdomains of  $I \times U$  such that  $\bar{V} \subseteq V'$ . If  $u^*$  vanishes outside of  $V$ , then there exists a pair  $(w, w^*) \in L^2(I; H_0^1(U)) \times L^2(I; H^{-1}(U))$  satisfying (3-24)–(3-25) for a constant  $C(V, V', I, U, d) < \infty$ , and such that  $w$  and  $w^*$  vanish outside of  $V'$ .

*Proof.* Denote by  $\Delta_{\square_n}^{-1}$  the solution operator for the Laplacian in  $\square_n$  with null Dirichlet boundary condition. We observe that

$$(u, v) \mapsto \int_{I \times U} (|U|^{-\frac{2}{d}} uv + \nabla u \cdot \nabla v + \nabla \Delta_U^{-1} \partial_t u \cdot \nabla \Delta_U^{-1} \partial_t v)$$

is a scalar product for the Hilbert space  $H_{\text{par}}^1(I \times U)$ . By the Riesz representation theorem, there exists a unique  $u \in H_{\text{par}}^1(I \times U)$  such that, for every  $v \in H_{\text{par}}^1(I \times U)$ ,

$$\int_{I \times U} u^* v = \int_{I \times U} (|U|^{-\frac{2}{d}} uv + \nabla u \cdot \nabla v + \nabla \Delta_U^{-1} \partial_t u \cdot \nabla \Delta_U^{-1} \partial_t v),$$

and moreover, by testing this identity with  $v = u$ , we obtain

$$\|u\|_{H_{\text{par}}^1(I \times U)} \leq C \|u^*\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)}. \quad (3-26)$$

We set

$$w := \Delta_U^{-1} \partial_t u \quad \text{and} \quad w^* := |U|^{-\frac{2}{d}} u - \Delta u.$$

The estimate (3-25) follows from (3-26). For  $v \in H_{\text{par}}^1(I \times U)$  with compact support in  $I \times U$ , we have

$$\int_{I \times U} u^* v = \int_{I \times U} (w^* v + \partial_t w v).$$

Since  $u^* \in L^2(I \times U)$ , we can argue by density to infer that  $w^* + \partial_t w \in L^2(I \times U)$ . The identity above then implies (3-24).

If  $u^*$  vanishes outside of  $V \subseteq I \times U$ , then we select a smooth cutoff function  $\eta$  such that  $\eta \equiv 1$  on  $V$  and  $\eta \equiv 0$  outside of  $V'$ , and we write

$$u^* = \eta u^* = \eta(\partial_t w + w^*) = \partial_t(\eta w) + \eta w^* - w \partial_t \eta.$$

This decomposition yields the second part of the statement, by standard comparisons of norms. □

*Proof of Proposition 3.10.* By scaling, we may fix  $n = 0$ . In order to localize an element  $S \in \mathcal{S}(\square_1)$  into an element of  $\mathcal{C}_0(\square_1)$  and thus be able to use the orthogonality property in the definition of the set  $\mathcal{S}(\square_1)$ , see (2-15), we introduce a version of the Helmholtz–Hodge decomposition of  $S$  which is adapted to the parabolic setting. In order to minimize difficulties due to boundary conditions, we work with functions which are periodic in the space variable. In the course of the proof, we will use the elementary variant of Proposition 2.1 for the standard heat operator with space-periodic boundary condition.

We decompose the proof into four steps. The first two steps are devoted to the construction of the Helmholtz–Hodge decomposition of  $S$ , and its estimation in relevant norms. The last step uses this representation to localize  $S$  to an element of  $\mathcal{C}_0(\square_1)$  and concludes the proof.

*Step 1.* We write  $S = (\nabla u, \mathbf{g}) \in \mathcal{S}(\square_1)$ . We recall that since  $\mathcal{S}(\square_1) \subseteq \mathcal{C}(\square_1)$ , we have  $\partial_t u = \nabla \cdot \mathbf{g}$ ; see (2-9). The function  $u$  is determined up to an additive constant, which we fix so that

$$\int_{\square_{3/4}} u = 0.$$

Let  $\eta \in C_c^\infty(\square_1)$  be a smooth function with compact support in  $\square_{3/4}$  such that  $0 \leq \eta \leq 1$  and  $\eta \equiv 1$  on  $\square_{1/2}$ . We set

$$\tilde{u} := \eta u, \quad \text{and} \quad \text{for all } j \in \{1, \dots, d\}, \quad \tilde{\mathbf{g}}_j := \eta \mathbf{g}_j.$$

For each  $j \in \{1, \dots, d\}$ , let  $T_{0j} \in H_{\text{par},\#}^1(\square_1)$  be the unique solution of

$$\begin{cases} (\partial_t - \Delta)T_{0j} = \tilde{\mathbf{g}}_j - \partial_{x_j} \tilde{u} & \text{in } \square_1, \\ T_{0j} = 0 & \text{on } \{-\frac{9}{2}\} \times \square_1. \end{cases} \tag{3-27}$$

By Lemma 3.11 there exist  $(w_j, w_j^*) \in L^2(I_1; H_0^1(\square_1)) \times L^2(I_1; H^{-1}(\square_1))$  which vanish in a neighborhood of  $\partial \square_1$  and satisfy

$$\tilde{\mathbf{g}}_j - \partial_{x_j} \tilde{u} = \partial_t w_j + w_j^*,$$

with

$$\|w_j\|_{L^2(I_1; H^1(\square_1))} + \|w_j^*\|_{L^2(I_1; H^{-1}(\square_1))} \leq C \|\tilde{\mathbf{g}}_j - \partial_{x_j} \tilde{u}\|_{\hat{H}_{\text{par}}^{-1}(\square_1)}.$$

Since  $(w_j, w_j^*)$  vanish in a neighborhood of  $\partial \square_1$ , we can interpret this pair as an element of

$$L^2(I_1; H_{\#}^1(\square_1)) \times L^2(I_1; H_{\#}^{-1}(\square_1)),$$

with the estimate

$$\|w_j\|_{L^2(I_1; H_{\#}^1(\square_1))} + \|w_j^*\|_{L^2(I_1; H_{\#}^{-1}(\square_1))} \leq C \|\tilde{\mathbf{g}}_j - \partial_{x_j} \tilde{u}\|_{\hat{H}_{\text{par}}^{-1}(\square_1)}.$$

Since  $\tilde{\mathbf{g}}_j - \partial_{x_j} \tilde{u} \in L^2(\square_1)$ , it is clear that  $\partial_t w_j \in L^2(I_1; H_{\#}^{-1}(\square_1))$ , and we therefore deduce that  $w_j \in H_{\#, \text{par}, \sqcup}^1(\square_1)$ . Moreover, by Proposition 2.1 applied to the standard heat operator, there exist a constant  $C(d) < \infty$  and  $T'_{0j} \in H_{\#, \text{par}, \sqcup}^1(\square_1)$  such that

$$(\partial_t - \Delta)T'_{0j} = -\Delta w_j + w_j^*,$$

with

$$\begin{aligned} \|T'_{0j}\|_{L^2(\square_1)} + \|\nabla T'_{0j}\|_{L^2(\square_1)} &\leq C(\|w_j\|_{L^2(I_1; H^1(\square_1))} + \|w_j^*\|_{L^2(I_1; H^{-1}(\square_1))}) \\ &\leq C \|\tilde{\mathbf{g}}_j - \partial_{x_j} \tilde{u}\|_{\hat{H}_{\text{par}}^{-1}(\square_1)}. \end{aligned}$$

We thus have

$$(\partial_t - \Delta)(w_j + T'_{0j} - T_{0j}) = 0,$$

with  $w_j + T'_{0j} - T_{0j} \in H_{\#, \text{par}, \sqcup}^1(\square_1)$ . Therefore,

$$T_{0j} = w_j + T'_{0j},$$

and

$$\|T_{0j}\|_{L^2(\square_1)} + \|\nabla T_{0j}\|_{L^2(\square_1)} \leq C \|\tilde{\mathbf{g}}_j - \partial_{x_j} \tilde{u}\|_{\hat{H}_{\text{par}}^{-1}(\square_1)}.$$

It is clear that  $\|\tilde{\mathbf{g}}_j\|_{\hat{H}_{\text{par}}^{-1}(\square_1)} \leq C \|\mathbf{g}_j\|_{H_{\text{par}}^{-1}(\square_1)}$ . By Proposition 3.7, Remark 3.8 and the comparison

$$\|\tilde{u}\|_{\hat{H}_{\text{par}}^{-1}(\square_1)} \leq \|\tilde{u}\|_{L^2(\square_1)} \leq \|u\|_{L^2(\square_{3/4})} = \|u - (u)_{\square_{3/4}}\|_{L^2(\square_{3/4})},$$

we obtain

$$\|T_{0j}\|_{L^2(\square_1)} + \|\nabla T_{0j}\|_{L^2(\square_1)} \leq C(\|\nabla u\|_{\hat{H}_{\text{par}}^{-1}(\square_1)} + \|\mathbf{g}\|_{\hat{H}_{\text{par}}^{-1}(\square_1)}). \quad (3-28)$$

*Step 2.* For each  $i, j \in \{1, \dots, d\}$ , we define  $T_{ij} \in H_{\text{par}, \#}^1(\square_1)$  as the solution of

$$\begin{cases} (\partial_t - \Delta)T_{ij} = \partial_{x_i} \tilde{\mathbf{g}}_j - \partial_{x_j} \tilde{\mathbf{g}}_i & \text{in } \square_1, \\ T_{ij} = 0 & \text{on } \{-\frac{9}{2}\} \times \square_1. \end{cases}$$

The solution  $T_{ij}$  is well-defined since the right-hand side belongs to  $L^2(I_1; H_{\#}^{-1}(\square_1))$ . We now estimate the  $L^2$  norm of  $T_{ij}$  using Lemma 3.9 and duality. We define  $\phi_{ij} \in H_{\text{par}, \#}^1(\square_1)$  to be the solution of the backwards heat equation

$$\begin{cases} -(\partial_t + \Delta)\phi_{ij} = T_{ij} & \text{in } \square_1, \\ \phi_{ij} = 0 & \text{on } \{\frac{9}{2}\} \times \square_1. \end{cases} \quad (3-29)$$

By Lemma 3.9, we have

$$\|\nabla\phi_{ij}\|_{H^1_{\text{par},\#}(\square_1)} \leq C\|T_{ij}\|_{L^2(\square_1)}.$$

Testing (3-29) against  $T_{ij}$ , we get

$$\|T_{ij}\|_{L^2(\square_1)}^2 = \int_{\square_1} (\tilde{\mathbf{g}}_i \partial_{x_j} \phi_{ij} - \tilde{\mathbf{g}}_j \partial_{x_i} \phi_{ij}) \leq \|\nabla\phi_{ij}\|_{H^1_{\text{par},\#}(\square_1)} \|\tilde{\mathbf{g}}\|_{\hat{H}^{-1}_{\text{par}}(\square_1)}.$$

Combining the two previous displays yields

$$\|T_{ij}\|_{L^2(\square_1)} \leq C\|\tilde{\mathbf{g}}\|_{\hat{H}^{-1}_{\text{par}}(\square_1)} \leq \|\mathbf{g}\|_{\hat{H}^{-1}_{\text{par}}(\square_1)}. \tag{3-30}$$

*Step 3.* For notational convenience, for each  $j \in \{1, \dots, d\}$ , we set  $T_{j0} := -T_{0j}$ ,  $\partial_0 := \partial_t$ ,  $\partial_j := \partial_{x_j}$ ,

$$R_0 := \tilde{u} - \sum_{j=1}^d \partial_j T_{0j},$$

and, for each  $i \in \{1, \dots, d\}$ ,

$$R_i := \tilde{\mathbf{g}}_i - \sum_{j=0}^d \partial_j T_{ij}.$$

In this step, we show that

$$\|R_0\|_{L^2(\square_1)} + \|\nabla R_0\|_{L^2(\square_1)} \leq C\|\nabla u\|_{\hat{H}^{-1}_{\text{par}}(\square_1)} + C\|\mathbf{g}\|_{\hat{H}^{-1}_{\text{par}}(\square_1)}, \tag{3-31}$$

and that for every  $i \in \{1, \dots, d\}$ ,

$$\|R_i\|_{L^2(\square_1)} \leq C\|\nabla u\|_{\hat{H}^{-1}_{\text{par}}(\square_1)} + C\|\mathbf{g}\|_{\hat{H}^{-1}_{\text{par}}(\square_1)}. \tag{3-32}$$

Recalling that  $\partial_t u = \nabla \cdot \mathbf{g}$ , we note that

$$(\partial_t - \Delta)R_0 = (\partial_t - \Delta)\tilde{u} - \sum_{j=1}^d \partial_j (\tilde{\mathbf{g}}_j - \partial_j \tilde{u}) = u \partial_t \eta + \mathbf{g} \cdot \nabla \eta,$$

and, for each  $i \in \{1, \dots, d\}$ ,

$$\begin{aligned} (\partial_t - \Delta)R_i &= (\partial_t - \Delta)\tilde{\mathbf{g}}_i - \partial_t (\tilde{\mathbf{g}}_i - \partial_i \tilde{u}) - \sum_{j=1}^d \partial_j (\partial_i \tilde{\mathbf{g}}_j - \partial_j \tilde{\mathbf{g}}_i) \\ &= (\partial_i u)(\partial_t \eta) + (\partial_t u)(\partial_i \eta) + (\nabla \cdot \mathbf{g})(\partial_i \eta) + (\partial_i \mathbf{g}) \cdot \nabla \eta. \end{aligned}$$

Moreover, it is clear from their definitions that  $T_{0j}$  and  $T_{ij}$  vanish in a neighborhood of the initial time slice  $\{-\frac{\rho}{2}\} \times \square_1$ . The estimates (3-31) and (3-32) are thus obtained by following the steps to the derivations of (3-28) and (3-30) respectively.

*Step 4.* We now select a cutoff function  $\chi \in C_c^\infty(\square_1)$  such that  $\chi \equiv 1$  on  $\square_0$  and  $\chi \equiv 0$  outside of  $\square_{1/2}$ , and observe that

$$\begin{pmatrix} \sum_{j=1}^d \nabla \partial_j (\chi^4 T_{0j}) \\ \sum_{j=0}^d \partial_j (\chi^4 T_{ij}) \end{pmatrix} \in \mathcal{C}_0(\square_1),$$

where we understand that the second component above denotes a  $d$ -dimensional vector field with components indexed by  $i \in \{1, \dots, d\}$ . By the definition of  $\mathcal{S}(\square_1)$ , we deduce that

$$\int_{\square_1} \begin{pmatrix} \sum_{j=1}^d \nabla \partial_j (\chi^4 T_{0j}) \\ \sum_{j=0}^d \partial_j (\chi^4 T_{ij}) \end{pmatrix} \cdot \mathbf{A} \begin{pmatrix} \nabla u \\ \mathbf{g} \end{pmatrix} = 0. \quad (3-33)$$

In the display above, the first vector is of dimension  $2d$ : the gradient appearing on the first row carries the first  $d$  components, while the other  $d$  components are represented by the second row and indexed by  $i \in \{1, \dots, d\}$ . Applying the chain rule in the identity (3-33) yields a number of terms, one of which is

$$\begin{aligned} \int_{\square_1} \chi^4 \begin{pmatrix} \sum_{j=1}^d \nabla \partial_j T_{0j} \\ \sum_{j=0}^d \partial_j T_{ij} \end{pmatrix} \cdot \mathbf{A} \begin{pmatrix} \nabla u \\ \mathbf{g} \end{pmatrix} &= \int_{\square_1} \chi^4 \begin{pmatrix} \nabla \tilde{u} \\ \tilde{\mathbf{g}} \end{pmatrix} \cdot \mathbf{A} \begin{pmatrix} \nabla u \\ \mathbf{g} \end{pmatrix} + \int_{\square_1} \chi^4 \begin{pmatrix} \nabla R_0 \\ R_i \end{pmatrix} \cdot \mathbf{A} \begin{pmatrix} \nabla u \\ \mathbf{g} \end{pmatrix} \\ &= \int_{\square_1} \chi^4 \begin{pmatrix} \nabla u \\ \mathbf{g} \end{pmatrix} \cdot \mathbf{A} \begin{pmatrix} \nabla u \\ \mathbf{g} \end{pmatrix} + \int_{\square_1} \chi^4 \begin{pmatrix} \nabla R_0 \\ R_i \end{pmatrix} \cdot \mathbf{A} \begin{pmatrix} \nabla u \\ \mathbf{g} \end{pmatrix}. \end{aligned}$$

We are interested in estimating the first term in this sum. By the uniform boundedness of  $\mathbf{A}$ , the absolute value of the second term in this sum is bounded by a constant times

$$\left( \|\nabla R_0\|_{L^2(\square_1)} + \sum_{i=1}^d \|R_i\|_{L^2(\square_1)} \right) \left( \int_{\square_1} \chi^4 \begin{pmatrix} \nabla u \\ \mathbf{g} \end{pmatrix} \cdot \mathbf{A} \begin{pmatrix} \nabla u \\ \mathbf{g} \end{pmatrix} \right)^{\frac{1}{2}}.$$

When applying the chain rule in the identity (3-33), the leftover terms are

$$4 \int_{\square_1} \chi^2 \begin{pmatrix} \sum_{j=1}^d T_{0j} (3\nabla \chi \partial_j \chi + \chi \nabla \partial_j \chi) + 2\nabla T_{0j} \chi \partial_j \chi \\ \sum_{j=0}^d T_{ij} \chi \partial_j \chi \end{pmatrix} \cdot \mathbf{A} \begin{pmatrix} \nabla u \\ \mathbf{g} \end{pmatrix}.$$

Using once more the uniform boundedness of  $\mathbf{A}$ , we obtain that the absolute value of the quantity above is bounded by a constant times

$$\left( \sum_{j=1}^d (\|T_{0j}\|_{L^2(\square_1)} + \|\nabla T_{0j}\|_{L^2(\square_1)}) + \sum_{j=0}^d \|T_{ij}\|_{L^2(\square_1)} \right) \left( \int_{\square_1} \chi^4 \begin{pmatrix} \nabla u \\ \mathbf{g} \end{pmatrix} \cdot \mathbf{A} \begin{pmatrix} \nabla u \\ \mathbf{g} \end{pmatrix} \right)^{\frac{1}{2}}.$$

Combining the previous displays with the estimates (3-28), (3-30), (3-31) and (3-32), we arrive at

$$\left( \int_{\square_1} \chi^4 \begin{pmatrix} \nabla u \\ \mathbf{g} \end{pmatrix} \cdot \mathbf{A} \begin{pmatrix} \nabla u \\ \mathbf{g} \end{pmatrix} \right)^{\frac{1}{2}} \leq C (\|\nabla u\|_{\widehat{H}_{\text{par}}^{-1}(\square_1)} + \|\mathbf{g}\|_{\widehat{H}_{\text{par}}^{-1}(\square_1)}).$$

By the uniform ellipticity of  $\mathbf{A}$ , the left side is an upper bound for  $\|S\|_{L^2(\square_0)}$ , up to a multiplicative constant, and therefore the proof is complete.  $\square$

#### 4. Convergence of subadditive quantities

We obtain an algebraic rate of convergence for the limits of the subadditive quantities by adapting the approach of [Armstrong and Smart 2016; Armstrong and Mourrat 2016], following the presentation of [Armstrong et al. 2017b, Chapter 2].

We let  $\bar{A}$  be the  $2d$ -by- $2d$  matrix characterized by the limit

$$\lim_{n \rightarrow \infty} \mathbb{E}[\mu(\square_n, X)] = \frac{1}{2} X \cdot \bar{A} X. \tag{4-1}$$

Note that the existence of the limit on the left side follows from the subadditivity of  $\mu(\cdot, X) = J(\cdot, X, 0)$  and stationarity, which together ensure that  $\mathbb{E}[\mu(\square_n, X)]$  is a nonincreasing sequence. The fact that  $X \mapsto \mu(\square_n, X)$  is quadratic ensures that the limit is also quadratic in  $X$  and can therefore be represented by a matrix. Moreover, by Lemma 2.4, there exists  $C(\Lambda) < \infty$  such that

$$\frac{1}{C} I_{2d} \leq \bar{A} \leq C I_{2d}, \tag{4-2}$$

where  $I_{2d}$  is the  $2d$ -by- $2d$  identity matrix. It is convenient to define

$$\bar{J}(X, X^*) := \frac{1}{2} X \cdot \bar{A} X + \frac{1}{2} X^* \cdot \bar{A}^{-1} X^* - X \cdot X^*.$$

The goal of this section is to prove the following theorem.

**Theorem 4.1** (convergence of  $J$ ). *There exist an exponent  $\beta(d, \Lambda) > 0$  and, for each  $s \in (0, 2 + d)$ , a constant  $C(s, d, \Lambda) < \infty$  such that, for every  $X, X^* \in B_1$  and  $n \in \mathbb{N}$ , we have*

$$|J(\square_n, X, X^*) - \bar{J}(X, X^*)| \leq C 3^{-n\beta(2+d-s)} + \mathcal{O}_1(C 3^{-ns}). \tag{4-3}$$

The next lemma, which should be compared to [Armstrong et al. 2017b, Lemma 2.7], allows us to reduce Theorem 4.1 to an estimate on the quantity  $J(\square_n, X, \bar{A}X)$ . Note that, in view of Lemma 2.3, a control on the size of  $\inf_{X^*} J(V, X, X^*)$  can be interpreted as information on the ‘‘convex duality defect’’ between the quantities  $\mu$  and  $\mu^*$ , quantifying how close these functions are to a convex dual pair.

**Lemma 4.2** (reduction to minimal set). *For each  $\Gamma \geq 1$ , there exists a constant  $C(\Gamma, d, \Lambda) < \infty$  such that, for every  $2d$ -by- $2d$  symmetric matrix  $\tilde{A}$  satisfying*

$$\Gamma^{-1} I_{2d} \leq \tilde{A} \leq \Gamma I_{2d} \tag{4-4}$$

and every parabolic cylinder  $V \subseteq \mathbb{R}^{d+1}$ , we have

$$\sup_{X, X^* \in B_1} \left| J(V, X, X^*) - \left( \frac{1}{2} X \cdot \tilde{A} X + \frac{1}{2} X^* \cdot \tilde{A}^{-1} X^* - X \cdot X^* \right) \right| \leq C \sup_{X \in B_1} (J(V, X, \tilde{A}X))^{\frac{1}{2}}. \tag{4-5}$$

*Proof.* Since the domain  $V$  plays no role in the argument, we drop the explicit dependence on  $V$ . Define

$$\delta^2 := \sup_{X \in B_1} J(X, \tilde{A}X).$$

To avoid a conflict in the notation, we denote the Legendre–Fenchel transform (convex dual function) of  $\mu$  by

$$H(X^*) := \sup_{X \in \mathbb{R}^{2d}} (X \cdot X^* - \mu(X)).$$

It is clear from (2-14) that

$$H(X^*) \leq \mu^*(X^*).$$

Thus, by (2-18), for every  $X \in B_1$ ,

$$0 \leq \mu(X) + H(\tilde{A}X) - X \cdot \tilde{A}X \leq \mu(X) + \mu^*(\tilde{A}X) - X \cdot \tilde{A}X = J(X, \tilde{A}X) \leq \delta^2. \tag{4-6}$$

This implies that, for every  $X \in B_1$ ,

$$|\mu^*(\tilde{A}X) - H(\tilde{A}X)| \leq \delta^2. \quad (4-7)$$

For each  $X \in \mathbb{R}^{2d}$ , the minimum of the map  $X^* \mapsto \mu(X) + H(X^*) - X \cdot X^*$  is zero and it is achieved at  $X^*$  for which  $X^* = \nabla\mu(X)$ . By uniform convexity (quadratic response) and (4-6), we deduce that, for every  $X \in B_1$ ,

$$|\tilde{A}X - \nabla\mu(X)|^2 \leq C\delta^2. \quad (4-8)$$

Using the expression

$$\mu(X) = \frac{1}{2}X \cdot \nabla\mu(X)$$

we obtain, for every  $X \in B_1$ ,

$$\left| \frac{1}{2}X \cdot \tilde{A}X - \mu(X) \right| \leq C\delta.$$

From this, uniform convexity and (4-4), we obtain, for every  $X^* \in B_1$ ,

$$\left| \frac{1}{2}X^* \cdot \tilde{A}^{-1}X^* - H(X^*) \right| \leq C\delta.$$

Hence by (4-7), (4-4) again,

$$\left| \frac{1}{2}X^* \cdot \tilde{A}^{-1}X^* - \mu^*(X^*) \right| \leq C\delta.$$

The formula (2-18) now yields the lemma.  $\square$

We decompose the estimate for  $J(\square_n, X, \bar{A}X)$  into three steps. In the first step, we identify a convenient finite-volume approximation of the homogenized matrix  $\bar{A}$ . We next control the expectation of  $J(\square_n, X, \bar{A}X)$  in Section 4B. We finally use the subadditivity of  $J$  in Section 4C to deduce a control of the fluctuations of  $J(\square_n, X, \bar{A}X)$ , and complete the proof of Theorem 4.1.

**4A. The coarsened mapping.** Recall that  $S(\cdot, V, X, X^*)$  denotes the unique maximizer in the definition of  $J(V, X, X^*)$ ; see (2-16). We let  $\bar{A}_V \in \mathbb{R}^{2d \times 2d}$  be the symmetric matrix such that, for every  $X^* \in \mathbb{R}^{2d}$ ,

$$\mathbb{E}[J(V, 0, X^*)] = \frac{1}{2}X^* \cdot \bar{A}_V^{-1}X^*.$$

By (2-24), there exists  $C(d, \Lambda) < \infty$  such that

$$\frac{1}{C}I_{2d} \leq \bar{A}_V \leq C I_{2d},$$

and by (2-29),

$$\mathbb{E}\left[ \int_V S(\cdot, V, 0, X^*) \right] = \bar{A}_V^{-1}X^*.$$

Recalling also (2-30) and the linearity of the mapping  $(X, X^*) \mapsto S(\cdot, V, X, X^*)$ , we thus see that the matrix  $\bar{A}_V$  is such that, for every  $X \in \mathbb{R}^{2d}$ ,

$$\mathbb{E}\left[ \int_V S(\cdot, V, X, \bar{A}_V X) \right] = 0. \quad (4-9)$$

We note that by Lemmas 2.3 and 2.4, for each  $X \in \mathbb{R}^{2d}$ , the mapping

$$X^* \mapsto \mathbb{E}[J(V, X, X^*)] = \mathbb{E}[\mu(V, X)] + \mathbb{E}[\mu^*(V, X^*)] - X \cdot X^*$$

is uniformly convex, and achieves its unique minimum at  $X^*$  satisfying

$$\mathbb{E}[\nabla_{X^*} \mu^*(V, X^*)] = X.$$

Moreover, the latter condition is equivalent to  $X^* = \bar{A}_V X$ . We thus deduce that for every  $X, X^* \in \mathbb{R}^{2d}$ ,

$$\mathbb{E}[J(V, X, \bar{A}_V X)] \leq \mathbb{E}[J(V, X, X^*)] \leq \mathbb{E}[J(V, X, \bar{A}_V X)] + C|X^* - \bar{A}_V X|^2. \tag{4-10}$$

We use the shorthand notation

$$\bar{A}_n := \bar{A}_{\square_n}. \tag{4-11}$$

**4B. Control of the expectation of  $J$ .** The goal of this subsection is to prove the following proposition.

**Proposition 4.3** (decay of  $\mathbb{E}[J]$ ). *There exist  $\beta(d, \Lambda) > 0$  and  $C(d, \Lambda) < \infty$  such that, for every  $n \in \mathbb{N}$  and  $X \in B_1(\mathbb{R}^{2d})$ ,*

$$\mathbb{E}[J(\square_n, X, \bar{A}X)] \leq C3^{-\beta n}. \tag{4-12}$$

The main step to prove this result is to control the size of  $J$  near  $(X, \bar{A}X)$  in terms of the expected ‘‘additivity defect’’ of  $J$  between successive triadic scales. We measure the latter using the quantity

$$\tau_n := \sup_{X, X^* \in B_1} (\mathbb{E}[J(\square_n, X, X^*)] - \mathbb{E}[J(\square_{n+1}, X, X^*)]). \tag{4-13}$$

**Proposition 4.4.** *There exist  $\alpha(d) < \infty$  and  $C(d, \Lambda) < \infty$  such that, for every  $n \in \mathbb{N}$  and  $X \in B_1(\mathbb{R}^{2d})$ ,*

$$\mathbb{E}[J(\square_n, X, \bar{A}_n X)] \leq C 3^{-\alpha n} \left( 1 + \sum_{k=0}^{n-1} 3^{\alpha k} \tau_k \right).$$

As will be explained below, Proposition 4.3 follows from Proposition 4.4 by iteration, in analogy with an ODE argument. We focus for now on the proof of Proposition 4.4, and start by rewriting the quadratic response (2-27) in a more convenient form.

**Lemma 4.5** (quadratic response). *There exists a constant  $C(d, \Lambda) < \infty$  such that the following holds. Let  $V, V_1, \dots, V_k$  be parabolic cylinders such that  $\{V_1, \dots, V_k\}$  forms a partition of  $V$ , up to a set of null measure. For every  $X, X^* \in \mathbb{R}^{2d}$ , we have*

$$\sum_{j=1}^k \frac{|V_j|}{|V|} \|S(\cdot, V, X, X^*) - S(\cdot, V_j, X, X^*)\|_{L^2(V_j)}^2 \leq C \sum_{j=1}^k \frac{|V_j|}{|V|} (J(V_j, X, X^*) - J(V, X, X^*)).$$

*Proof.* Define  $T := S(\cdot, V, X, X^*)$ . Applying (2-27) on the subdomain  $V_j$  for each  $j \in \{1, \dots, k\}$ , we get

$$\frac{1}{C} \int_{V_j} |T - S(\cdot, V_j, X, X^*)|^2 \leq |V_j| J(V_j, X, X^*) - \int_{V_j} \left( -\frac{1}{2} T \cdot AT - X \cdot AT + X^* \cdot T \right).$$

Summing over  $j$  and recalling (2-16) yields the result. □

We next show that the spatial averages of  $S$  can be controlled by an expression involving the additivity defects of  $J$  on all smaller length scales. We define

$$\bar{S}_n(X, X^*) := \mathbb{E} \left[ \int_{\square_n} S(\cdot, \square_n, X, X^*) \right].$$



**Lemma 4.6.** *There exist  $\alpha(d) < \infty$  and  $C(d, \Lambda) < \infty$  such that, for every  $n \in \mathbb{N}$  and  $X, X^* \in B_1(\mathbb{R}^{2d})$ , we have*

$$\mathbb{E} \left[ \left| \int_{\square_n} S(\cdot, \square_n, X, X^*) - \bar{S}_n(X, X^*) \right|^2 \right] \leq C 3^{-\alpha n} \left( 1 + \sum_{k=0}^{n-1} 3^{\alpha k} \tau_k \right). \quad (4-14)$$

*Proof.* For any  $X' \in B_1$ ,  $m \leq n$  and  $z \in \mathcal{Z}_m$ , the first variation (2-26) gives

$$\begin{aligned} X' \cdot \int_{z+\square_m} (S(\cdot, \square_n, X, X^*) - S(\cdot, z+\square_m, X, X^*)) \\ = \int_{z+\square_m} S(\cdot, z+\square_m, 0, X') \cdot A(S(\cdot, \square_n, X, X^*) - S(\cdot, z+\square_m, X, X^*)). \end{aligned}$$

Averaging over  $z \in \mathcal{Z}_m$  and using the Cauchy–Schwarz inequality yields

$$\begin{aligned} |\mathcal{Z}_m|^{-1} \cdot \left| X' \cdot \sum_{z \in \mathcal{Z}_m} \int_{z+\square_m} (S(\cdot, \square_n, X, X^*) - S(\cdot, z+\square_m, X, X^*)) \right| \\ \leq \left( |\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} \int_{z+\square_m} |A S(\cdot, z+\square_m, 0, X')|^2 \right)^{\frac{1}{2}} \\ \cdot \left( |\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} \int_{z+\square_m} |S(\cdot, \square_n, X, X^*) - S(\cdot, z+\square_m, X, X^*)|^2 \right)^{\frac{1}{2}}. \quad (4-15) \end{aligned}$$

The first term on the right side is bounded by a constant  $C(d, \Lambda) < \infty$ . We use Lemma 4.5 to bound the second term and obtain

$$\begin{aligned} \left| |\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} \int_{z+\square_m} |S(\cdot, \square_n, X, X^*) - S(\cdot, z+\square_m, X, X^*)|^2 \right| \\ \leq \frac{C}{|\mathcal{Z}_m|} \sum_{z \in \mathcal{Z}_m} (J(z+\square_m, X, X^*) - J(\square_n, X, X^*)). \end{aligned}$$

Now, we can estimate the variance of  $\int_{\square_n} S(\cdot, \square_n, X, X^*)$  using those at scale  $m$ :

$$\begin{aligned} \text{var} \left[ \int_{\square_n} S(\cdot, \square_n, X, X^*) \right] \leq 2 \text{var} \left[ 3^{-(d+2)(n-m)} \sum_{z \in \mathcal{Z}_m} \int_{z+\square_m} S(\cdot, z+\square_m, X, X^*) \right] \\ + C \mathbb{E}[J(\square_m, X, X^*) - J(\square_n, X, X^*)]. \end{aligned}$$

For  $m, n \in \mathbb{N}$ ,  $m \leq n$ , we can decompose  $\mathcal{Z}_m$  into a union of  $2^{d+1}$  “checkerboard” subsets  $\mathcal{Z}^{(1)}, \dots, \mathcal{Z}^{(2^{1+d})}$  to ensure that for each  $i \in \{1, \dots, 2^{1+d}\}$ ,

$$(z, z') \in \mathcal{Z}^{(i)} \implies \text{dist}(z+\square_m, z'+\square_m) \geq 1.$$

For example, to any  $i \in \{1, \dots, 2^{d+1}\}$  we can associate  $(i_0, \dots, i_d) \in \{0, 1\}^{d+1}$ , and then set

$$\mathcal{Z}^{(i)} := ((i_0 3^{2m}, i_1 3^m, \dots, i_d 3^m) + 2((3^{2m} \mathbb{Z}) \times (3^m \mathbb{Z}^d))) \cap \square_n. \quad (4-16)$$

Thus, we obtain the bound

$$\text{var} \left[ \sum_{z \in \mathcal{Z}_m} \int_{z+\square_m} S(\cdot, z+\square_m, X, X^*) \right] \leq C(d) \sum_{i=1}^{2^{d+1}} \text{var} \left[ \sum_{z \in \mathcal{Z}^{(i)}} \int_{z+\square_m} S(\cdot, z+\square_m, X, X^*) \right],$$

and by independence at distance larger than 1 and stationarity,

$$\text{var} \left[ \sum_{z \in \mathbb{Z}_m} \int_{z + \square_m} S(\cdot, z + \square_m, X, X^*) \right] \leq C 3^{(d+2)(n-m)} \text{var} \left[ \int_{\square_m} S(\cdot, \square_m, X, X^*) \right].$$

We can now estimate the variance of the spatial average of  $S$  at scale  $n$  by the variance at smaller scales:

$$\text{var} \left[ \int_{\square_n} S(\cdot, \square_n, X, X^*) \right] \leq C 3^{-(d+2)(n-m)} \text{var} \left[ \int_{\square_m} S(\cdot, \square_m, X, X^*) \right] + C \mathbb{E}[J(\square_m, X, X^*) - J(\square_n, X, X^*)]. \quad (4-17)$$

Selecting  $\ell$  to be the smallest integer such that  $C 3^{-(d+2)\ell} \leq \frac{1}{3}$ , we get

$$\text{var} \left[ \int_{\square_{m+\ell}} S(\cdot, \square_{m+\ell}, X, X^*) \right] \leq \frac{1}{3} \text{var} \left[ \int_{\square_m} S(\cdot, \square_m, X, X^*) \right] + C \mathbb{E}[J(\square_m, X, X^*) - J(\square_{m+\ell}, X, X^*)].$$

We introduce

$$u_n := \text{var} \left[ \int_{\square_n} S(\cdot, \square_n, X, X^*) \right],$$

and  $v_n := u_{n\ell}$ . We have

$$v_n \leq \frac{1}{3} v_{n-1} + C \sum_{k=(n-1)\ell}^{n\ell-1} \tau_k,$$

and, by induction,

$$v_n \leq 3^{-n} u_0 + C \sum_{i=1}^n 3^{-i} \sum_{k=(n-i)\ell}^{(n-i+1)\ell-1} \tau_k.$$

Defining  $\alpha := 1/\ell$  (recall that  $\ell$  only depends on  $d$ ), we get

$$v_n \leq C \left( 3^{-n} + \sum_{k=0}^{n\ell-1} 3^{-\alpha(n\ell-k)} \tau_k \right).$$

Thus, if  $n$  is a multiple of  $\ell$ , we have

$$u_n \leq C \left( 3^{-\alpha n} + \sum_{k=0}^{n-1} 3^{-\alpha(n-k)} \tau_k \right),$$

and for  $n = \ell n' + m$ , with  $0 \leq m < \ell$ , another application of (4-17) gives the same estimate, so finally

$$\text{var} \left[ \int_{\square_n} S(\cdot, \square_n, X, X^*) \right] \leq C \left( 3^{-\alpha n} + \sum_{k=0}^{n-1} 3^{-\alpha(n-k)} \tau_k \right),$$

which is (4-14). □

We can now sum the scales and deduce that  $S(\cdot, \square_n, X, X^*)$  is close to a constant in a weak sense, provided that a weighted norm of  $(\tau_k)_{k < n}$  is small.

**Lemma 4.7** (weak control of  $S$ ). *There exist  $\alpha(d) < \infty$  and  $C(d, \Lambda) < \infty$  such that, for every  $n \in \mathbb{N}$  and  $X, X^* \in B_1(\mathbb{R}^{2d})$ ,*

$$\mathbb{E}\left[\|S(\cdot, \square_n, X, X^*) - \bar{S}_n(X, X^*)\|_{\bar{H}_{\text{par}}^{-1}(\square_n)}^2\right] \leq C3^{(2-\alpha)n} \left(1 + \sum_{k=0}^{n-1} 3^{\alpha k} \tau_k\right).$$

*Proof.* We decompose the proof into three steps.

*Step 1.* To begin, we show that there exists a constant  $C(d, \Lambda) < \infty$  such that, for every  $m, n \in \mathbb{N}$ ,  $m \leq n$ , and  $X, X^* \in B^1(\mathbb{R}^{2d})$ ,

$$|\bar{S}_n(X, X^*) - \bar{S}_m(X, X^*)|^2 \leq C \sum_{k=m}^{n-1} \tau_k. \quad (4-18)$$

Indeed, recalling the definition of  $\mathcal{Z}_m$  in (3-8) (which depends implicitly on  $n$ ), we use Jensen's inequality, Lemma 4.5 and stationarity to get

$$\begin{aligned} & \left| \mathbb{E} \left[ \int_{\square_n} S(\cdot, \square_n, X, X^*) - |\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} \int_{z + \square_m} S(\cdot, z + \square_m, X, X^*) \right] \right|^2 \\ & \leq \mathbb{E} \left[ |\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} \int_{z + \square_m} \left| S(\cdot, \square_n, X, X^*) - S(\cdot, z + \square_m, X, X^*) \right|^2 \right] \\ & \leq C(J(\square_n, X, X^*) - J(\square_m, X, X^*)) \leq C \sum_{k=m}^{n-1} \tau_k, \end{aligned}$$

and this implies (4-18).

*Step 2.* In this step, we show that there exists a constant  $C(d, \Lambda) < \infty$  such that, for every  $m \in \mathbb{N}$ ,  $m \leq n$ ,

$$\mathbb{E} \left[ |\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} \left| (S(\cdot, \square_n, X, X^*))_{z + \square_m} - \bar{S}_n(X, X^*) \right|^2 \right] \leq C \left( 3^{-\alpha m} + \sum_{k=0}^m 3^{\alpha(k-m)} \tau_k + \sum_{k=m}^{n-1} \tau_k \right). \quad (4-19)$$

By Lemma 4.5, we have

$$\begin{aligned} & \sum_{z \in \mathcal{Z}_m} \|S(\cdot, \square_n, X, X^*) - S(\cdot, z + \square_m, X, X^*)\|_{L^2(z + \square_m)}^2 \\ & \leq C \sum_{z \in \mathcal{Z}_m} (J(\square_n, X, X^*) - J(z + \square_m, X, X^*)). \quad (4-20) \end{aligned}$$

Taking expectations, using stationarity and Jensen's inequality, we deduce that

$$\mathbb{E} \left[ \sum_{z \in \mathcal{Z}_m} \left| (S(\cdot, \square_n, X, X^*))_{z + \square_m} - (S(\cdot, z + \square_m, X, X^*))_{z + \square_m} \right|^2 \right] \leq C |\mathcal{Z}_m| \sum_{k=m}^{n-1} \tau_k. \quad (4-21)$$

Moreover, by stationarity and Lemma 4.6, we have

$$\mathbb{E} \left[ |\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} \left| (S(\cdot, z + \square_m, X, X^*))_{z + \square_m} - \bar{S}_m(X, X^*) \right|^2 \right] \leq C \left( 3^{-\alpha m} + \sum_{k=0}^m 3^{\alpha(k-m)} \tau_k \right). \quad (4-22)$$

Since

$$\begin{aligned} |(S(\cdot, \square_n, X, X^*))_{z+\square_m} - \bar{S}_n(X, X^*)|^2 &\leq 3|(S(\cdot, \square_n, X, X^*))_{z+\square_m} - (S(\cdot, z + \square_m, X, X^*))_{z+\square_m}|^2 \\ &\quad + 3|(S(\cdot, z + \square_m, X, X^*))_{z+\square_m} - \bar{S}_m(X, X^*)|^2 \\ &\quad + 3|\bar{S}_n(X, X^*) - \bar{S}_m(X, X^*)|^2, \end{aligned}$$

we obtain (4-19) by combining (4-21), (4-22) and (4-18).

*Step 3.* We now combine Proposition 3.6 with the result of the previous step to obtain

$$\|S(\cdot, \square_n, X, X^*) - \bar{S}_n(X, X^*)\|_{\underline{H}^{\text{par}}(\square_n)}^2 \leq C \left( 1 + \left( \sum_{m=0}^{n-1} 3^m Z_m^{\frac{1}{2}} \right)^2 \right), \tag{4-23}$$

where  $Z_m$  is a random variable satisfying

$$\mathbb{E}[Z_m] \leq C \left( 3^{-\alpha m} + \sum_{k=0}^m 3^{\alpha(k-m)} \tau_k + \sum_{k=m}^{n-1} \tau_k \right). \tag{4-24}$$

By Hölder’s inequality, we have

$$\left( \sum_{m=0}^{n-1} 3^m Z_m^{\frac{1}{2}} \right)^2 \leq \left( \sum_{m=0}^{n-1} 3^m \right) \left( \sum_{m=0}^{n-1} 3^m Z_m \right) \leq C 3^n \sum_{m=0}^{n-1} 3^m Z_m. \tag{4-25}$$

Taking expectations and using (4-24), we get

$$\mathbb{E} \left[ \left( \sum_{m=0}^{n-1} 3^m Z_m^{\frac{1}{2}} \right)^2 \right] \leq C 3^n \sum_{m=0}^{n-1} 3^m \left( 3^{-\alpha m} + \sum_{k=0}^m 3^{\alpha(k-m)} \tau_k + \sum_{k=m}^{n-1} \tau_k \right).$$

For the last two terms, we reverse the order of the sums to find

$$\sum_{m=0}^{n-1} 3^m \sum_{k=0}^m 3^{\alpha(k-m)} \tau_k = \sum_{k=0}^{n-1} 3^{\alpha k} \tau_k \sum_{m=k}^{n-1} 3^{(1-\alpha)m} \leq C 3^{(1-\alpha)n} \sum_{k=0}^{n-1} 3^{\alpha k} \tau_k,$$

and

$$\sum_{m=0}^{n-1} 3^m \sum_{k=m}^{n-1} \tau_k = \sum_{k=0}^{n-1} \sum_{m=0}^k 3^m \tau_k \leq C \sum_{k=0}^{n-1} 3^k \tau_k. \tag{4-26}$$

The second sum is bounded by the first; thus combining the above displays yields

$$\mathbb{E} \left[ \left( \sum_{m=0}^{n-1} 3^m Z_m^{\frac{1}{2}} \right)^2 \right] \leq C 3^{(2-\alpha)n} \left( 1 + \sum_{k=0}^{n-1} 3^{\alpha k} \tau_k \right). \quad \square$$

We next complete the proof of Proposition 4.4 and then of Proposition 4.3.

*Proof of Proposition 4.4.* According to Lemma 4.7, Proposition 3.10 and (4-9), we have

$$\mathbb{E} \left[ \|S(\cdot, \square_n, X, \bar{A}_n X)\|_{\underline{L}^2(\square_{n-1})}^2 \right] \leq C 3^{-\alpha n} \left( 1 + \sum_{k=0}^{n-1} 3^{\alpha k} \tau_k \right).$$

By Lemma 4.5, we deduce

$$\mathbb{E}[\|S(\cdot, \square_{n-1}, X, \bar{A}_n X)\|_{\underline{L}^2(\square_{n-1})}^2] \leq C3^{-\alpha n} \left(1 + \sum_{k=0}^{n-1} 3^{\alpha k} \tau_k\right).$$

Recall that  $(z + \square_{n-1})_{z \in \mathcal{Z}_{n-1}}$  is a partition of  $\square_n$ , up to a set of null measure. Moreover, by stationarity, the previous display implies that, for every  $z \in \mathcal{Z}_{n-1}$ ,

$$\mathbb{E}[\|S(\cdot, z + \square_{n-1}, X, \bar{A}_n X)\|_{\underline{L}^2(z + \square_{n-1})}^2] \leq C3^{-\alpha n} \left(1 + \sum_{k=0}^{n-1} 3^{\alpha k} \tau_k\right).$$

Applying Lemma 4.5 once more and summing over  $z \in \mathcal{Z}_{n-1}$ , we obtain the result.  $\square$

*Proof of Proposition 4.3.* We denote by  $\mathcal{B}$  the set of canonical basis elements of  $\mathbb{R}^{2d}$ , and observe that there exists a constant  $C(d) < \infty$  such that if  $X \mapsto B(X)$  is a nonnegative quadratic form over  $\mathbb{R}^{2d}$ , then

$$\sup_{X \in B_1} B(X) \leq C \sum_{X \in \mathcal{B}} B(X). \quad (4-27)$$

Indeed, a quadratic form is associated to a nonnegative symmetric matrix with largest eigenvalue bounded by its trace; this trace is equal to the right side above.

By the definition of  $\tau_n$ , see (4-13), and Lemma 2.3, we have

$$\tau_n \leq \sup_{X \in B_1} (\mathbb{E}[\mu(\square_n, X)] - \mathbb{E}[\mu(\square_{n+1}, X)]) + \sup_{X^* \in B_1} (\mathbb{E}[\mu^*(\square_n, X^*)] - \mathbb{E}[\mu^*(\square_{n+1}, X^*)]).$$

Since  $X \mapsto \mathbb{E}[\mu(V, X)]$  and  $X^* \mapsto \mathbb{E}[\mu(V, X^*)]$  are nonnegative quadratic forms, and since this property is stable under linear changes of coordinates, it follows from (4-27) that

$$\tau_n \leq C \sum_{X \in \mathcal{B}} (\mathbb{E}[\mu(\square_n, X)] - \mathbb{E}[\mu(\square_{n+1}, X)]) + C \sum_{X \in \mathcal{B}} (\mathbb{E}[\mu^*(\square_n, \bar{A}_n X)] - \mathbb{E}[\mu^*(\square_{n+1}, \bar{A}_n X)]),$$

and thus by Lemma 2.3,

$$\tau_n \leq C \sum_{X \in \mathcal{B}} (\mathbb{E}[J(\square_n, X, \bar{A}_n X)] - \mathbb{E}[J(\square_{n+1}, X, \bar{A}_n X)]).$$

By (4-10), we have

$$\mathbb{E}[J(\square_{n+1}, X, \bar{A}_{n+1} X)] \leq \mathbb{E}[J(\square_{n+1}, X, \bar{A}_n X)],$$

and therefore

$$\tau_n \leq C \sum_{X \in \mathcal{B}} (\mathbb{E}[J(\square_n, X, \bar{A}_n X)] - \mathbb{E}[J(\square_{n+1}, X, \bar{A}_{n+1} X)]). \quad (4-28)$$

This motivates the definition of

$$D_n := \sum_{X \in \mathcal{B}} \mathbb{E}[J(\square_n, X, \bar{A}_n X)].$$

Proposition 4.4 asserts that

$$D_n \leq C3^{-\alpha n} \left(1 + \sum_{k=0}^{n-1} 3^{\alpha k} \tau_k\right).$$

Setting

$$\tilde{D}_n := 3^{-\frac{\alpha}{2}n} \sum_{k=0}^n 3^{\frac{\alpha}{2}k} D_k,$$

we deduce that

$$\begin{aligned} \tilde{D}_n &\leq C 3^{-\frac{\alpha}{2}n} \sum_{m=0}^n 3^{-\frac{\alpha}{2}m} \left( 1 + \sum_{k=0}^m 3^{\alpha k} \tau_k \right) \\ &\leq C 3^{-\frac{\alpha}{2}n} + C 3^{-\frac{\alpha}{2}n} \sum_{k=0}^n \sum_{m=k}^n 3^{-\frac{\alpha}{2}m} 3^{\alpha k} \tau_k \leq C 3^{-\frac{\alpha}{2}n} \left( 1 + \sum_{k=0}^n 3^{\frac{\alpha}{2}k} \tau_k \right). \end{aligned} \tag{4-29}$$

Since  $D_0 \leq C$ , we also have

$$\tilde{D}_n - \tilde{D}_{n+1} \geq 3^{-\frac{\alpha}{2}n} \sum_{k=0}^n 3^{\frac{\alpha}{2}k} (D_k - D_{k+1}) - C 3^{-\frac{\alpha}{2}n}.$$

Combining this with (4-28) yields

$$\tilde{D}_n - \tilde{D}_{n+1} \geq C^{-1} 3^{-\frac{\alpha}{2}n} \sum_{k=0}^n 3^{\frac{\alpha}{2}k} \tau_k - C 3^{-\frac{\alpha}{2}n}.$$

From this and (4-29), we obtain that there exists an exponent  $\beta(d, \Lambda) \in (0, \frac{\alpha}{2})$  such that

$$\tilde{D}_{n+1} \leq 3^{-\beta} \tilde{D}_n + C 3^{-\frac{\alpha}{2}n};$$

introducing  $v_n := 3^{\beta n} \tilde{D}_n$  and multiplying the previous identity by  $3^{(\beta+1)n}$  gives

$$v_{n+1} \leq v_n + C 3^{(\beta-\frac{\alpha}{2})n}.$$

Summing this inequality over  $n$  yields

$$v_n \leq v_0 + \frac{C}{1 - (3^{\beta-\frac{\alpha}{2}})^n}.$$

That is,  $v$  is bounded; i.e.,

$$\tilde{D}_n \leq C 3^{-\beta n}.$$

By (4-28), we also obtain

$$\tau_n \leq C 3^{-\beta n}.$$

By the definition of  $\bar{A}_n$ , we have

$$|\bar{A}_n - \bar{A}_{n+1}| \leq C \tau_n,$$

so that, setting

$$\bar{A} := \lim_{n \rightarrow \infty} \bar{A}_n, \tag{4-30}$$

we get

$$|\bar{A}_n - \bar{A}| \leq \sum_{m=n}^{\infty} |\bar{A}_m - \bar{A}_{m+1}| \leq C \sum_{m=n}^{\infty} \tau_m \leq C 3^{-\beta n}.$$

Combining the last displays with (4-10) yields

$$\sup_{X \in B_1} \mathbb{E}[J(\square_n, X, \bar{A}X)] \leq C3^{-\beta n}.$$

By an application of Lemma 4.2, we can verify that the matrix  $\bar{A}$  defined in (4-30) coincides with that defined in (4-1). □

**4C. Control of the fluctuations of  $J$ .** In this subsection, we prove Theorem 4.1. In view of Lemma 4.2, the main point is to obtain a control on the fluctuations of  $J(\square_n, X, \bar{A}X)$ , which we do using subadditivity.

*Proof of Theorem 4.1. Step 1.* In this first step, we show that there exists an exponent  $\beta(d, \Lambda) > 0$  and a constant  $C(d, \Lambda) > 0$  such that, for every  $X \in B_1(\mathbb{R}^{2d})$  and  $m, n \in \mathbb{N}$ ,  $m \leq n$ , we have

$$3^{-(2+d)(n-m)} \log \mathbb{E}[\exp(C^{-1}3^{(2+d)(n-m)} J(\square_n, X, \bar{A}X))] \leq C3^{-\beta m}. \tag{4-31}$$

For  $m, n \in \mathbb{N}$ ,  $m \leq n$ , recall that the cube  $\square_n$  is partitioned into a union of  $2^{d+1}$  “checkerboard” subsets, see (4-16), to ensure that for each  $i \in \{1, \dots, 2^{1+d}\}$ ,

$$z, z' \in \mathcal{Z}^{(i)} \implies \text{dist}(z + \square_n, z' + \square_n) \geq 1.$$

In particular, for each fixed  $i \in \{1, \dots, 2^{1+d}\}$ , the random variables  $(z + \square_n)_{z \in \mathcal{Z}_m^{(i)}}$  are independent. By subadditivity, for each  $X \in B_1(\mathbb{R}^{2d})$  and  $t > 0$ , we have

$$\log \mathbb{E}[\exp(t3^{(2+d)(n-m)} J(\square_n, X, \bar{A}X))] \leq \log \mathbb{E} \left[ \exp \left( t \sum_{z \in \mathcal{Z}_m} J(z + \square_m, X, \bar{A}X) \right) \right],$$

and by Hölder’s inequality and independence, the latter is bounded by

$$\begin{aligned} &\leq 2^{-(1+d)} \sum_{i=1}^{2^{1+d}} \log \mathbb{E} \left[ \exp \left( t2^{1+d} \sum_{z \in \mathcal{Z}_m^{(i)}} J(z + \square_m, X, \bar{A}X) \right) \right] \\ &\leq 2^{-(1+d)} \sum_{z \in \mathcal{Z}_m} \log \mathbb{E}[\exp(t2^{1+d} J(z + \square_m, X, \bar{A}X))]. \end{aligned}$$

By stationarity, the summands above do not depend on  $z \in \mathcal{Z}_m$ . Since

$$J(\square_m, X, \bar{A}X) \leq C(d, \Lambda),$$

we can choose  $t(d, \Lambda) > 0$  sufficiently small and use the elementary inequalities

$$\begin{aligned} \exp(s) &\leq 1 + 2s \quad \text{for all } 0 \leq s \leq 1, \\ \log(1 + s) &\leq s \quad \text{for all } s \geq 0 \end{aligned}$$

to obtain

$$\log \mathbb{E}[\exp(C^{-1}3^{(2+d)(n-m)} J(\square_n, X, \bar{A}X))] \leq C3^{(2+d)(n-m)} \mathbb{E}[J(\square_m, X, \bar{A}X)].$$

Inequality (4-31) then follows by an application of Proposition 4.3.

Step 2. Set

$$\rho_n := \sup_{X \in B_1} J(\square_n, X, \bar{A}X).$$

In this step, we show that there exists an exponent  $\beta(d, \Lambda) > 0$  and, for every  $s \in (0, 2 + d)$ , a constant  $C(s, d, \Lambda) < \infty$  such that, for every  $n \in \mathbb{N}$ ,

$$\rho_n \leq C3^{-\beta(2+d-s)n} + \mathcal{O}_1(C3^{-sn}). \quad (4-32)$$

By (4-27) and Hölder's inequality, the relation (4-31) can be improved to

$$3^{-(2+d)(n-m)} \log \mathbb{E}[\exp(C^{-1}3^{(2+d)(n-m)}\rho_n)] \leq C3^{-\beta m}. \quad (4-33)$$

By Chebyshev's inequality, for every  $t \geq 0$ ,

$$\begin{aligned} \mathbb{P}[\rho_n \geq t] &\leq \exp(-C^{-1}3^{(2+d)(n-m)}t) \mathbb{E}[\exp(C^{-1}3^{(2+d)(n-m)}\rho_n)] \\ &\leq \exp(-C^{-1}3^{(2+d)(n-m)}t + C3^{(2+d)(n-m)-\beta m}). \end{aligned}$$

Replacing  $t$  by  $C3^{-\beta m} + t$  gives

$$\mathbb{P}[\rho_n \geq C3^{-\beta m} + t] \leq \exp(-C^{-1}3^{(2+d)(n-m)}t).$$

Choosing

$$m := \left\lfloor \frac{2+d-s}{2+d}n \right\rfloor$$

yields

$$\mathbb{P}[\rho_n \geq C3^{-\beta \frac{2+d-s}{2+d}n} + t] \leq \exp(-C^{-1}3^{sn}t).$$

By (1-19), this is (4-32), up to a redefinition of  $\beta(d, \Lambda) > 0$ .

Step 3. We now combine Lemma 4.2, (4-33) and the elementary inequality

$$\text{for all } a, b > 0, \quad (a+b)^{\frac{1}{2}} \leq a^{\frac{1}{2}} + \frac{1}{2}a^{-\frac{1}{2}}b \quad (4-34)$$

to get

$$\begin{aligned} \sup_{X, X^* \in B_1} \left| J(\square_n, X, X^*) - \left( \frac{1}{2}X \cdot \bar{A}X + \frac{1}{2}X^* \cdot \bar{A}^{-1}X^* - X \cdot X^* \right) \right| \\ \leq 3^{-\frac{\beta}{2}(2+d-s)n} + \mathcal{O}_1(C3^{-(s-\frac{\beta}{2}(2+d-s))n}). \end{aligned} \quad (4-35)$$

For every  $s' \in (0, 2 + d)$ , if we set

$$s := \frac{2s' + \beta(2+d)}{2+\beta} \in (0, 2+d),$$

then the right side of (4-35) can be rewritten as

$$3^{-\frac{\beta}{2+\beta}(2+d-s')n} + \mathcal{O}_1(C3^{-s'n}).$$

We have thus obtained (4-3), up to a redefinition of  $\beta(d, \Lambda) > 0$ . □



**Proposition 4.8.** *There exist  $C(d, \Lambda) < \infty$  and a matrix  $\bar{\mathbf{a}} \in \mathbb{R}^{d \times d}$  satisfying*

$$\text{for all } \xi \in \mathbb{R}^d, \quad \xi \cdot \bar{\mathbf{a}} \xi \geq \frac{1}{C} |\xi|^2 \quad \text{and} \quad |\bar{\mathbf{a}} \xi| \leq C |\xi| \quad (4-36)$$

such that, for every  $p, q \in \mathbb{R}^d$ , we have the equivalence

$$\frac{1}{2} \begin{pmatrix} p \\ q \end{pmatrix} \cdot \bar{\mathbf{A}} \begin{pmatrix} p \\ q \end{pmatrix} - p \cdot q = 0 \quad \iff \quad q = \bar{\mathbf{a}} p. \quad (4-37)$$

*Proof. Step 1.* We show that, for every  $p, q \in \mathbb{R}^d$ ,

$$\bar{A}(p, q) := \frac{1}{2} \begin{pmatrix} p \\ q \end{pmatrix} \cdot \bar{\mathbf{A}} \begin{pmatrix} p \\ q \end{pmatrix} \geq p \cdot q. \quad (4-38)$$

By Lemma 2.2, we have for every  $S = (\nabla u, \mathbf{g}) \in \mathcal{C}_0(I \times U)$  and  $p, q \in \mathbb{R}^d$  that

$$\int A(p + \nabla u, q + \mathbf{g}, \cdot) \geq \int (p + \nabla u) \cdot (q + \mathbf{g}) = p \cdot q.$$

By the definition of  $\mu$  in (2-5), we deduce that

$$\mu(I \times U, p, q) \geq p \cdot q,$$

and thus (4-38) follows from (4-1).

*Step 2.* We show that, for every  $q^*, p^* \in \mathbb{R}^d$ ,

$$\frac{1}{2} \begin{pmatrix} q^* \\ p^* \end{pmatrix} \cdot \bar{\mathbf{A}}^{-1} \begin{pmatrix} q^* \\ p^* \end{pmatrix} \geq q^* \cdot p^*. \quad (4-39)$$

Fix  $p^* \in \mathbb{R}^d$ . For every  $u \in \ell_{p^*} + H_{\text{par}, \square}^1(I \times U)$  and  $\mathbf{g} \in L^2(I \times U; \mathbb{R}^d)$  satisfying  $-\nabla \cdot \mathbf{g} = -\partial_t u$ , we have  $(\nabla u, \mathbf{g}) \in \mathcal{C}(I \times U)$ , as well as

$$\int_{I \times U} \nabla u = p^*$$

and

$$\int_{I \times U} (p^* - \nabla u) \cdot \mathbf{g} = \frac{1}{|I|} \int_U (u - \ell_p)^2 \geq 0.$$

Therefore, for every  $q^* \in \mathbb{R}^d$ ,

$$\begin{aligned} \mu^*(I \times U, q^*, p^*) &\geq \int_{I \times U} (-A(\nabla u, \mathbf{g}, \cdot) + q^* \cdot \nabla u + p^* \cdot \mathbf{g}) \\ &\geq \int_{I \times U} (-A(\nabla u, \mathbf{g}, \cdot) + q^* \cdot p^* + \nabla u \cdot \mathbf{g}) \\ &= q^* \cdot p^* - \int_{I \times U} (A(\nabla u, \mathbf{g}, \cdot) - \nabla u \cdot \mathbf{g}). \end{aligned}$$

By the solvability of the Cauchy–Dirichlet problem (Proposition A.1), for every  $p^* \in \mathbb{R}^d$ ,

$$0 = \inf \left\{ \int_{I \times U} (A(\nabla u, \mathbf{g}, \cdot) - \nabla u \cdot \mathbf{g}) : u \in \ell_{p^*} + H_{\text{par}, \square}^1(I \times U), \mathbf{g} \in L^2(I \times U; \mathbb{R}^d), -\nabla \cdot \mathbf{g} = -\partial_t u \right\}.$$

Combining the above yields

$$\mu^*(I \times U, q^*, p^*) \geq q^* \cdot p^*.$$

According to Theorem 4.1, we have the  $\mathbb{P}$ -a.s. limit

$$\lim_{n \rightarrow \infty} \mu^*(\square_n, X^*) = \frac{1}{2} X^* \cdot \bar{A}^{-1} X^*.$$

We therefore obtain (4-39).

*Step 3.* We argue that, for every  $p \in \mathbb{R}^d$ ,

$$\inf_{q \in \mathbb{R}^d} (\bar{A}(p, q) - p \cdot q) = 0. \tag{4-40}$$

We have already shown in (4-38) that the infimum on the left is nonnegative. The infimum is attained, by the quadratic growth of  $q \mapsto \bar{A}(p, q)$ . To see that it is equal to zero, we fix  $p \in \mathbb{R}^d$  and select  $q \in \mathbb{R}^d$  achieving the infimum. Then

$$\bar{A} \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} * \\ p \end{pmatrix}.$$

Let  $q^* \in \mathbb{R}^d$  denote the “\*” in the previous line, so that

$$\bar{A} \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} q^* \\ p \end{pmatrix}. \tag{4-41}$$

Then using (4-41), we find that

$$\begin{aligned} \frac{1}{2} \begin{pmatrix} q^* \\ p \end{pmatrix} \bar{A}^{-1} \begin{pmatrix} q^* \\ p \end{pmatrix} &= \sup_{p', q' \in \mathbb{R}^d} \left( \begin{pmatrix} q^* \\ p \end{pmatrix} \cdot \begin{pmatrix} p' \\ q' \end{pmatrix} - \frac{1}{2} \begin{pmatrix} p' \\ q' \end{pmatrix} \cdot \bar{A} \begin{pmatrix} p' \\ q' \end{pmatrix} \right) \\ &= \begin{pmatrix} q^* \\ p \end{pmatrix} \cdot \begin{pmatrix} p \\ q \end{pmatrix} - \frac{1}{2} \begin{pmatrix} p \\ q \end{pmatrix} \cdot \bar{A} \begin{pmatrix} p \\ q \end{pmatrix}. \end{aligned}$$

By the previous inequality and (4-39), we discover that

$$p \cdot q^* \leq \begin{pmatrix} q^* \\ p \end{pmatrix} \cdot \begin{pmatrix} p \\ q \end{pmatrix} - \frac{1}{2} \begin{pmatrix} p \\ q \end{pmatrix} \cdot \bar{A} \begin{pmatrix} p \\ q \end{pmatrix} = p \cdot q^* + p \cdot q - \frac{1}{2} \begin{pmatrix} p \\ q \end{pmatrix} \cdot \bar{A} \begin{pmatrix} p \\ q \end{pmatrix}.$$

Rearranging, this yields  $\bar{A}(p, q) \leq p \cdot q$ , which in view of (4-38) allows us to deduce that  $\bar{A}(p, q) = p \cdot q$  and completes the proof of (4-40).

*Step 4.* We define  $\bar{a}$  to be the matrix associated to the linear mapping taking  $p$  to the  $q$  achieving the infimum in (4-40). That the infimum is achieved at a unique minimum point is a consequence of the uniform convexity of  $q \mapsto \bar{A}(p, q)$ . That this mapping is linear is due to the fact that  $q \mapsto \bar{A}(p, q)$  is quadratic. The bounds (4-36) are a consequence of (4-2). This completes the proof of the proposition.  $\square$

### 5. Quantitative homogenization of the Cauchy–Dirichlet problem

In this section, we demonstrate the passage from the convergence of  $J$  to the homogenization of the parabolic operator. In particular, we complete the proof of Theorem 1.1 on the quantitative homogenization

of the Cauchy–Dirichlet problem. The argument is completely deterministic in the sense that the only probabilistic ingredient is the appeal to Theorem 4.1. The argument proceeds in four steps: (i) we show that convergence of  $J$  implies convergence of  $S$  and  $AS$  in  $H^{-1}$ ; (ii) we use Remark 2.7 to show that there are “finite-volume correctors” which can be found hiding in  $S$  and  $AS$  and we obtain estimates on them; (iii) we use the finite-volume correctors and a quantitative version of the standard two-scale expansion argument to pass from estimates on the correctors to estimates on the homogenization error for a general Cauchy–Dirichlet problem.

**5A. Convergence of  $J$  maximizers.** We use the multiscale Poincaré inequality (Proposition 3.6) to obtain information about the weak convergence of  $S(\square_n, X, X^*)$  as  $n \rightarrow \infty$ . It is useful to define the quantity

$$\mathcal{E}(V) := \sup_{X, X^* \in B_1} |J(V, X, X^*) - \bar{J}(X, X^*)|,$$

which keeps track of the convergence of  $J$ . We also define, given  $X, X^* \in \mathbb{R}^{2d}$ ,

$$\bar{S}(X, X^*) := \bar{A}^{-1} X^* - X.$$

Note that  $\bar{S} = \nabla_{X^*} \bar{J}$  and therefore, by (2-29) and the fact that  $J$  and  $\bar{J}$  are quadratic, we have, for every  $X, X^* \in \mathbb{R}^{2d}$ ,

$$\begin{aligned} \left| \bar{S}(X, X^*) - \int_V S(\cdot, V, X, X^*) \right| &= |\nabla_{X^*} \bar{J}(X, X^*) - \nabla_{X^*} J(V, X, X^*)| \\ &\leq C(|X| + |X^*|)\mathcal{E}(V). \end{aligned} \tag{5-1}$$

Similarly,

$$\begin{aligned} \left| \bar{A}\bar{S}(X, X^*) - \int_V AS(\cdot, V, X, X^*) \right| &= |\nabla_X \bar{J}(X, X^*) - \nabla_X J(V, X, X^*)| \\ &\leq C(|X| + |X^*|)\mathcal{E}(V). \end{aligned} \tag{5-2}$$

That is, we can control the spatial averages of  $S(\cdot, V, X, X^*)$  and  $AS(\cdot, V, X, X^*)$  in terms of the random variable  $\mathcal{E}(V)$ . The combination of this observation and Proposition 3.6 yields the following result.

**Proposition 5.1** (weak convergence of  $(S, AS)$ ). *There exists  $C(d, \Lambda) < \infty$  such that, for every  $X, X^* \in B_1$  and  $n \in \mathbb{N}$ ,*

$$3^{-n} \|S(\cdot, \square_n, X, X^*) - \bar{S}(X, X^*)\|_{\widehat{H}_{\text{par}}^{-1}(\square_n)} \leq C3^{-n} + C \sum_{m=0}^{n-1} 3^{m-n} \left( |\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} (\mathcal{E}(z + \square_m)) \right)^{\frac{1}{2}} \tag{5-3}$$

and

$$3^{-n} \|AS(\cdot, \square_n, X, X^*) - \bar{A}\bar{S}(X, X^*)\|_{\widehat{H}_{\text{par}}^{-1}(\square_n)} \leq C3^{-n} + C \sum_{m=0}^{n-1} 3^{m-n} \left( |\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} (\mathcal{E}(z + \square_m)) \right)^{\frac{1}{2}}. \tag{5-4}$$

*Proof.* We fix  $X, X^* \in B_1$  and, since it plays no role in the argument, we drop explicit display of the dependence on  $(X, X^*)$ . According to Proposition 3.6,

$$\|S(\cdot, \square_n) - \bar{S}\|_{\widehat{H}_{\text{par}}^{-1}(\square_n)} \leq C \|S(\cdot, \square_n) - \bar{S}\|_{L^2(\square_n)} + C \sum_{m=0}^{n-1} 3^m \left( |\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} |(S(\cdot, \square_n))_{z+\square_m} - \bar{S}|^2 \right)^{\frac{1}{2}}.$$

To estimate the first term on the right side, we just observe that

$$\|S(\cdot, \square_n) - \bar{S}\|_{L^2(\square_n)} \leq \|S(\cdot, \square_n)\|_{L^2(\square_n)} + |\bar{S}| \leq C.$$

We next estimate the second term. By the triangle inequality, (5-1) and Lemma 4.5,

$$\begin{aligned} \sum_{z \in \mathcal{Z}_m} |(S(\cdot, \square_n))_{z+\square_m} - \bar{S}|^2 &\leq 2 \sum_{z \in \mathcal{Z}_m} (|(S(\cdot, z+\square_m))_{z+\square_m} - \bar{S}|^2 + \|S(\cdot, \square_n) - S(\cdot, z+\square_m)\|_{L^2(z+\square_m)}^2) \\ &\leq C \sum_{z \in \mathcal{Z}_m} \mathcal{E}(z+\square_m) + C \sum_{z \in \mathcal{Z}_m} (J(z+\square_m) - J(\square_n)) \\ &\leq C \sum_{z \in \mathcal{Z}_m} (\mathcal{E}(z+\square_m) + \mathcal{E}(\square_n)). \end{aligned}$$

Thus

$$\begin{aligned} \sum_{m=0}^{n-1} 3^m \left( |\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} |(S(\cdot, \square_n))_{z+\square_m} - \bar{S}|^2 \right)^{\frac{1}{2}} &\leq C \sum_{m=0}^{n-1} 3^m \left( C\mathcal{E}(\square_n) + |\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} (\mathcal{E}(z+\square_m)) \right)^{\frac{1}{2}} \\ &\leq C \sum_{m=0}^{n-1} 3^m \left( |\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} (\mathcal{E}(z+\square_m)) \right)^{\frac{1}{2}}. \end{aligned}$$

Combining the above yields (5-3). The estimate (5-4) is obtained similarly; we just need to use (5-2) instead of (5-1). □

We next give an estimate of the random variable appearing on the right side of (5-3) and (5-4), which is a straightforward consequence of Theorem 4.1. This is the only place in this section where Theorem 4.1 or any other stochastic ingredient is used.

**Proposition 5.2.** *There exists  $\beta(d, \Lambda)$  and, for every  $s \in (0, 2 + d)$ , a constant  $C(s, d, \Lambda) < \infty$  such that, for every  $n \in \mathbb{N}$ ,*

$$\sum_{m=0}^{n-1} 3^{m-n} \left( |\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} (\mathcal{E}(z+\square_m)) \right)^{\frac{1}{2}} \leq C 3^{-n\beta(2+d-s)} + \mathcal{O}_1(C 3^{-ns}). \tag{5-5}$$

*Proof.* Fix  $s' := \frac{1}{3}(2s + 2 + d)$  and  $s'' := \frac{1}{3}(s + 2(2 + d))$  so that  $s < s' < s'' < 2 + d$  with equally sized gaps between these numbers. By Theorem 4.1 and (1-20), we have

$$|\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} (\mathcal{E}(z+\square_m)) \leq C 3^{-m\beta(2+d-s'')} + \mathcal{O}_1(C 3^{-ms''}).$$

Using the elementary inequality (4-34), we deduce that

$$\left( |\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} (\mathcal{E}(z+\square_m)) \right)^{\frac{1}{2}} \leq C 3^{-m\beta \frac{2+d-s''}{2}} + \mathcal{O}_1(C 3^{m\beta \frac{2+d-s''}{2} - ms''}).$$

Redefining  $\beta$  to be smaller if necessary, we get

$$\left( |\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} (\mathcal{E}(z+\square_m)) \right)^{\frac{1}{2}} \leq C 3^{-m\beta(2+d-s')} + \mathcal{O}_1(C 3^{-ms'}).$$

As the left side of the previous line is bounded by  $C$ , we can apply [Armstrong et al. 2017b, Lemma A.3] to obtain

$$\left( |\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} (\mathcal{E}(z + \square_m)) \right)^{\frac{1}{2}} \leq C 3^{-\frac{m\beta}{2+d-s'}} + \mathcal{O}_{d+2}(C 3^{-\frac{ms'}{d+2}}).$$

Since  $s'/(d+2) \leq 1-c$ , we may apply (1-20) again to obtain

$$\sum_{m=0}^{n-1} 3^{m-n} \left( |\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} (\mathcal{E}(z + \square_m)) \right)^{\frac{1}{2}} \leq C 3^{-\frac{n\beta}{2+d-s'}} + \mathcal{O}_{d+2}(C 3^{-\frac{ns'}{d+2}}).$$

We conclude by observing that, for any nonnegative random variable  $X$ ,

$$X \leq C 3^{-n\beta(2+d-s')} + \mathcal{O}_{d+2}(C 3^{-\frac{ns'}{d+2}}) \implies X \leq C 3^{-n\beta(2+d-s)} + \mathcal{O}_1(C 3^{-ns}), \quad (5-6)$$

where  $\beta(d, \Lambda)$  in the second statement may be smaller than in the first. To see this, we compute

$$\begin{aligned} X &\leq X + 3^{-n\beta(2+d-s)(1+d)} \leq 3^{n\beta(2+d-s)(1+d)} (X + 3^{-n\beta(2+d-s)})^{2+d} \\ &\leq C 3^{n\beta(2+d-s)(1+d)} (X^{2+d} + 3^{-n\beta(2+d-s')(2+d)}) \\ &\leq C 3^{n\beta(2+d-s)(1+d)} (C 3^{-n\beta(2+d-s)(2+d)} + \mathcal{O}_1(C 3^{-ns'})) \\ &\leq C 3^{-n\beta(2+d-s)} + \mathcal{O}_1(C 3^{-ns}), \end{aligned}$$

provided that  $\beta$  is small enough that  $\beta(2+d-s)(2+d) \leq s' - s$ . It suffices to require  $\beta \leq 1/(3(2+d))$ . This completes the proof of (5-6) and of the proposition.  $\square$

**5B. Construction of finite-volume correctors.** We next give the construction of the (finite-volume) correctors. The usage of the term ‘‘corrector’’ in stochastic homogenization is typically reserved for a function with stationary, mean-zero gradient which is the difference of a solution of the equation in the full space and an affine function. For our purposes, it is more convenient to work with a finite-volume approximation of the corrector which will be defined on a large cylinder  $\square_n$ , because this is what comes most easily and naturally out of the estimates we have already proved above. These correctors will be obtained in a simple way from  $S(\cdot, X, X^*)$  and  $AS(\cdot, X, X^*)$  and Remark 2.7; the estimates we need for them will be easy consequences of (5-3) and (5-4). The fact that these correctors are not stationary functions defined in the whole space does not create any complication in the proof of Theorem 1.1.

The corrector with slope  $e \in \mathbb{R}^d$  on the cylinder  $\square_n$  with  $n \in \mathbb{N}$  will be denoted by  $\phi_{e,n}$ . We define it from the maximizers of  $J(\square_m, X, 0)$ , studied in the previous section. We first must make an appropriate choice of  $X$ , depending on  $e$ . This is a linear algebra exercise using Proposition 4.8. We set

$$X_e := - \begin{pmatrix} e \\ \bar{a}e \end{pmatrix}$$

and observe from (4-37) that we have

$$\bar{A}X_e = - \begin{pmatrix} \bar{a}e \\ e \end{pmatrix}. \quad (5-7)$$

To check the previous line, we note (see Proposition 4.8) that the map

$$q \mapsto \frac{1}{2} \begin{pmatrix} e \\ q \end{pmatrix} \cdot \bar{A} \begin{pmatrix} e \\ q \end{pmatrix} - e \cdot q \quad \text{attains its minimum at } q = \bar{a}e$$

and the map

$$p \mapsto \frac{1}{2} \begin{pmatrix} p \\ \bar{a}e \end{pmatrix} \cdot \bar{A} \begin{pmatrix} p \\ \bar{a}e \end{pmatrix} - p \cdot \bar{a}e \quad \text{attains its minimum at } p = e.$$

Differentiating in  $p$  and  $q$ , respectively, gives (5-7).

We next take  $u_{e,n} \in H^1_{\text{par}}(\square_{n+1})$  to be the element  $u \in \mathcal{A}(\square_{n+1})$  in the representation of  $S(\cdot, \square_{n+1}, X_e, 0)$  given in Lemma 2.6, with additive constant chosen so that  $(u_{e,n})_{\square_n} = 0$ . Equivalently, in view of Remark 2.7, we can define  $u_{e,n}$  to be the function on  $\square_{n+1}$  with mean zero on  $\square_n$  with gradient given by

$$\nabla u_{e,n} = \frac{1}{2}(\pi_1 S(\cdot, \square_{n+1}, X_e, 0) + \pi_2 \mathbf{A} S(\cdot, \square_{n+1}, X_e, 0)), \tag{5-8}$$

where  $\pi_1$  and  $\pi_2$  denote the projections  $\mathbb{R}^{2d} \rightarrow \mathbb{R}^d$  onto the first and second  $d$ -variables, respectively (that is,  $\pi_1(x, y) = x$  and  $\pi_2(x, y) = y$  for  $x, y \in \mathbb{R}^d$ ). Note that, by Remark 2.7, we also have the formula

$$\mathbf{a} \nabla u_{e,n} = \frac{1}{2}(\pi_2 S(\cdot, \square_{n+1}, X_e, 0) + \pi_1 \mathbf{A} S(\cdot, \square_{n+1}, X_e, 0)). \tag{5-9}$$

By Proposition 5.1, (5-7), (5-8) and (5-9), we have

$$\begin{aligned} 3^{-n} \|\nabla u_{e,n} - e\|_{\widehat{H}^{-1}_{\text{par}}(\square_{n+1})} + 3^{-n} \|\mathbf{a} \nabla u_{e,n} - \bar{a}e\|_{\widehat{H}^{-1}_{\text{par}}(\square_{n+1})} \\ \leq C 3^{-n} + C \sum_{m=0}^n 3^{m-n} \left( |\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} (\mathcal{E}(z + \square_m)) \right)^{\frac{1}{2}}. \end{aligned} \tag{5-10}$$

Since  $u_{e,n} \in \mathcal{A}(\square_{n+1})$ , we have that  $u_{e,n}$  is a solution of

$$\partial_t u_{e,n} - \nabla \cdot (\mathbf{a} \nabla u_{e,n}) = 0 \quad \text{in } \square_{n+1}. \tag{5-11}$$

The approximate first-order corrector  $\phi_{e,n}$  is defined by subtracting the affine function  $x \mapsto e \cdot x$  from  $u_{e,n}$ :

$$\phi_{e,n}(x) := u_{e,n}(x) - e \cdot x.$$

Summarizing, we therefore have that  $\phi_{e,n}$  is a solution of

$$\partial_t \phi_{e,n} - \nabla \cdot (\mathbf{a}(e + \nabla \phi_{e,n})) = 0 \quad \text{in } \square_{n+1}, \tag{5-12}$$

and satisfies the estimates

$$\begin{aligned} 3^{-n} (\|\nabla \phi_{e,n}\|_{\widehat{H}^{-1}_{\text{par}}(\square_{n+1})} + \|\mathbf{a}(e + \nabla \phi_{e,n}) - \bar{a}e\|_{\widehat{H}^{-1}_{\text{par}}(\square_{n+1})}) \\ \leq C 3^{-n} + C \sum_{m=0}^n 3^{m-n} \left( |\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} (\mathcal{E}(z + \square_m)) \right)^{\frac{1}{2}}. \end{aligned} \tag{5-13}$$

By the previous two displays, Proposition 3.7 and  $(\phi_{e,n})_{\square_n} = 0$ , we also have

$$3^{-n} \|\phi_{e,n}\|_{L^2(\square_n)} \leq C 3^{-n} + C \sum_{m=0}^n 3^{m-n} \left( |\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} (\mathcal{E}(z + \square_m)) \right)^{\frac{1}{2}}. \tag{5-14}$$

**5C. The proof of Theorem 1.1.** The main step in the proof of Theorem 1.1 is the following proposition. It is a deterministic estimate of the homogenization error in terms of the error in the convergence of the correctors defined in the previous subsection. Since we have already estimated the latter in (5-5), (5-13) and (5-14), this is sufficient to imply the theorem. It is convenient to define, for every  $m \in \mathbb{N}$ ,

$$\mathcal{E}'(m) := 3^{-m} \sum_{k=1}^d \left( \|\phi_{e_k, m}\|_{L^2(\square_m)} + \|\nabla \phi_{e_k, m}\|_{\widehat{H}_{\text{par}}^{-1}(\square_m)} + \|\mathbf{a}(e_k + \nabla \phi_{e_k, m}) - \bar{\mathbf{a}}e_k\|_{\widehat{H}_{\text{par}}^{-1}(\square_m)} \right).$$

We also set, for each  $\varepsilon > 0$ ,

$$\mathbf{a}^\varepsilon(t, x) := \mathbf{a}\left(\frac{t}{\varepsilon^2}, \frac{x}{\varepsilon}\right) \quad \text{and} \quad \phi_{e, n}^\varepsilon(t, x) := \varepsilon \phi_{e, n}\left(\frac{t}{\varepsilon^2}, \frac{x}{\varepsilon}\right). \quad (5-15)$$

**Proposition 5.3.** Fix a bounded interval  $I := (I_-, 0) \subseteq (-\frac{1}{4}, 0)$ , a bounded Lipschitz domain  $U \subseteq \square_0$ , a small parameter  $\varepsilon \in (0, \frac{1}{2}]$ , an exponent  $\delta > 0$  and an initial-boundary condition  $f \in W_{\text{par}}^{1, 2+\delta}(I \times U)$ . Let

$$u^\varepsilon, u \in f + H_{\text{par}, \square}^1(I \times U)$$

respectively denote the solutions of

$$\begin{cases} \partial_t u^\varepsilon - \nabla \cdot (\mathbf{a}^\varepsilon \nabla u^\varepsilon) = 0 & \text{in } I \times U, \\ u^\varepsilon = f & \text{on } \partial_\square(I \times U), \end{cases} \quad (5-16)$$

and

$$\begin{cases} \partial_t u - \nabla \cdot (\bar{\mathbf{a}} \nabla u) = 0 & \text{in } I \times U, \\ u = f & \text{on } \partial_\square(I \times U). \end{cases} \quad (5-17)$$

Let  $n \in \mathbb{N}$  be such that  $3^{-n} \leq \varepsilon < 3^{-(n+1)}$ . Then there exist  $\beta(\delta, d, \Lambda) > 0$  and  $C(I, U, \delta, d, \Lambda) < \infty$  such that, for every  $r \in (0, 1)$ , we have the estimate

$$\begin{aligned} \|\nabla u^\varepsilon - \nabla u\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)} + \|\mathbf{a}^\varepsilon \nabla u^\varepsilon - \bar{\mathbf{a}} \nabla u\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)} + \|u^\varepsilon - u\|_{L^2(I \times U)} \\ \leq C \|f\|_{W_{\text{par}}^{1, 2+\delta}(I \times U)} \left( r^\beta + \frac{1}{r^{3+(2+d)/2}} \mathcal{E}'(n) \right). \end{aligned} \quad (5-18)$$

*Proof.* With  $n$  fixed as in the statement of the proposition, we let  $\phi_e = \phi_{e, n}$  denote, for each  $e \in \mathbb{R}^d$ , the (finite-volume) corrector defined in the previous subsection (we will not display its dependence on  $n$ ). We also use the notation  $\phi_e^\varepsilon = \phi_{e, n}^\varepsilon$  as in (5-15).

We will argue that  $u^\varepsilon$  is close to its modified two-scale expansion suitably cut off near the boundary. The latter is defined by

$$\begin{aligned} w^\varepsilon(t, x) &:= u(t, x) + \varepsilon \zeta_r(t, x) \sum_{k=1}^d \partial_{x_k} u(t, x) \phi_{e_k} \left( \frac{t}{\varepsilon^2}, \frac{x}{\varepsilon} \right) \\ &= u(t, x) + \zeta_r(t, x) \sum_{k=1}^d \partial_{x_k} u(t, x) \phi_{e_k}^\varepsilon(t, x), \end{aligned} \quad (5-19)$$

where  $r \in (0, 1)$  is the free parameter (representing a mesoscopic scale) given in the proposition, and we define

$$U_r := \{x \in U : \text{dist}(x, \partial U) > r\} \quad \text{and} \quad I_r := (I_- + r^2, I_+),$$

where the cutoff function  $\zeta_r$  is selected so that

$$\begin{aligned} 0 \leq \zeta_r \leq 1, \quad \zeta_r = 1 \quad \text{in } I_{2r} \times U_{2r}, \\ \zeta_r \equiv 0 \quad \text{in } (I \times U) \setminus (I_r \times U_r), \\ \text{for all } k, l \in \mathbb{N}, \quad |\nabla^k \partial_t^l \zeta_r| \leq C_{k+2l} r^{-(k+2l)}. \end{aligned} \tag{5-20}$$

Note that the constant  $C_m$  here depends on  $(I, U, d)$  in addition to  $m \in \mathbb{N}$ .

*Step 0.* We record some standard estimates from the deterministic regularity theory for uniformly parabolic equations that are needed below. The global Meyers estimate (see Proposition B.2) gives us  $\delta_0(U, d, \Lambda) > 0$  such that  $\delta \leq \delta_0$  implies

$$\|\nabla u^\varepsilon\|_{L^{2+\delta}(I \times U)} + \|\nabla u\|_{L^{2+\delta}(I \times U)} \leq C \|f\|_{W_{\text{par}}^{1,2+\delta}(I \times U)}. \tag{5-21}$$

We henceforth assume without loss of generality that  $\delta \leq \delta_0$  so that (5-21) holds. We also need pointwise derivative estimates for constant-coefficient parabolic equations. These can be found for instance in [Evans 2010, Section 2.3.3.c] (note that estimates for the operator  $\partial_t - \nabla \cdot \bar{\mathbf{a}} \nabla$  are implied by estimates for the heat equation by a simple affine change of variables), and they yield, for every  $m, l \in \mathbb{N}$ ,

$$\begin{aligned} \|\partial_t^l \nabla^m u\|_{L^\infty(I_r \times U_r)} &\leq C_{m+2l} r^{-m-2l} r^{-\frac{2+d}{2}} \|u\|_{L^2(I \times U)} \\ &\leq C_{m+2l} r^{-m-2l} r^{1-\frac{2+d}{2}} \|\nabla u\|_{L^2(I \times U)}. \end{aligned} \tag{5-22}$$

Here  $C_k$  depends only on  $(d, \Lambda)$  in addition to  $k \in \mathbb{N}$ .

The main step in the proof is to obtain an estimate on  $\|u^\varepsilon - w^\varepsilon\|_{H_{\text{par}}^1(I \times U)}$ , which is stated below in (5-25).

*Step 1.* We plug  $w^\varepsilon$  into the heterogeneous equation and estimate the error. The claim is that we can write  $(\partial_t - \nabla \cdot \mathbf{a}^\varepsilon \nabla)w^\varepsilon$  in the form

$$(\partial_t - \nabla \cdot \mathbf{a}^\varepsilon \nabla)w^\varepsilon = \partial_t F + G,$$

where  $F \in H_{\text{par}, \square}^1(I \times U)$  and  $G \in L^2(I; H^{-1}(U))$  satisfy the estimates

$$\|F\|_{L^2(I; H^1(U))} + \|G\|_{L^2(I; H^{-1}(U))} \leq C (r^{\frac{\delta}{4+2\delta}} + r^{-3-\frac{2+d}{2}} \mathcal{E}'(n)) \|f\|_{W_{\text{par}}^{1,2+\delta}(I \times U)}. \tag{5-23}$$

We begin by computing

$$\begin{aligned} \nabla w^\varepsilon &= \zeta_r \sum_{k=1}^d (e_k + \nabla \phi_{e_k}^\varepsilon) \partial_{x_k} u + \sum_{k=1}^d \phi_{e_k}^\varepsilon \nabla (\zeta_r \partial_{x_k} u) + (1 - \zeta_r) \nabla u, \\ \partial_t w^\varepsilon &= \partial_t u + \zeta_r \sum_{k=1}^d \partial_t \phi_{e_k}^\varepsilon \partial_{x_k} u + \sum_{k=1}^d \phi_{e_k}^\varepsilon \partial_t (\zeta_r \partial_{x_k} u). \end{aligned}$$

According to (5-11), the map  $\hat{u}_e^\varepsilon(t, x) := e \cdot x + \phi_e^\varepsilon(t, x)$  is a solution of the equation

$$\partial_t \hat{u}_e^\varepsilon - \nabla \cdot (\mathbf{a}^\varepsilon \nabla \hat{u}_e^\varepsilon) = 0 \quad \text{in } I \times U.$$



Therefore we find that

$$\partial_t w^\varepsilon - \nabla \cdot (\mathbf{a}^\varepsilon \nabla w^\varepsilon) = \partial_t u + \sum_{k=1}^d \phi_{e_k}^\varepsilon \partial_t (\zeta_r \partial_{x_k} u) - \sum_{k=1}^d \nabla (\zeta_r \partial_{x_k} u) \cdot \mathbf{a}^\varepsilon (e_k + \nabla \phi_{e_k}^\varepsilon) - \nabla \cdot \left( \mathbf{a}^\varepsilon \left( \sum_{k=1}^d \phi_{e_k}^\varepsilon \nabla (\zeta_r \partial_{x_k} u) + (1 - \zeta_r) \nabla u \right) \right).$$

Since  $u$  satisfies the homogenized equation, we have furthermore that

$$\partial_t u = \nabla \cdot \bar{\mathbf{a}} \nabla u = \sum_{k=1}^d \nabla (\zeta_r \partial_{x_k} u) \cdot \bar{\mathbf{a}} e_k + \nabla \cdot ((1 - \zeta_r) \bar{\mathbf{a}} \nabla u),$$

and this gives us the identity

$$\partial_t w^\varepsilon - \nabla \cdot (\mathbf{a}^\varepsilon \nabla w^\varepsilon) = \sum_{k=1}^d \phi_{e_k}^\varepsilon \partial_t (\zeta_r \partial_{x_k} u) - \sum_{k=1}^d \nabla (\zeta_r \partial_{x_k} u) \cdot (\mathbf{a}^\varepsilon (e_k + \nabla \phi_{e_k}^\varepsilon) - \bar{\mathbf{a}} e_k) - \nabla \cdot \left( \mathbf{a}^\varepsilon \sum_{k=1}^d \phi_{e_k}^\varepsilon \nabla (\zeta_r \partial_{x_k} u) \right) - \nabla \cdot ((\mathbf{a}^\varepsilon - \bar{\mathbf{a}})(1 - \zeta_r) \nabla u).$$

According to Lemma 3.11, we can find  $v \in L^2(I; H_0^1(U))$  and  $v^* \in L^2(I; H^{-1}(U))$  such that

$$f^* := - \sum_{k=1}^d \nabla (\zeta_r \partial_{x_k} u) \cdot (\mathbf{a}^\varepsilon (e_k + \nabla \phi_{e_k}^\varepsilon) - \bar{\mathbf{a}} e_k) = \partial_t v + v^* \quad (5-24)$$

and

$$\|v\|_{L^2(I; H^1(U))} + \|v^*\|_{L^2(I; H^{-1}(U))} \leq C \|f^*\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)}.$$

The lemma allows us to take  $v$  and  $v^*$  to vanish in a neighborhood of the parabolic boundary of  $I \times U$ . Since the left side of (5-24) belongs to  $L^2(I; H^{-1}(U))$ , we have also that  $v \in H_{\text{par}, \square}^1(I \times U)$ . Therefore we obtain

$$\partial_t w^\varepsilon - \nabla \cdot (\mathbf{a}^\varepsilon \nabla w^\varepsilon) = \partial_t F + G,$$

where

$$F := v$$

and

$$G := v^* + \sum_{k=1}^d \phi_{e_k}^\varepsilon \partial_t (\zeta_r \partial_{x_k} u) - \nabla \cdot \left( \mathbf{a}^\varepsilon \sum_{k=1}^d \phi_{e_k}^\varepsilon \nabla (\zeta_r \partial_{x_k} u) \right) - \nabla \cdot ((\mathbf{a}^\varepsilon - \bar{\mathbf{a}})(1 - \zeta_r) \nabla u).$$

It is clear that

$$\begin{aligned} & \|F\|_{L^2(I; H^1(U))} + \|G\|_{L^2(I; H^{-1}(U))} \\ & \leq C \sum_{k=1}^d \|\phi_{e_k}^\varepsilon \partial_t (\zeta_r \partial_{x_k} u)\|_{L^2(I \times U)} + C \sum_{k=1}^d \|\nabla (\zeta_r \partial_{x_k} u) \cdot (\mathbf{a}^\varepsilon (e_k + \nabla \phi_{e_k}^\varepsilon) - \bar{\mathbf{a}} e_k)\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)} \\ & \quad + C \left\| \mathbf{a}^\varepsilon \sum_{k=1}^d \phi_{e_k}^\varepsilon \nabla (\zeta_r \partial_{x_k} u) \right\|_{L^2(I \times U)} + C \|(\mathbf{a}^\varepsilon - \bar{\mathbf{a}})(1 - \zeta_r) \nabla u\|_{L^2(I \times U)} \\ & = T_1 + T_2 + T_3 + T_4. \end{aligned}$$

We will now show that each of the four terms  $T_i$  can be estimated by the right side of (5-23), using the definition of  $\mathcal{E}'(n)$  and the bounds (5-20), (5-21) and (5-22). For  $T_1$ , we use (5-20) and (5-22) to find that, for each  $e \in \partial B_1$ ,

$$\begin{aligned} \|\phi_{e_k}^\varepsilon \partial_t(\zeta_r \partial_{x_k} u)\|_{L^2(I \times U)} &\leq C \|\partial_t(\zeta_r \partial_{x_k} u)\|_{L^\infty(I \times U)} \|\phi_e^\varepsilon\|_{L^2(I \times U)} \\ &\leq Cr^{-3-\frac{2+d}{2}} \mathcal{E}'(n) \|f\|_{H_{\text{par}}^1(I \times U)}. \end{aligned}$$

For  $T_2$ , we have

$$\begin{aligned} \|\nabla(\zeta_r \partial_{x_k} u) \cdot (\mathbf{a}^\varepsilon(e + \nabla \phi_e^\varepsilon) - \bar{\mathbf{a}}e)\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)} &\leq C \|\nabla(\zeta_r \partial_{x_k} u)\|_{W^{1,\infty}(I \times U)} \|\mathbf{a}^\varepsilon(e + \nabla \phi_e^\varepsilon) - \bar{\mathbf{a}}e\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)} \\ &\leq Cr^{-3-\frac{2+d}{2}} \mathcal{E}'(n) \|f\|_{H_{\text{par}}^1(I \times U)}. \end{aligned}$$

For  $T_3$ , we use (5-20) and (5-22) again to get

$$\begin{aligned} \|\mathbf{a}^\varepsilon \phi_e^\varepsilon \nabla(\zeta_r \partial_{x_k} u)\|_{L^2(I \times U)} &\leq C \|\nabla(\zeta_r \partial_{x_k} u)\|_{L^\infty(I \times U)} \|\phi_e^\varepsilon\|_{L^2(I \times U)} \\ &\leq Cr^{-2-\frac{2+d}{2}} \mathcal{E}'(n) \|f\|_{H_{\text{par}}^1(I \times U)}. \end{aligned}$$

Finally, for  $T_4$ , we use (5-20), (5-21) and Hölder's inequality to get

$$\begin{aligned} \|(\mathbf{a}^\varepsilon - \bar{\mathbf{a}})(1 - \zeta_r) \nabla u\|_{L^2(I \times U)} &\leq C |\{x \in I \times U : \zeta_r(x) \neq 1\}|^{\frac{\delta}{4+2\delta}} \|\nabla u\|_{L^{2+\delta}(I \times U)} \\ &\leq Cr^{\frac{\delta}{4+2\delta}} \|f\|_{W_{\text{par}}^{1,2+\delta}(I \times U)}. \end{aligned}$$

This completes the proof of (5-23).

*Step 2.* We deduce that

$$\|u^\varepsilon - w^\varepsilon\|_{L^2(I; H^1(U))} \leq C \left( r^{\frac{\delta}{4+2\delta}} + r^{-3-\frac{2+d}{2}} \mathcal{E}'(n) \right) \|f\|_{W_{\text{par}}^{1,2+\delta}(I \times U)}. \quad (5-25)$$

This is an immediate consequence of the estimate (5-23) proved in the previous step, the fact that  $u^\varepsilon - w^\varepsilon, F \in H_{\text{par},\square}^1(I \times U)$  and the estimate (A-12) proved in Appendix A.

At this point, we have succeeded in comparing  $u^\varepsilon$  to  $w^\varepsilon$ . What is left is to compare  $w^\varepsilon$  to  $u$  by showing that the second term on the right side of (5-19) is small. This is relatively straightforward to obtain from (5-13) and (5-14).

*Step 3.* We show that

$$\begin{aligned} \|u - w^\varepsilon\|_{L^2(I \times U)} + \|\nabla u - \nabla w^\varepsilon\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)} + \|\bar{\mathbf{a}} \nabla u - \mathbf{a}^\varepsilon \nabla w^\varepsilon\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)} \\ \leq C \left( r^{\frac{\delta}{4+2\delta}} + r^{-3-\frac{2+d}{2}} \mathcal{E}'(n) \right) \|f\|_{W_{\text{par}}^{1,2+\delta}(I \times U)}. \end{aligned} \quad (5-26)$$

We use the formula

$$\nabla w^\varepsilon(t, x) - \nabla u(t, x) = \varepsilon \sum_{k=1}^d \nabla(\zeta_r \partial_{x_k} u)(t, x) \phi_{e_k} \left( \frac{t}{\varepsilon^2}, \frac{x}{\varepsilon} \right) + \zeta_r(t, x) \sum_{k=1}^d \partial_{x_k} u(t, x) \nabla \phi_{e_k} \left( \frac{t}{\varepsilon^2}, \frac{x}{\varepsilon} \right)$$

to get

$$\begin{aligned} \|\nabla u - \nabla w^\varepsilon\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)} \\ \leq \|\nabla(\zeta_r \nabla u)\|_{L^\infty(I \times U)} \varepsilon \sum_{k=1}^d \|\phi_{e_k}\|_{L^2(Q_{\varepsilon^{-1}})} + C \|\zeta_r \nabla u\|_{W^{1,\infty}(I \times U)} \left\| \nabla \phi_{e_k} \left( \frac{\cdot}{\varepsilon^2}, \frac{\cdot}{\varepsilon} \right) \right\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)} \\ \leq Cr^{-2} \|f\|_{H_{\text{par}}^1(I \times U)} \mathcal{E}'(n). \end{aligned}$$

For the fluxes, we find it convenient to use coordinates. We have

$$\begin{aligned}
 (\mathbf{a}^\varepsilon \nabla w^\varepsilon)_i(t, x) &= \sum_{j,k=1}^d \zeta_r(t, x) \mathbf{a}_{ij}^\varepsilon(t, x) \partial_{x_k} u(t, x) \left( \delta_{jk} + \partial_{x_j} \phi_{e_k} \left( \frac{t}{\varepsilon^2}, \frac{x}{\varepsilon} \right) \right) \\
 &\quad + \varepsilon \sum_{j,k=1}^d \mathbf{a}_{ij}^\varepsilon(t, x) \partial_{x_j} (\zeta_r \partial_{x_k} u)(t, x) \phi_{e_k} \left( \frac{t}{\varepsilon^2}, \frac{x}{\varepsilon} \right).
 \end{aligned}$$

Thus

$$\begin{aligned}
 &(\mathbf{a}^\varepsilon \nabla w^\varepsilon)_i(t, x) - (\bar{\mathbf{a}} \nabla u)_i(t, x) \\
 &= \sum_{j,k=1}^d \zeta_r(t, x) \partial_{x_k} u(t, x) \left( \mathbf{a}_{ij}^\varepsilon(t, x) \left( \delta_{jk} + \partial_{x_j} \phi_{e_k} \left( \frac{t}{\varepsilon^2}, \frac{x}{\varepsilon} \right) \right) - \bar{\mathbf{a}}_{ik} \right) \\
 &\quad + \sum_{j,k=1}^d (1 - \zeta_r(t, x)) \bar{\mathbf{a}}_{ik} \partial_{x_k} u(t, x) + \varepsilon \sum_{j,k=1}^d \mathbf{a}_{ij}^\varepsilon(t, x) \partial_{x_j} (\zeta_r \partial_{x_k} u)(t, x) \phi_{e_k} \left( \frac{t}{\varepsilon^2}, \frac{x}{\varepsilon} \right).
 \end{aligned}$$

We can easily estimate the last two terms on the right side using (5-20), (5-21), (5-22) and the Hölder inequality. We have

$$\begin{aligned}
 \sum_{j,k=1}^d \|(1 - \zeta_r) \bar{\mathbf{a}}_{ik} \partial_{x_k} u\|_{L^2(I \times U)} &\leq C \|\nabla u\|_{L^2((I \times U) \setminus (I_r \times U_r))} \\
 &\leq Cr^{\frac{\delta}{4+2\delta}} \|\nabla u\|_{L^{2+\delta}(I \times U)} \leq Cr^{\frac{\delta}{4+2\delta}} \|f\|_{W_{\text{par}}^{1,2+\delta}(I \times U)}
 \end{aligned}$$

and

$$\begin{aligned}
 \varepsilon \sum_{j,k=1}^d \left\| \mathbf{a}_{ij}^\varepsilon \partial_{x_j} (\zeta_r \partial_{x_k} u) \phi_{e_k} \left( \frac{\cdot}{\varepsilon^2}, \frac{\cdot}{\varepsilon} \right) \right\|_{L^2(I \times U)} &\leq C \|\nabla (\zeta_r \nabla u)\|_{L^\infty(I \times U)} \varepsilon \sum_{k=1}^d \|\phi_{e_k}\|_{L^2(Q_{1/2})} \\
 &\leq Cr^{-2-\frac{2+d}{2}} \|f\|_{H_{\text{par}}^1(I \times U)} \mathcal{E}'(n).
 \end{aligned}$$

For the first term, we have

$$\begin{aligned}
 &\sum_{j,k=1}^d \left\| \zeta_r \partial_{x_k} u \left( \mathbf{a}_{ij}^\varepsilon \left( \delta_{jk} + \partial_{x_j} \phi_{e_k} \left( \frac{\cdot}{\varepsilon^2}, \frac{\cdot}{\varepsilon} \right) \right) - \bar{\mathbf{a}}_{ik} \right) \right\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)} \\
 &\leq C \sum_{j,k=1}^d \|\zeta_r \partial_{x_k} u\|_{W^{1,\infty}(I \times U)} \left\| \mathbf{a}_{ij}^\varepsilon \left( \delta_{jk} + \partial_{x_j} \phi_{e_k} \left( \frac{\cdot}{\varepsilon^2}, \frac{\cdot}{\varepsilon} \right) \right) - \bar{\mathbf{a}}_{ik} \right\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)} \\
 &\leq Cr^{-3} \|f\|_{H_{\text{par}}^1(I \times U)} \sum_{k=1}^d \|\mathbf{a}^\varepsilon(e_k + \nabla \phi_{e_k}) - \bar{\mathbf{a}} e_k\|_{\widehat{H}_{\text{par}}^{-1}(Q_{1/2})} \\
 &\leq Cr^{-3-\frac{2+d}{2}} \|f\|_{H_{\text{par}}^1(I \times U)} \mathcal{E}'(n).
 \end{aligned}$$

Combining the previous four displays, we obtain

$$\|\mathbf{a}^\varepsilon \nabla w^\varepsilon - \bar{\mathbf{a}} \nabla u\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)} \leq C(r^{\frac{\delta}{4+2\delta}} + r^{-3-\frac{2+d}{2}} \mathcal{E}'(n)) \|f\|_{W_{\text{par}}^{1,2+\delta}(I \times U)}.$$

Finally, for the estimate of  $w^\varepsilon - u$ , we have

$$\begin{aligned} \|w^\varepsilon - u\|_{L^2(I \times U)} &\leq C \|\nabla u\|_{L^\infty((I \times U) \setminus (I_r \times U_r))} \mathcal{E} \sum_{k=1}^d \left\| \phi_{e_k} \left( \frac{\cdot}{\varepsilon^2}, \frac{\cdot}{\varepsilon} \right) \right\|_{L^2(I \times U)} \\ &\leq C r^{-1 - \frac{2+d}{2}} \|f\|_{H_{\text{par}}^1(I \times U)} \mathcal{E}'(n). \end{aligned}$$

This completes the proof of (5-26).

*Step 4.* We summarize and conclude the argument. According to (5-25), (5-26) and the triangle inequality, we have

$$\begin{aligned} \|\nabla u^\varepsilon - \nabla u\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)} &\leq \|\nabla u^\varepsilon - \nabla w^\varepsilon\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)} + \|\nabla w^\varepsilon - \nabla u\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)} \\ &\leq C \|u^\varepsilon - w^\varepsilon\|_{L^2(I; H^1(U))} + \|\nabla w^\varepsilon - \nabla u\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)} \\ &\leq C (r^{\frac{\delta}{4+2\delta}} + r^{-3 - \frac{2+d}{2}} \mathcal{E}'(n)) \|f\|_{W_{\text{par}}^{1,2+\delta}(I \times U)}. \end{aligned}$$

Similarly, for the fluxes we have

$$\begin{aligned} \|\mathbf{a}^\varepsilon \nabla u^\varepsilon - \bar{\mathbf{a}} \nabla u\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)} &\leq \|\mathbf{a}^\varepsilon \nabla u^\varepsilon - \mathbf{a}^\varepsilon \nabla w^\varepsilon\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)} + \|\mathbf{a}^\varepsilon \nabla w^\varepsilon - \bar{\mathbf{a}} \nabla u\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)} \\ &\leq C \|u^\varepsilon - w^\varepsilon\|_{L^2(I; H^1(U))} + \|\mathbf{a}^\varepsilon \nabla w^\varepsilon - \bar{\mathbf{a}} \nabla u\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)} \\ &\leq C (r^{\frac{\delta}{4+2\delta}} + r^{-3 - \frac{2+d}{2}} \mathcal{E}'(n)) \|f\|_{W_{\text{par}}^{1,2+\delta}(I \times U)}, \end{aligned}$$

and, for the homogenization error, we have

$$\begin{aligned} \|u^\varepsilon - u\|_{L^2(I \times U)} &\leq \|u^\varepsilon - w^\varepsilon\|_{L^2(I \times U)} + \|w^\varepsilon - u\|_{L^2(I \times U)} \\ &\leq C (r^{\frac{\delta}{4+2\delta}} + r^{-3 - \frac{2+d}{2}} \mathcal{E}'(n)) \|f\|_{W_{\text{par}}^{1,2+\delta}(I \times U)}. \end{aligned} \quad \square$$

To complete the proof of Theorem 1.1, we just need to estimate the random variables on the right side of (5-18) using Proposition 5.2 and the estimates (5-13) and (5-14) for the correctors.

*Proof of Theorem 1.1.* Fix  $s \in (0, 2 + d)$  and put  $s' := \frac{1}{2}s + \frac{1}{2}(2 + d)$  and  $s'' := \frac{1}{2}s' + \frac{1}{2}(2 + d)$ . Thus  $s < s' < s'' < 2 + d$  and the gaps are at least of size  $\frac{1}{4}(2 + d - s)$ . Observe that (5-13) and (5-14) imply

$$\mathcal{E}'(n) \leq C 3^{-n} + C \sum_{m=0}^{n-1} 3^{m-n} \left( |\mathcal{Z}_m|^{-1} \sum_{z \in \mathcal{Z}_m} (\mathcal{E}(z + \square_m)) \right)^{\frac{1}{2}}.$$

Thus, by Proposition 5.2,

$$\mathcal{E}'(n) \leq C 3^{-n\beta(2+d-s'')} + \mathcal{O}_1(C 3^{-ns''}).$$

Hence

$$3^{ns'} (\mathcal{E}'(n) - C 3^{-n\beta(2+d-s'')})_+ \leq \mathcal{O}_1(C 3^{-n(s''-s')}).$$

By (1-20),

$$\mathcal{X} := \sum_{n \in \mathbb{N}} 3^{ns'} (\mathcal{E}'(n) - C 3^{-n\beta(2+d-s'')})_+ \leq \mathcal{O}_1(C).$$

Proposition 5.3 yields therefore that, for every  $\varepsilon \in (0, \frac{1}{2}]$  and  $r \in (0, 1)$ ,

$$\begin{aligned} & \|\nabla u^\varepsilon - \nabla u\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)} + \|\mathbf{a}^\varepsilon \nabla u^\varepsilon - \bar{\mathbf{a}} \nabla u\|_{\widehat{H}_{\text{par}}^{-1}(I \times U)} + \|u^\varepsilon - u\|_{\underline{L}^2(I \times U)} \\ & \leq C \|f\|_{W_{\text{par}}^{1,p}(I \times U)} \left( r^\beta + \frac{1}{r^{3+(2+d)/2}} (\varepsilon^{\beta(2+d-s'')} + \varepsilon^{s'} \mathcal{X}') \right). \end{aligned}$$

We now select  $r \in (0, 1)$  as small as possible (it must be no larger than a positive power of  $\varepsilon$ ) such that  $r^{-3-(2+d)/2} \varepsilon^{s'} \leq \varepsilon^s$  and  $r^{-3-(2+d)/2} \leq \varepsilon^{-\beta(2+d-s'')/2}$ . We can take for example

$$r := \varepsilon^{\beta(2+d-s'')/(3+(2+d)/2)} \vee \varepsilon^{(s'-s)/(3+(2+d)/2)}.$$

Recalling that  $2+d-s'' \geq \frac{1}{4}(2+d-s)$  and  $s'-s \geq \frac{1}{4}(2+d-s)$ , we obtain the theorem.  $\square$

## 6. Regularity theory

In this section, we sketch the proof of Theorem 1.2, following along the lines of the argument given in the proof of [Armstrong et al. 2017b, Theorem 3.6] in the elliptic case. We do not give full details, since this would involve an almost verbatim repetition of the proof of the latter.

We begin by reformulating Theorem 1.1 in a slightly different way in terms of *caloric approximation*, which is more convenient for its application in this section. The next statement can be compared to its elliptic analogue in [Armstrong et al. 2017b, Proposition 3.2].

**Proposition 6.1** (caloric approximation). *Fix  $s \in (0, 2+d)$ . There exist an exponent  $\alpha(d, \Lambda) > 0$ , a constant  $C(s, d, \Lambda) < \infty$ , and a random variable  $\mathcal{X}_s : \Omega \rightarrow [1, \infty]$  satisfying the estimate*

$$\mathcal{X}_s = \mathcal{O}_s(C) \tag{6-1}$$

such that the following holds: for every  $R \geq \mathcal{X}_s$  and weak solution  $u \in H_{\text{par}}^1(Q_R)$  of

$$\partial_t u - \nabla \cdot (\mathbf{a} \nabla u) = 0 \quad \text{in } Q_R, \tag{6-2}$$

there exists a solution  $\bar{u} \in H_{\text{par}}^1(Q_{R/2})$  of the equation

$$\partial_t \bar{u} - \nabla \cdot (\bar{\mathbf{a}} \nabla \bar{u}) = 0 \quad \text{in } Q_{R/2}$$

such that

$$\|u - \bar{u}\|_{\underline{L}^2(Q_{R/2})} \leq C R^{-\alpha(2+d-s)} \|u - (u)_{Q_R}\|_{\underline{L}^2(Q_R)}. \tag{6-3}$$

*Proof.* This is a simple application of Theorem 1.1 combined with the parabolic Meyers estimate. The argument is almost the same as in the elliptic case presented in [Armstrong et al. 2017b, Proposition 3.2]; we just need to replace the elliptic interior Meyers estimate with its parabolic analogue proved in Proposition B.1 below. The latter gives us  $\delta(d, \Lambda) > 0$  and  $C(d, \Lambda) > 0$  such that, for every  $u \in H_{\text{par}}^1(Q_R)$  satisfying (6-2), we have  $\nabla u \in L^{2+\delta}(Q_{R/2})$  and the estimate

$$\|\nabla u\|_{\underline{L}^{2+\delta}(Q_{R/2})} \leq \frac{C}{R} \|u - (u)_{Q_R}\|_{\underline{L}^2(Q_R)}. \tag{6-4}$$

Following the proof of [Armstrong et al. 2017b, Proposition 3.2], using (6-4), substituting Theorem 1.1 in place of Theorem 2.16 there and making obvious changes to the notation, we obtain the proposition.  $\square$

We next state a parabolic counterpart of [Armstrong et al. 2017b, Lemma 3.5].

**Lemma 6.2.** *Fix  $\alpha \in [0, 1]$ ,  $K \geq 1$  and  $X \geq 1$ . Let  $R \geq 2X$  and  $u \in L^2(Q_R)$  have the property that, for every  $r \in [X, R]$ , there exists  $w_r \in H_{\text{par}}^1(Q_{r/2})$  which is a solution of*

$$\partial_t w_r - \nabla \cdot (\bar{\mathbf{a}} \nabla w_r) = 0 \quad \text{in } Q_{r/2}$$

and satisfies

$$\|u - w_r\|_{\underline{L}^2(Q_{r/2})} \leq K r^{-\alpha} \|u - (u)_{Q_r}\|_{\underline{L}^2(Q_r)}. \quad (6-5)$$

Then, for every  $k \in \mathbb{N}$ , there exists  $\theta(\alpha, k, d, \Lambda) \in (0, \frac{1}{2})$  and  $C(\alpha, k, d, \Lambda) < \infty$  such that, for every  $r \in [X, R]$ ,

$$\inf_{p \in \bar{\mathcal{A}}_k(Q_\infty)} \|u - p\|_{\underline{L}^2(Q_{\theta r})} \leq \frac{1}{4} \theta^{k+1-\frac{\alpha}{2}} \inf_{p \in \bar{\mathcal{A}}_k(Q_\infty)} \|u - p\|_{\underline{L}^2(Q_r)} + C K r^{-\alpha} \|u - (u)_{Q_r}\|_{\underline{L}^2(Q_r)}. \quad (6-6)$$

*Proof.* The proof is essentially the same as that of [Armstrong et al. 2017b, Lemma 3.5]. We just have to substitute balls for parabolic cylinders and use Proposition 6.1 in place of its elliptic version. These changes cause no additional complexity in the proof.  $\square$

With Lemma 6.2 in hand, the proof of Theorem 1.2 is now completed in the same way as the one of [Armstrong et al. 2017b, Theorem 3.6], by following the argument almost verbatim and making only obvious modifications. We refer to that book for the details.

## Appendix A. Variational structure of uniformly parabolic equations

The aim of this appendix is to show that the solution of the parabolic equation (2-1) can be obtained as the minimizer of a uniformly convex functional. We will prove this result in the more general context of uniformly monotone operators, since this causes no modification to the proof. Although our statement differs in detail, it is close to the main result of [Ghoussoub and Tzou 2004]; see also [Ghoussoub 2009]. The proof we give is also relatively close to that of [Ghoussoub and Tzou 2004]; we hope that the reader will appreciate the short and self-contained presentation in this appendix. The fact that a parabolic equation can be cast as the first variation of a uniformly convex integral functional was first discovered in [Brezis and Ekeland 1976a; 1976b].

Let  $I := (0, T) \subseteq \mathbb{R}$  and  $U \subseteq \mathbb{R}^d$  be a bounded Lipschitz domain. For a given right-hand side  $w^*$  and boundary condition (both of which will be made precise below), we study the solvability of the parabolic equation

$$\partial_t u - \nabla \cdot (\mathbf{a}(\nabla u, \cdot)) = w^* \quad \text{in } I \times U, \quad (\text{A-1})$$

where the dot represents the time-space variable in  $I \times U \subseteq \mathbb{R}^{1+d}$ , and  $\mathbf{a} \in L_{\text{loc}}^\infty(\mathbb{R}^d \times \mathbb{R}^{1+d}; \mathbb{R}^d)$  is Lipschitz and uniformly monotone in its first argument. That is, we assume that there exists a constant  $\lambda < \infty$  such that, for every  $p_1, p_2 \in \mathbb{R}^d$  and  $z \in \mathbb{R}^{1+d}$ ,

$$\begin{aligned} |\mathbf{a}(p_1, z) - \mathbf{a}(p_2, z)| &\leq \lambda |p_1 - p_2|, \\ (\mathbf{a}(p_1, z) - \mathbf{a}(p_2, z)) \cdot (p_1 - p_2) &\geq \lambda^{-1} |p_1 - p_2|^2. \end{aligned} \quad (\text{A-2})$$

As a first step, we introduce a variational representation of the mapping  $p \mapsto \mathbf{a}(p, z)$ , for each  $z \in \mathbb{R}^{1+d}$ . This idea is often attributed to Fitzpatrick [1988], although it actually appeared in [Krylov 1982] several years earlier.

By [Armstrong and Mourrat 2016, Theorem 2.9], there exists  $A \in L_{\text{loc}}^{\infty}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{1+d})$  satisfying the following properties for  $\Lambda := 2\lambda + 1$  and for each  $z \in \mathbb{R}^{1+d}$ :

- The mapping

$$(p, q) \mapsto A(p, q, z) - \frac{1}{2\Lambda}(|p|^2 + |q|^2) \quad \text{is convex.} \quad (\text{A-3})$$

- The mapping

$$(p, q) \mapsto A(p, q, z) - \frac{\Lambda}{2}(|p|^2 + |q|^2) \quad \text{is concave.} \quad (\text{A-4})$$

- For every  $p, q \in \mathbb{R}^d$ , we have

$$A(p, q, z) \geq p \cdot q, \quad (\text{A-5})$$

and

$$A(p, q, z) = p \cdot q \iff q = \mathbf{a}(p, z). \quad (\text{A-6})$$

In the particular case when  $p \mapsto \mathbf{a}(p, z)$  is linear, we can define the mapping  $(p, q) \mapsto A(p, q, z)$  according to (2-3); see Lemma 2.2. Another familiar example is when  $\mathbf{a}(\cdot, z)$  is the gradient of a uniformly convex Lagrangian  $L(p, z)$ ; that is,  $\mathbf{a}(\cdot, z) = \nabla_p L(\cdot, z)$ , where  $p \mapsto L(p, z)$  is uniformly convex. In this case, we can take

$$A(p, q, z) := L(p, z) + L^*(q, z),$$

where  $L^*$  is the Legendre–Fenchel transform of  $L$ . We remark that the choice of  $A$  is in general not unique.

We define the function space

$$Z(I \times U) := \{(u, u^*) : u \in L^2(I; H^1(U)) \text{ and } (u^* - \partial_t u) \in L^2(I; H^{-1}(U))\},$$

with norm

$$\|(u, u^*)\|_{Z(I \times U)} := \|u\|_{L^2(I; H^1(U))} + \|u^* - \partial_t u\|_{L^2(I; H^{-1}(U))}.$$

The function space  $H_{\text{par}}^1(I \times U)$  is defined in (1-9)–(1-10). We denote by  $H_{\text{par}, \sqcup}^1(I \times U)$  the closure in  $H_{\text{par}}^1(I \times U)$  of the set of smooth functions with compact support in  $(0, T] \times U$ . For every  $(u, u^*) \in Z(I \times U)$ , we set

$$\mathcal{J}[u, u^*] := \inf \left\{ \int_{I \times U} (A(\nabla u, \mathbf{g}, \cdot) - \nabla u \cdot \mathbf{g}) : -\nabla \cdot \mathbf{g} = u^* - \partial_t u \right\}. \quad (\text{A-7})$$

In the infimum above, we understand that  $\mathbf{g} \in L^2(I \times U; \mathbb{R}^d)$ , and the last condition is interpreted as

$$\text{for all } \phi \in L^2(I; H_0^1(U)), \quad \int_{I \times U} \nabla \phi \cdot \mathbf{g} = \int_{I \times U} \phi (u^* - \partial_t u). \quad (\text{A-8})$$

Note that the set of candidates for  $\mathbf{g}$  is not empty; indeed, denoting by  $\Delta_U^{-1}$  the solution operator for the Laplacian in  $U$  with a null Dirichlet boundary condition, we verify that

$$\mathbf{g} = \nabla \Delta_U^{-1}(u^* - \partial_t u)$$

is a suitable candidate, by the assumption of  $u^* - \partial_t u \in L^2(I; H^{-1}(U))$ .

The goal of this appendix is to prove the following proposition.

**Proposition A.1.** *For each  $(w, w^*) \in Z(I \times U)$ , the mapping*

$$\begin{aligned} w + H_{\text{par}, \sqcup}^1(I \times U) &\rightarrow \mathbb{R}, \\ u &\mapsto \mathcal{J}[u, w^*], \end{aligned} \tag{A-9}$$

*is uniformly convex. Moreover, its minimum is zero, and the associated minimizer is the unique  $u \in w + H_{\text{par}, \sqcup}^1(I \times U)$  solution of (A-1), in the sense that*

$$\text{for all } \phi \in L^2(I; H_0^1(U)), \quad \int_{I \times U} \nabla \phi \cdot \mathbf{a}(\nabla u, \cdot) = \int_{I \times U} \phi(w^* - \partial_t u).$$

**Remark A.2.** By the inclusion

$$H_{\text{par}}^1(I \times U) \times L^2(I; H^{-1}(U)) \subseteq Z(I \times U), \tag{A-10}$$

Proposition A.1 ensures in particular the solvability of the parabolic equation (A-1) for every right-hand side  $w^* \in L^2(I; H^{-1}(U))$  and every boundary condition  $w \in H_{\text{par}}^1(I \times U)$ ; the solution thus obtained then belongs to  $H_{\text{par}}^1(I \times U)$ .

More generally, for every  $w^*$  of the form

$$w^* = \partial_t f + v, \quad f \in L^2(I; H_0^1(U)), \quad v \in L^2(I; H^{-1}(U)), \tag{A-11}$$

we have  $(f, w^*) \in Z(I \times U)$  and hence Proposition A.1 yields the existence of a unique solution  $u \in f + H_{\text{par}, \sqcup}^1(I \times U)$  of (A-1) which satisfies the estimate

$$\|u - f\|_{H_{\text{par}}^1(I \times U)} \leq C(\|f\|_{L^2(I; H^1(U))} + \|w^* - \partial_t f\|_{L^2(I; H^{-1}(U))}). \tag{A-12}$$

In other words, we have identified a mapping

$$\partial_t f + v \mapsto f + P(f, v), \tag{A-13}$$

where  $f \in L^2(I; H_0^1(U))$ ,  $v \in L^2(I; H^{-1}(U))$ , and  $P$  is a bounded linear operator from  $L^2(I; H_0^1(U)) \times L^2(I; H^{-1}(U))$  to  $H_{\text{par}, \sqcup}^1(I \times U)$ . If we moreover restrict our attention, say, to the set of functions  $f$  which vanish in a neighborhood of  $\{0\} \times U$ , then this mapping provides us with a notion of a solution of (A-1) with null Dirichlet boundary condition on the parabolic boundary of  $I \times U$ . This additional regularity assumption on the behavior of  $f$  near the initial time can of course be weakened as desired.

Note that every  $w^*$  of the form (A-11) belongs to  $\widehat{H}_{\text{par}}^{-1}(I \times U)$ , but the latter space is strictly larger than the set of such  $w^*$ . This may at first glance appear at odds with Lemma 3.11; however that lemma required that  $u^*$  belong to  $L^2(I \times U)$ . This hypothesis rules out certain singular distributions which belong to  $\widehat{H}_{\text{par}}^{-1}(I \times U)$  but cannot be written in the form (A-11).



**Remark A.3.** One may wonder if, in analogy with the elliptic setting, one can identify a reflexive subspace  $E$  of the space of distributions such that the standard heat operator  $(\partial_t - \Delta)$  maps  $E$  to its dual  $E^*$  surjectively. This is however not possible, as we now explain briefly. Observe first that by Proposition A.1, the heat operator is a bijective mapping from  $H_{\text{par},\sqcup}^1(I \times U)$  to  $L^2(I; H^{-1}(U))$ , and that  $L^2(I; H^{-1}(U))$  is strictly smaller than the dual of  $H_{\text{par},\sqcup}^1(I \times U)$ . Indeed, the dual of  $H_{\text{par},\sqcup}^1(I \times U)$  contains all elements of the form  $\partial_t v$  for  $v \in L^2(I; H_0^1(U))$ . Hence, the space  $E$  should be strictly between the spaces  $H_{\text{par},\sqcup}^1(I \times U)$  and  $L^2(I; H_0^1(U))$ . Using the decomposition of the solution operator in (A-13), one can then verify that such a space  $E$  does not exist.

Before turning to the proof of Proposition A.1, we first recall the following continuity result for elements of a space intermediate between  $H_{\text{par},\sqcup}^1(I \times U)$  and  $H_{\text{par}}^1(I \times U)$  where the null boundary condition is only imposed in the space direction. We refer to [Temam 1979, Section III.1.4] for a proof.

**Lemma A.4.** *Let  $u \in L^2(I; H_0^1(U))$  be such that  $\partial_t u \in L^2(I; H^{-1}(U))$ . There exists  $\tilde{u} \in C(\bar{I}; L^2(U))$  such that, for almost every  $t \in I$ , we have  $u(t, \cdot) = \tilde{u}(t, \cdot)$ .*

From now on, whenever a function  $u$  satisfies the conditions of Lemma A.4, we identify it with its continuous representative.

*Proof of Proposition A.1.* We decompose the proof into four steps.

*Step 1.* We show that the mapping in (A-9) is uniformly convex. We will in fact prove the stronger statement that the mapping

$$(u, \mathbf{g}) \mapsto \int_{I \times U} (A(\nabla u, \mathbf{g}, \cdot) - \nabla u \cdot \mathbf{g}), \quad (\text{A-14})$$

defined over all pairs  $(u, \mathbf{g})$  in the set

$$\{(u, \mathbf{g}) \in (w + H_{\text{par},\sqcup}^1(I \times U)) \times L^2(I \times U; \mathbb{R}^d) \text{ and } -\nabla \cdot \mathbf{g} = w^* - \partial_t u\}, \quad (\text{A-15})$$

is uniformly convex. We first show that the mapping

$$(u, \mathbf{g}) \mapsto - \int_{I \times U} \nabla u \cdot \mathbf{g}$$

is convex over the set defined in (A-15). By (A-8) (with  $u^*$  replaced by  $w^*$ ), we have

$$\begin{aligned} - \int_{I \times U} \nabla u \cdot \mathbf{g} &= - \int_{I \times U} \nabla w \cdot \mathbf{g} + \int_{I \times U} (w - u)(w^* - \partial_t u) \\ &= - \int_{I \times U} \nabla w \cdot \mathbf{g} + \int_{I \times U} (w - u)(w^* - \partial_t w) + \frac{1}{2} \|(u - w)(T, \cdot)\|_{L^2(U)}^2. \end{aligned} \quad (\text{A-16})$$

This expression is clearly convex in the pair  $(u, \mathbf{g})$ . We now complete this step by showing that the mapping

$$(u, \mathbf{g}) \mapsto \int_{I \times U} A(\nabla u, \mathbf{g}, \cdot)$$

is uniformly convex over the set defined in (A-15). By (A-3), for every  $(u, \mathbf{g})$  in the set defined in (A-15) and

$$(v, \mathbf{h}) \in H_{\text{par}, \sqcup}^1(I \times U) \times L^2(I \times U; \mathbb{R}^d) \quad \text{such that } \nabla \cdot \mathbf{h} = \partial_t v, \quad (\text{A-17})$$

we have

$$\frac{1}{2}A(\nabla(u+v), \mathbf{g} + \mathbf{h}, \cdot) + \frac{1}{2}A(\nabla(u-v), \mathbf{g} - \mathbf{h}, \cdot) - A(\nabla u, \mathbf{g}, \cdot) \geq \frac{1}{2\Lambda}(|\nabla v|^2 + |\mathbf{h}|^2).$$

Moreover, by (A-17),

$$\begin{aligned} \|\partial_t v\|_{L^2(I; H^{-1}(U))} &= \sup \left\{ \int_{I \times U} \phi \partial_t v : \phi \in L^2(I; H_0^1(U)), \|\nabla \phi\|_{L^2(I \times U)} \leq 1 \right\} \\ &= \sup \left\{ \int_{I \times U} \nabla \phi \cdot \mathbf{h} : \phi \in L^2(I; H_0^1(U)), \|\nabla \phi\|_{L^2(I \times U)} \leq 1 \right\} \leq \|\mathbf{h}\|_{L^2(I \times U)}. \end{aligned}$$

We have thus shown that

$$\begin{aligned} \int_{I \times U} \left( \frac{1}{2}A(\nabla(u+v), \mathbf{g} + \mathbf{h}, \cdot) + \frac{1}{2}A(\nabla(u-v), \mathbf{g} - \mathbf{h}, \cdot) - A(\nabla u, \mathbf{g}, \cdot) \right) \\ \geq \frac{1}{4\Lambda} (\|\nabla v\|_{L^2(I \times U)}^2 + \|\partial_t v\|_{L^2(I; H^{-1}(U))}^2 + \|\mathbf{h}\|_{L^2(I \times U)}^2), \end{aligned}$$

so the proof of uniform convexity is complete.

*Step 2.* By the result of the previous step, there exists a unique pair  $(u_0, \mathbf{g}_0)$  in the set defined by (A-15) which minimizes the functional in (A-14). In order to complete the proof, it suffices to show that

$$\int_{I \times U} (A(\nabla u_0, \mathbf{g}_0, \cdot) - \nabla u_0 \cdot \mathbf{g}_0) = 0. \quad (\text{A-18})$$

Indeed, by (A-5), the identity (A-18) implies

$$\mathbf{g}_0 = \mathbf{a}(\nabla u_0, \cdot) \quad \text{a.e. in } I \times U,$$

and moreover, by (A-15),

$$\nabla \cdot \mathbf{g}_0 = w^* - \partial_t u_0,$$

so that  $u_0$  indeed solves

$$\partial_t u_0 - \nabla \cdot (\mathbf{a}(\nabla u_0, \cdot)) = w^*$$

in the weak sense. Our goal is therefore to show (A-18). The fact that the left side of (A-18) is nonnegative is immediate from (A-5). There remains to show that this quantity is nonpositive; that is,

$$\inf_{u \in H_{\text{par}, \sqcup}^1(I \times U)} \mathcal{J}[w + u, w^*] \leq 0. \quad (\text{A-19})$$

In order to do so, we consider the perturbed convex minimization problem defined for every  $u^* \in L^2(I; H^{-1}(U))$  by

$$G(u^*) := \inf_{u \in H_{\text{par}, \sqcup}^1(I \times U)} \left( \mathcal{J}[w + u, w^* + u^*] + \int_{I \times U} u u^* \right).$$

Note that (A-19) is equivalent to the statement that  $G(0) \leq 0$ . By the computation in (A-16), for every  $u^* \in L^2(I; H^{-1}(U))$  and

$$(u, \mathbf{g}) \in H_{\text{par}, \sqcup}^1(I \times U) \times L^2(I \times U; \mathbb{R}^d) \quad \text{such that} \quad -\nabla \cdot \mathbf{g} = w^* + u^* - \partial_t(w + u), \quad (\text{A-20})$$

we have

$$\begin{aligned} & \int_{I \times U} (A(\nabla(w + u), \mathbf{g}, \cdot) - \nabla(w + u) \cdot \mathbf{g}) + \int_{I \times U} u u^* \\ &= \int_{I \times U} (A(\nabla(w + u), \mathbf{g}, \cdot) - \nabla w \cdot \mathbf{g} - u(w^* - \partial_t w)) + \frac{1}{2} \|u(T, \cdot)\|_{L^2(U)}^2, \end{aligned} \quad (\text{A-21})$$

and hence the function  $G$  is convex over  $L^2(I; H^{-1}(U))$ . Moreover, one can check that it is also locally bounded above, which implies that  $G$  is lower semicontinuous, by convexity; see, e.g., [Ekeland and Temam 1976, Lemma I.2.1 and Corollary I.2.2]. Denoting by  $G^*$  the convex dual of  $G$ , defined for every  $v \in L^2(I; H_0^1(U))$  by

$$G^*(v) := \sup_{u^* \in L^2(I; H^{-1}(U))} \left( -G(u^*) + \int_{I \times U} v u^* \right),$$

and by  $G^{**}$  its bidual, we deduce that  $G = G^{**}$ , see [Ekeland and Temam 1976, Proposition I.4.1], and in particular,

$$G(0) = G^{**}(0) = \sup_{v \in L^2(I; H_0^1(U))} (-G^*(v)).$$

The statement (A-19) is therefore equivalent to

$$\text{for all } v \in L^2(I; H_0^1(U)), \quad G^*(v) \geq 0. \quad (\text{A-22})$$

The proof of this fact occupies the next two steps.

*Step 3.* For each  $v \in L^2(I; H_0^1(U))$ , we have  $G^*(v) \in \mathbb{R} \cup \{+\infty\}$ . In this step, we show that

$$G^*(v) < +\infty \quad \implies \quad \partial_t v \in L^2(I; H^{-1}(U)). \quad (\text{A-23})$$

We note that

$$G^*(v) = \sup \left\{ \int_{I \times U} ((v - u) u^* - A(\nabla(w + u), \mathbf{g}, \cdot) + \nabla(w + u) \cdot \mathbf{g}) : \right. \\ \left. u^* \in L^2(I; H^{-1}(U)), (u, \mathbf{g}) \text{ satisfy (A-20)} \right\}. \quad (\text{A-24})$$

Restricting to  $u^* = \partial_t u$  and to a fixed  $\mathbf{g} \in L^2(I \times U; \mathbb{R}^d)$  satisfying  $-\nabla \cdot \mathbf{g} = w^* - \partial_t w$  (which can be constructed as the gradient of the solution of a Dirichlet problem) yields the lower bound

$$G^*(v) \geq \sup \left\{ \int_{I \times U} (v \partial_t u - A(\nabla(w + u), \mathbf{g}, \cdot) + \nabla(w + u) \cdot \mathbf{g}) - \frac{1}{2} \|u(T, \cdot)\|_{L^2(U)}^2 : u \in H_{\text{par}, \sqcup}^1(I \times U) \right\}.$$

The assumption of  $G^*(v) < \infty$  thus implies

$$\sup \left\{ \int_{I \times U} v \partial_t u : u \in H_{\text{par}, \sqcup}^1(I \times U), \|\nabla u\|_{L^2(I \times U)} \leq 1, \|u(T, \cdot)\|_{L^2(U)} \leq 1 \right\} < \infty.$$

Denoting the supremum above by  $C < \infty$ , we infer that for every smooth test function  $u$  with compact support in  $I \times U$ ,

$$\left| \int_{I \times U} u \partial_t v \right| \leq C \|\nabla u\|_{L^2(I \times U)}.$$

By density, we deduce that  $\partial_t v$  can be identified with an element of the dual of  $L^2(I; H_0^1(U))$ . Since this dual space is  $L^2(I; H^{-1}(U))$ , the proof of (A-23) is complete.

*Step 4.* In this step, we show that

$$v \in L^2(I; H_0^1(U)) \text{ and } \partial_t v \in L^2(I; H^{-1}(U)) \implies G^*(v) \geq 0. \quad (\text{A-25})$$

Together with (A-23), this would complete the proof of (A-22) and therefore of the proposition.

The fact that  $G^*(v) \geq 0$  would follow immediately from (A-24) if we could choose  $u = v$  and then ensure the equality of the last two terms under the integral. The difficulty we face is that the function  $u$  is allowed to range in  $H_{\text{par}, \sqcup}^1(I \times U)$ , while the function  $v$  does not belong to this space in general, due to the boundary condition at the initial time. We therefore wish to argue that this constraint on  $u$  can be relaxed.

Replacing  $u^*$  by  $u^* + \partial_t u$  in the supremum in (A-24), we can rewrite  $G^*(v)$  as

$$G^*(v) = \sup \left\{ \int_{I \times U} ((v - u)(u^* + \partial_t u) - A(\nabla(w + u), \mathbf{g}, \cdot) + \nabla(w + u) \cdot \mathbf{g}) \right\}, \quad (\text{A-26})$$

where the supremum is taken over every  $u^* \in L^2(I; H^{-1}(U))$ ,  $u \in H_{\text{par}, \sqcup}^1(I \times U)$  and  $\mathbf{g} \in L^2(I \times U; \mathbb{R}^d)$  satisfying

$$-\nabla \cdot \mathbf{g} = w^* + u^* - \partial_t w. \quad (\text{A-27})$$

Integrating by parts, we can rewrite the term involving  $\partial_t u$  on the right side of (A-26) as

$$\int_{I \times U} (v - u) \partial_t u = - \int_{I \times U} u \partial_t v + \int_U u(T, \cdot) v(T, \cdot) - \frac{1}{2} \|u(T, \cdot)\|_{L^2(U)}^2.$$

The functional under the supremum in (A-26) can thus be decomposed into the sum of

$$I_1(u, u^*, \mathbf{g}) := \int_{I \times U} ((v - u)u^* - u \partial_t v - A(\nabla(w + u), \mathbf{g}, \cdot) + \nabla(w + u) \cdot \mathbf{g}) \quad (\text{A-28})$$

and

$$I_2(u(T, \cdot)) := \int_U u(T, \cdot) v(T, \cdot) - \frac{1}{2} \|u(T, \cdot)\|_{L^2(U)}^2. \quad (\text{A-29})$$

Moreover, for each given  $u^* \in L^2(I; H^{-1}(U))$  and  $\mathbf{g} \in L^2(I \times U; \mathbb{R}^d)$ , the mapping  $u \mapsto I_1(u, u^*, \mathbf{g})$  is continuous for the topology of  $L^2(I; H^1(U))$ . For any given  $b \in H_0^1(U)$  and  $\tilde{u} \in L^2(I; H_0^1(U))$ , one can find elements of the space

$$\{u \in H_{\text{par}, \sqcup}^1(I \times U) : u(T, \cdot) = b\}$$

which approximate  $\tilde{u}$  with arbitrary precision for the topology of  $L^2(I; H^1(U))$ . Hence, for each given  $u^* \in L^2(I; H^{-1}(U))$  and  $\mathbf{g} \in L^2(I \times U; \mathbb{R}^d)$ , we have

$$\begin{aligned} & \sup\{I_1(u, u^*, \mathbf{g}) + I_2(u(T, \cdot)) : u \in H_{\text{par}, \square}^1(I \times U)\} \\ & \geq \sup\{I_1(u, u^*, \mathbf{g}) + I_2(b) : u \in L^2(I; H_0^1(U)) \text{ and } b \in H_0^1(U)\}. \end{aligned}$$

Moreover, the mapping  $b \mapsto I_2(b)$  is continuous for the topology of  $L^2(U)$ , and thus we have in fact

$$\begin{aligned} & \sup\{I_1(u, u^*, \mathbf{g}) + I_2(u(T, \cdot)) : u \in H_{\text{par}, \square}^1(I \times U)\} \\ & = \sup\{I_1(u, u^*, \mathbf{g}) + I_2(b) : u \in L^2(I; H_0^1(U)) \text{ and } b \in L^2(U)\}. \end{aligned}$$

Selecting  $u = v$  and  $b = v(T, \cdot)$ , we have thus shown

$$G^*(v) \geq \sup\left\{ \frac{1}{2} \|v(T, \cdot)\|_{L^2(U)}^2 + \int_{I \times U} (-v \partial_t v - A(\nabla(w+v), \mathbf{g}, \cdot) + \nabla(w+v) \cdot \mathbf{g}) \right\},$$

where the supremum is taken over every  $u^* \in L^2(I; H^{-1}(U))$  and  $\mathbf{g} \in L^2(I \times U; \mathbb{R}^d)$  satisfying (A-27). Note that

$$\frac{1}{2} \|v(T, \cdot)\|_{L^2(U)}^2 - \int_{I \times U} v \partial_t v = \frac{1}{2} \|v(0, \cdot)\|_{L^2(U)}^2 \geq 0.$$

Selecting  $u^*$  such that

$$-\nabla \cdot (\mathbf{a}(\nabla(w+v), \cdot)) = w^* + u^* - \partial_t w,$$

and then

$$\mathbf{g} = \mathbf{a}(\nabla(w+v), \cdot),$$

ensures that the constraint (A-27) is satisfied, and by (A-6), that

$$\int_{I \times U} (A(\nabla(w+v), \mathbf{g}, \cdot) - \nabla(w+v) \cdot \mathbf{g}) = 0.$$

The proof of (A-25) is therefore complete.  $\square$

## Appendix B. Meyers-type estimates

In this appendix, we present local and global versions of the Meyers improvement of integrability estimate for gradients of solutions of linear, uniformly parabolic equations with measurable coefficients.

The interior Meyers estimate in the parabolic case was first proved in [Giaquinta and Struwe 1982]. We follow their argument to obtain Proposition B.1 below, which is included for completeness and since the same ideas are needed to prove the global version in Proposition B.2. The statement of the latter will certainly not come as a surprise to experts, but we do not believe it has appeared before. Global versions of the Meyers estimate in the parabolic setting have been previously considered in [Parviainen 2009], but the statement of Proposition B.2 is stronger than the results of that paper since we do not require any additional regularity of the boundary condition in time — a modest technical improvement, but it gives a more natural statement and one which is useful for the application in this paper.

In what follows, we use the same notation for parabolic cylinders as in Section 6; see (1-7). That is, for  $(t, x) \in \mathbb{R} \times \mathbb{R}^d$ , we define

$$\tilde{I}_r := (-r^2, 0], \quad Q_r(t, x) := (t, x) + \tilde{I}_r \times B_r, \quad \text{and} \quad Q_r := Q_r(0, 0).$$

We fix a coefficient field  $\mathbf{a} = \mathbf{a}(t, x)$  satisfying (1-2) for every  $(t, x) \in \mathbb{R} \times \mathbb{R}^d$ , and consider the linear parabolic equation

$$\partial_t u - \nabla \cdot (\mathbf{a}(t, x) \nabla u) = u^*. \quad (\text{B-1})$$

We remark that the argument we present only makes mild use of linearity and can be adapted to give similar estimates for solutions of nonlinear parabolic equations like the ones considered in Appendix A.

We first present the interior Meyers estimate. Recall that the space  $W_{\text{par}}^{1,p}$  is defined in (1-13) and (1-14).

**Proposition B.1** (interior Meyers estimate [Giaquinta and Struwe 1982, Theorem 2.1]). *Fix  $r > 0$ ,  $p \geq 2$  and suppose  $u \in H_{\text{par}}^1(Q_{2r})$  and  $u^* \in L^p(I_{2r}; W^{-1,p}(B_{2r}))$  satisfy (B-1) in  $Q_{2r}$ . There exist an exponent  $\delta(d, \Lambda) > 0$  and a constant  $C(d, \Lambda) < \infty$  such that  $u \in W_{\text{par}}^{1,p \wedge (2+\delta)}(Q_r)$  and we have the estimate*

$$\|\nabla u\|_{L^{p \wedge (2+\delta)}(Q_r)} \leq C(\|\nabla u\|_{L^2(Q_{2r})} + \|u^*\|_{L^{p \wedge (2+\delta)}(I_{2r}; W^{-1,p \wedge (2+\delta)}(B_{2r}))}). \quad (\text{B-2})$$

We next give a global statement of the Meyers estimate with respect to a Cauchy–Dirichlet initial-boundary condition.

**Proposition B.2** (global Meyers estimate). *Fix  $p \geq 2$ . Let  $U \in \mathbb{R}^d$  be a bounded Lipschitz domain,  $I \subseteq \mathbb{R}$  a bounded interval and set  $V := I \times U$ . Fix  $f \in W_{\text{par}}^{1,p}(V)$ ,  $u^* \in L^p(I; W^{-1,p}(V))$  and suppose*

$$u \in f + H_{\text{par}, \square}^1(V)$$

*is the unique solution of the Cauchy–Dirichlet problem*

$$\begin{cases} \partial_t u - \nabla \cdot (\mathbf{a} \nabla u) = u^* & \text{in } V, \\ u = f & \text{on } \partial_{\square} V. \end{cases}$$

*There exist  $\delta(V, d, \Lambda) > 0$  and a constant  $C(V, d, \Lambda) < \infty$  such that  $u \in W_{\text{par}}^{1,p \wedge (2+\delta)}(V)$  and we have the estimate*

$$\|u\|_{W_{\text{par}}^{1,p \wedge (2+\delta)}(V)} \leq C(\|f\|_{W_{\text{par}}^{1,p \wedge (2+\delta)}(V)} + \|u^*\|_{L^{p \wedge (2+\delta)}(I; W^{-1,p \wedge (2+\delta)}(V))}). \quad (\text{B-3})$$

The Meyers estimates are consequences of the Caccioppoli inequality, the most basic regularity estimate for divergence-form equations.

**Lemma B.3** (parabolic Caccioppoli inequality). *Suppose  $u \in H_{\text{par}}^1(Q_{2r})$  and  $u^* \in L^2(I_{2r}; H^{-1}(B_{2r}))$  satisfy*

$$\partial_t u - \nabla \cdot (\mathbf{a} \nabla u) = u^* \quad \text{in } Q_{2r}.$$

*Then there exists  $C(d, \Lambda) < \infty$  such that*

$$\|\nabla u\|_{L^2(Q_r)} \leq Cr^{-1} \|u\|_{L^2(Q_{2r})} + C \|u^*\|_{L^2(I_{2r}; H^{-1}(B_{2r}))} \quad (\text{B-4})$$

*and*

$$\sup_{s \in I_r} \|u(s, \cdot)\|_{L^2(B_r)} \leq C \|\nabla u\|_{L^2(Q_{2r})} + C \|u^*\|_{L^2(I_{2r}; H^{-1}(B_{2r}))}. \quad (\text{B-5})$$

*Proof.* We take  $\eta_r \in C_c^\infty(Q_{2r})$  to be a test function satisfying

$$0 \leq \eta \leq 1, \quad \eta \equiv 1 \quad \text{on } Q_r, \quad |\partial_t \eta| + |\nabla \eta|^2 \leq Cr^{-2}.$$

We test the weak formulation

$$\text{for all } \phi \in L^2(I_{2r}; H_0^1(B_{2r})), \quad \int_{Q_{2r}} \phi(u^* - \partial_t u) = \int_{Q_{2r}} \nabla \phi \cdot \mathbf{a} \nabla u$$

with the function  $\phi := \eta_r^2 u \in L^2(I_{2r}; H_0^1(B_{2r}))$ . We estimate the right side from below by

$$\begin{aligned} \int_{Q_{2r}} \nabla \phi \cdot \mathbf{a} \nabla u &\geq \frac{1}{\Lambda} \int_{Q_{2r}} \eta_r^2 |\nabla u|^2 - C \int_{Q_{2r}} \eta_r |\nabla \eta_r| |u| |\nabla u| \\ &\geq \frac{1}{2\Lambda} \int_{Q_{2r}} \eta_r^2 |\nabla u|^2 - C \int_{Q_{2r}} |\nabla \eta_r|^2 |u|^2 \\ &\geq \frac{1}{2\Lambda} \int_{Q_{2r}} \eta_r^2 |\nabla u|^2 - Cr^{-2} \int_{Q_{2r}} |u|^2 \end{aligned}$$

and the left side from above by

$$\begin{aligned} \int_{Q_{2r}} \eta_r^2 u(u^* - \partial_t u) &\leq - \int_{Q_{2r}} \partial_t \left( \frac{1}{2} \eta_r^2 u^2 \right) + \int_{Q_{2r}} \eta_r |\partial_t \eta_r| u^2 + \int_{-4r^2}^0 \|(\eta_r^2 u)(t, \cdot)\|_{H^1(B_{2r})} \|u^*(t, \cdot)\|_{H^{-1}(B_{2r})} dt \\ &\leq -\frac{1}{2} \int_{B_{2r}} \eta_r^2(0, x) u^2(0, x) dx + Cr^{-2} \int_{Q_{2r}} u^2 + C \|\eta_r^2 u\|_{L^2(I_{2r}; H^1(B_{2r}))} \|u^*\|_{L^2(I_{2r}; H^{-1}(B_{2r}))}. \end{aligned}$$

Using that

$$\|\eta_r^2 u\|_{L^2(I_{2r}; H^1(B_{2r}))} \leq Cr^{-1} \|u\|_{L^2(I_{2r} \times B_{2r})} + C \|\eta_r \nabla u\|_{L^2(I_{2r} \times B_{2r})},$$

we get

$$\begin{aligned} C \|\eta_r^2 u\|_{L^2(I_{2r}; H^1(B_{2r}))} \|u^*\|_{L^2(I_{2r}; H^{-1}(B_{2r}))} \\ \leq r^{-2} \|u\|_{L^2(I_{2r} \times B_{2r})}^2 + \frac{1}{4\Lambda} \|\eta_r \nabla u\|_{L^2(I_{2r} \times B_{2r})}^2 + C \|u^*\|_{L^2(I_{2r}; H^{-1}(B_{2r}))}^2. \end{aligned}$$

Combining the above, we get

$$\frac{1}{2} \int_{B_{2r}} \eta_r^2(0, x) u^2(0, x) dx + \frac{1}{4\Lambda} \int_{Q_{2r}} \eta_r^2 |\nabla u|^2 \leq Cr^{-2} \int_{Q_{2r}} |u|^2 + C \|u^*\|_{L^2(I_{2r}; H^{-1}(B_{2r}))}^2.$$

This yields (B-4).

By repeating the above computation, using instead the test function  $\phi := \eta_r^2 u \mathbb{1}_{\{t < s\}}$  for fixed  $s \in I_{2r}$ , and estimating the right side of the weak formulation from below differently, namely

$$\begin{aligned} \int_{Q_{2r}} \nabla \phi \cdot \mathbf{a} \nabla u &\geq -C \|\eta_r \nabla u\|_{L^2(Q_{2r})}^2 - C \|\nabla \eta_r \nabla u\|_{L^2(Q_{2r})} \|u \eta_r\|_{L^2(Q_{2r})} \\ &\geq -C \|\nabla u\|_{L^2(Q_{2r})}^2 - \frac{1}{16} r^{-2} \int_{Q_{2r}} \eta_r^2 u^2 \\ &\geq -C \|\nabla u\|_{L^2(Q_{2r})}^2 - \frac{1}{4} \sup_{t \in I_{2r}} \int_{B_{2r}} \eta_r^2(t, x) u^2(t, x) dx, \end{aligned}$$

we get the bound

$$\frac{1}{2} \int_{B_{2r}} \eta_r^2(s, x) u^2(s, x) dx \leq C \|\nabla u\|_{L^2(Q_{2r})}^2 + \frac{1}{4} \sup_{t \in I_{2r}} \int_{B_{2r}} \eta_r^2(t, x) u^2(t, x) dx + C \|u^*\|_{L^2(I_{2r}; H^{-1}(B_{2r}))}^2.$$

Taking the supremum over  $s \in I_{2r}$  and rearranging, we get (B-5).  $\square$

In the following statement, what is important is that  $q < 2$ . It is convenient to use the Sobolev exponent  $q := 2_*$ , although the choice  $q = 1$  in  $d = 2$  causes technical problems so in that case we just take  $q \in (\frac{5}{4}, \frac{7}{4})$ .

**Lemma B.4** (reverse Hölder inequality). *Suppose  $u \in H_{\text{par}}^1(Q_{4r})$  and  $u^* \in L^2(I_{4r}; H^{-1}(B_{4r}))$  satisfy*

$$\partial_t u - \nabla \cdot (\mathbf{a}(x) \nabla u) = u^* \quad \text{in } Q_{4r}.$$

*Set  $q := 2_* = 2d/(2+d)$  if  $d > 2$  or let  $q$  be any element of  $(\frac{5}{4}, \frac{7}{4})$  if  $d = 2$ . Then there exists  $C(d, \Lambda) < \infty$  such that, for every  $\alpha > 0$ ,*

$$\|\nabla u\|_{L^2(Q_r)}^2 \leq \frac{C}{\alpha} \|\nabla u\|_{L^q(Q_{4r})}^2 + \alpha \|\nabla u\|_{L^2(Q_{4r})}^2 + C \|u^*\|_{L^2(I_{4r}; H^{-1}(B_{4r}))}^2. \quad (\text{B-6})$$

*Proof.* By subtracting a constant, we may suppose  $(u)_{Q_{2r}} = 0$ . Let  $\xi \in C_c^\infty(B_r)$  with  $\int_{B_r} \xi = 1$  and  $|\nabla \xi| \leq Cr^{-1}$ . Define

$$v(t, x) := u(t, x) - w(t), \quad w(t) := \int_{B_r} \xi(y) u(t, y) dy.$$

Then  $v$  satisfies

$$\partial_t v - \nabla \cdot (\mathbf{a} \nabla v) = u^* - \partial_t w.$$

Applying (B-5) to  $v$ , we find that

$$\begin{aligned} \int_{Q_{2r}} |v|^2 &\leq \left( \sup_{s \in I_{2r}} \int_{B_{2r}} |v(s, x)|^2 dx \right)^{\frac{1}{2}} \int_{I_{2r}} \left( \int_{B_{2r}} |v(t, x)|^2 dx \right)^{\frac{1}{2}} dt \\ &\leq C (\|\nabla u\|_{L^2(Q_{4r})} + \|u^* - \partial_t w\|_{L^2(I_{4r}; H^{-1}(B_{4r}))}) \int_{I_{2r}} \left( \int_{B_{2r}} |v(t, x)|^2 dx \right)^{\frac{1}{2}} dt. \end{aligned}$$

Denote by  $q'$  the Hölder conjugate exponent to  $q$  and notice that  $q' = 2^*$  in  $d > 2$  and  $q' < \infty$  in  $d = 2$ . Using the Hölder and Sobolev inequalities, we find that

$$\begin{aligned} \int_{I_{2r}} \left( \int_{B_{2r}} |v(t, x)|^2 dx \right)^{\frac{1}{2}} dt &\leq \int_{I_{2r}} \left( \int_{B_{2r}} |v(t, x)|^q dx \right)^{\frac{1}{2q}} \left( \int_{B_{2r}} |v(t, x)|^{q'} dx \right)^{\frac{1}{2q'}} dt \\ &\leq Cr^{1+d(\frac{1}{4}-\frac{1}{2q})} \int_{I_{2r}} \left( \int_{B_{2r}} |\nabla v(t, x)|^q dx \right)^{\frac{1}{2q}} \left( \int_{B_{2r}} |\nabla v(t, x)|^2 dx \right)^{\frac{1}{4}} dt \\ &\leq Cr^{1+d(\frac{1}{4}-\frac{1}{2q})} \|\nabla v\|_{L^q(Q_{2r})}^{\frac{1}{2}} \left( \int_{I_{2r}} \left( \int_{B_{2r}} |\nabla v(t, x)|^2 dx \right)^{\frac{(2q)'}{4}} dt \right)^{\frac{1}{(2q)'}}. \end{aligned}$$



As  $\frac{1}{4}(2q)' \leq \frac{1}{2} < 1$ , we can use Hölder's inequality in time and then (B-4) and Lemma 3.1 to get

$$\begin{aligned} \left( \int_{I_{2r}} \left( \int_{B_{2r}} |\nabla v(t, x)|^2 dx \right)^{\frac{(2q)'}{4}} dt \right)^{\frac{2}{(2q)'}} &\leq Cr^{\frac{2}{(2q)'} - \frac{1}{2}} \left( \int_{I_{2r}} \int_{B_{2r}} |\nabla v(t, x)|^2 dx dt \right)^{\frac{1}{2}} \\ &\leq Cr^{\frac{2}{(2q)' - \frac{1}{2}}} (\|\nabla u\|_{L^2(Q_{4r})} + \|u^* - \partial_t w\|_{L^2(I_{4r}; H^{-1}(B_{4r}))}). \end{aligned}$$

Let  $\kappa := d\left(\frac{1}{4} - \frac{1}{2q}\right) + \frac{1}{(2q)'} + \frac{3}{4}$ . Combining the above, we get

$$\|v\|_{L^2(Q_{2r})}^2 \leq Cr^\kappa \|\nabla v\|_{L^q(Q_{2r})}^{\frac{1}{2}} (\|\nabla u\|_{L^2(Q_{4r})} + \|u^* - \partial_t w\|_{L^2(I_{4r}; H^{-1}(B_{4r}))})^{\frac{3}{2}}.$$

Combining (B-4) and the previous inequality, we obtain

$$\|\nabla v\|_{L^2(Q_r)}^2 \leq Cr^{\kappa-2} \|\nabla v\|_{L^q(Q_{2r})}^{\frac{1}{2}} (\|\nabla v\|_{L^2(Q_{4r})} + \|u^* - \partial_t w\|_{L^2(I_{4r}; H^{-1}(B_{4r}))})^{\frac{3}{2}} + C \|u^* - \partial_t w\|_{L^2(I_{4r}; H^{-1}(B_{4r}))}^2.$$

Normalizing the norms, we find that this is the same as

$$\|\nabla v\|_{\underline{L}^2(Q_r)}^2 \leq \|\nabla v\|_{\underline{L}^q(Q_{2r})}^{\frac{1}{2}} (\|\nabla v\|_{\underline{L}^2(Q_{4r})} + \|u^* - \partial_t w\|_{\underline{L}^2(I_{4r}; \underline{H}^{-1}(B_{4r}))})^{\frac{3}{2}} + C \|u^* - \partial_t w\|_{\underline{L}^2(I_{4r}; \underline{H}^{-1}(B_{4r}))}^2.$$

Applying Young's inequality, we obtain, for every  $\alpha > 0$ ,

$$\|\nabla v\|_{\underline{L}^2(Q_r)}^2 \leq \frac{C}{\alpha} \|\nabla v\|_{\underline{L}^q(Q_{4r})}^2 + \alpha \|\nabla v\|_{\underline{L}^2(Q_{4r})}^2 + C \|u^* - \partial_t w\|_{\underline{L}^2(I_{4r}; \underline{H}^{-1}(B_{4r}))}^2. \quad (\text{B-7})$$

It is not difficult to show, by using the equation and the definition of  $w$ , that

$$\|\partial_t w\|_{\underline{L}^2(I_{4r}; \underline{H}^{-1}(B_{4r}))} \leq C (\|\nabla u\|_{\underline{L}^1(Q_{4r})} + \|u^*\|_{\underline{L}^2(I_{4r}; \underline{H}^{-1}(B_{4r}))}).$$

Combining the previous two displays yields

$$\|\nabla v\|_{\underline{L}^2(Q_r)}^2 \leq \frac{C}{\alpha} \|\nabla v\|_{\underline{L}^q(Q_{4r})}^2 + \alpha \|\nabla v\|_{\underline{L}^2(Q_{4r})}^2 + C \|u^*\|_{\underline{L}^2(I_{4r}; \underline{H}^{-1}(B_{4r}))}^2.$$

Since  $\nabla v = \nabla u$ , this completes the argument.  $\square$

To complete the proof of the interior Meyers estimate, we need the following version of Gehring's lemma for parabolic cylinders which states that a reverse Hölder inequality implies an improvement of integrability. This result is standard and so we do not give the proof here. See for instance [Giaquinta and Modica 1979, Proposition 5.1], where the statement is given in cubes rather than parabolic cylinders (which makes no difference in its proof).

**Lemma B.5** (Gehring-type lemma). *Assume that  $R > 0$ ,  $q > 1$ ,  $F \in L^1(Q_{4R})$ ,  $G \in L^q(Q_{4R})$ ,  $m \in (0, 1)$  and  $A \in [1, \infty)$ . Suppose that, for every  $(t, x) \in Q_R$  and  $r \in (0, \frac{1}{2}R]$ ,*

$$\|F\|_{L^1(Q_r(t,x))} \leq A (\|F^m\|_{L^1(Q_{4r}(t,x))}^{\frac{1}{m}} + \|G\|_{L^1(Q_{4r}(t,x))}) + \varepsilon \|F\|_{L^1(Q_{4r}(t,x))}.$$

*Then there exists  $\varepsilon_0(d, m) \in (0, \frac{1}{2}]$  such that  $\varepsilon \leq \varepsilon_0$  implies the existence of an exponent  $\delta(\varepsilon, A, m, q, d) \in (0, \frac{1}{2}]$  and  $C(\varepsilon, A, m, d) < \infty$  such that  $F \in L^{1+\delta}(Q_R)$  and*

$$\|F\|_{L^{1+\delta}(Q_R)} \leq C (\|F\|_{L^1(Q_{4R})} + \|G\|_{L^{1+\delta}(Q_{4R})}).$$

The statement of Proposition B.1 can now be obtained as a consequence of Lemmas B.4 and B.5 and a routine covering argument. Indeed, any element of  $W^{-1,p}(U)$  can be represented as the divergence of an element of  $L^p(U; \mathbb{R}^d)$  by the Riesz representation theorem; see [Adams and Fournier 2003, Theorem 3.9]. This allows us to obtain the estimate (B-2). The statement that  $\partial_t u$  belongs to  $L^{p \wedge (2+\delta)}(I_r; W^{-1,p \wedge (2+\delta)}(B_r))$ , with an appropriate estimate, follows from (B-2) and (B-1).

We next give a sketch of the proof of Proposition B.2, which requires us to first revisit the proof of the Caccioppoli inequality to obtain a global version.

**Lemma B.6** (global Caccioppoli inequality). *Let  $U \subseteq \mathbb{R}^d$  be a bounded Lipschitz domain and define  $V := I_2 \times U$ . Suppose  $f \in H_{\text{par}}^1(V)$ ,  $u \in f + H_{\text{par}, \square}^1(V)$  and  $u^* \in L^2(I_2; H^{-1}(U))$  satisfy*

$$\partial_t u - \nabla \cdot (\mathbf{a}(x)\nabla u) = u^* \quad \text{in } V.$$

Then there exists  $C(V, d, \Lambda) < \infty$  such that, for every  $r \in (0, 1)$  and  $(t, x) \in I_1 \times U$ ,

$$\begin{aligned} \|\nabla(u - f)\|_{L^2(Q_r(t,x) \cap V)} & \\ & \leq Cr^{-1} \|u - f\|_{L^2(Q_{2r}(t,x) \cap V)} + C\|\nabla f\|_{L^2(Q_{2r}(t,x) \cap V)} + C\|u^*\|_{L^2((t+I_{2r}) \cap I_2; H^{-1}(B_{2r}(x) \cap U))} \end{aligned} \quad (\text{B-8})$$

and

$$\begin{aligned} \sup_{s \in (t+I_{2r}) \cap I_2} \|(u - f)(s, \cdot)\|_{L^2(B_r(x) \cap U)} & \\ & \leq C\|\nabla(u - f)\|_{L^2(Q_{2r}(t,x) \cap V)} + C\|\nabla f\|_{L^2(Q_{2r}(t,x) \cap V)} + C\|u^*\|_{L^2((t+I_{2r}) \cap I_2; H^{-1}(B_{2r}(x) \cap U))}. \end{aligned} \quad (\text{B-9})$$

*Proof.* By replacing  $u$  by  $\tilde{u} := u - f$  and  $u^*$  by  $\tilde{u}^* := u^* - (\partial_t - \nabla \cdot \mathbf{a}\nabla)f$ , we may assume without loss of generality that  $f = 0$ . The lemma is then obtained by repeating the argument of Lemma B.3 and making obvious adjustments to the notation.  $\square$

Following the proof of Lemma B.4, we obtain a global version of the reverse Hölder inequality.

**Lemma B.7** (reverse Hölder inequality). *Let  $U \subseteq \mathbb{R}^d$  be a bounded Lipschitz domain and define  $V := I_2 \times U$ . Suppose  $f \in H_{\text{par}}^1(V)$ ,  $u \in f + H_{\text{par}, \square}^1(V)$  and  $u^* \in L^2(I_2; H^{-1}(U))$  satisfy*

$$\partial_t u - \nabla \cdot (\mathbf{a}(x)\nabla u) = u^* \quad \text{in } V.$$

Set  $q := 2_* = 2d/(2+d)$  if  $d > 2$  or let  $q$  be any element of  $(\frac{5}{4}, \frac{7}{4})$  if  $d = 2$ . Then there exists  $C(V, d, \Lambda) < \infty$  such that, for every  $r \in (0, 1)$ ,  $(t, x) \in I_1 \times U$  and  $\alpha > 0$ ,

$$\begin{aligned} \|\nabla(u - f)\|_{L^2(Q_r(t,x) \cap V)}^2 & \leq \frac{C}{\alpha} \|\nabla(u - f)\|_{\underline{L}^q(Q_{4r}(t,x) \cap V)}^2 + \alpha \|\nabla(u - f)\|_{L^2(Q_{4r}(t,x) \cap V)}^2 \\ & \quad + \alpha \|\nabla f\|_{\underline{L}^2(Q_{4r}(t,x) \cap V)}^2 + C\|u^*\|_{L^2((t+I_{4r}) \cap I_2; H^{-1}(B_{4r}(x) \cap U))}^2. \end{aligned}$$

*Proof.* The argument is omitted, since it is an easy adaptation of the proof of Lemma B.4.  $\square$

Proposition B.2 is now a straightforward consequence of Lemmas B.5 and B.7.

## Acknowledgments

Armstrong was partially supported by the NSF Grant DMS-1700329. Mourrat was partially supported by the ANR grant LSD (ANR-15-CE40-0020-03).

## References

- [Adams and Fournier 2003] R. A. Adams and J. J. F. Fournier, *Sobolev spaces*, 2nd ed., Pure and Applied Mathematics **140**, Elsevier/Academic, Amsterdam, 2003. MR Zbl
- [Armstrong and Mourrat 2016] S. N. Armstrong and J.-C. Mourrat, “Lipschitz regularity for elliptic equations with random coefficients”, *Arch. Ration. Mech. Anal.* **219**:1 (2016), 255–348. MR Zbl
- [Armstrong and Smart 2016] S. N. Armstrong and C. K. Smart, “Quantitative stochastic homogenization of convex integral functionals”, *Ann. Sci. Éc. Norm. Supér. (4)* **49**:2 (2016), 423–481. MR Zbl
- [Armstrong et al. 2016] S. Armstrong, T. Kuusi, and J.-C. Mourrat, “Mesoscopic higher regularity and subadditivity in elliptic homogenization”, *Comm. Math. Phys.* **347**:2 (2016), 315–361. MR Zbl
- [Armstrong et al. 2017a] S. Armstrong, T. Kuusi, and J.-C. Mourrat, “The additive structure of elliptic homogenization”, *Invent. Math.* **208**:3 (2017), 999–1154. MR Zbl
- [Armstrong et al. 2017b] S. Armstrong, T. Kuusi, and J.-C. Mourrat, “Quantitative stochastic homogenization and large-scale regularity”, book preprint, 2017, available at <http://www.math.ens.fr/~mourrat/lecturenotes.pdf>.
- [Bella et al. 2018] P. Bella, A. Chiarini, and B. Fehrman, “A Liouville theorem for stationary and ergodic ensembles of parabolic systems”, *Probab. Theory Related Fields* (online publication March 2018). arXiv 1706.03440
- [Bensoussan et al. 1978] A. Bensoussan, J.-L. Lions, and G. Papanicolaou, *Asymptotic analysis for periodic structures*, Studies in Mathematics and its Applications **5**, North-Holland, Amsterdam, 1978. MR Zbl
- [Brezis and Ekeland 1976a] H. Brezis and I. Ekeland, “Un principe variationnel associé à certaines équations paraboliques: le cas indépendant du temps”, *C. R. Acad. Sci. Paris Sér. A-B* **282**:17 (1976), A971–A974. MR Zbl
- [Brezis and Ekeland 1976b] H. Brezis and I. Ekeland, “Un principe variationnel associé à certaines équations paraboliques: le cas dépendant du temps”, *C. R. Acad. Sci. Paris Sér. A-B* **282**:20 (1976), A1197–A1198. MR Zbl
- [Ekeland and Temam 1976] I. Ekeland and R. Temam, *Convex analysis and variational problems*, Studies in Mathematics and its Applications **1**, North-Holland, Amsterdam, 1976. MR Zbl
- [Evans 2010] L. C. Evans, *Partial differential equations*, 2nd ed., Graduate Studies in Mathematics **19**, American Mathematical Society, Providence, RI, 2010. MR Zbl
- [Fitzpatrick 1988] S. Fitzpatrick, “Representing monotone operators by convex functions”, pp. 59–65 in *Workshop/Miniconference on Functional Analysis and Optimization* (Canberra, 1988), edited by S. P. Fitzpatrick and J. R. Giles, Proc. Centre Math. Anal. Austral. Nat. Univ. **20**, Austral. Nat. Univ., Canberra, 1988. MR Zbl
- [Ghoussoub 2009] N. Ghoussoub, *Self-dual partial differential systems and their variational principles*, Springer, 2009. MR Zbl
- [Ghoussoub and Tzou 2004] N. Ghoussoub and L. Tzou, “A variational principle for gradient flows”, *Math. Ann.* **330**:3 (2004), 519–549. MR Zbl
- [Giaquinta and Modica 1979] M. Giaquinta and G. Modica, “Regularity results for some classes of higher order nonlinear elliptic systems”, *J. Reine Angew. Math.* **311/312** (1979), 145–169. MR Zbl
- [Giaquinta and Struwe 1982] M. Giaquinta and M. Struwe, “On the partial regularity of weak solutions of nonlinear parabolic systems”, *Math. Z.* **179**:4 (1982), 437–451. MR Zbl
- [Gloria and Otto 2011] A. Gloria and F. Otto, “An optimal variance estimate in stochastic homogenization of discrete elliptic equations”, *Ann. Probab.* **39**:3 (2011), 779–856. MR Zbl
- [Gloria and Otto 2012] A. Gloria and F. Otto, “An optimal error estimate in stochastic homogenization of discrete elliptic equations”, *Ann. Appl. Probab.* **22**:1 (2012), 1–28. MR Zbl

- [Gloria and Otto 2015] A. Gloria and F. Otto, “The corrector in stochastic homogenization: near-optimal rates with optimal stochastic integrability”, preprint, 2015. arXiv 1510.08290v2
- [Gloria and Otto 2016] A. Gloria and F. Otto, “The corrector in stochastic homogenization: optimal rates, stochastic integrability, and fluctuations”, preprint, 2016. arXiv 1510.08290v3
- [Gloria et al. 2015] A. Gloria, S. Neukamm, and F. Otto, “Quantification of ergodicity in stochastic homogenization: optimal bounds via spectral gap on Glauber dynamics”, *Invent. Math.* **199**:2 (2015), 455–515. MR Zbl
- [Jikov et al. 1994] V. V. Jikov, S. M. Kozlov, and O. A. Oleĭnik, *Homogenization of differential operators and integral functionals*, Springer, 1994. MR Zbl
- [Krylov 1982] N. V. Krylov, “Some properties of monotone mappings”, *Litovsk. Mat. Sb.* **22**:2 (1982), 80–87. In Russian; translated in *Lith. Math. J.* **22**:2 (1982), 140–145. MR Zbl
- [Nualart 2006] D. Nualart, *The Malliavin calculus and related topics*, 2nd ed., Springer, 2006. MR Zbl
- [Papanicolaou and Varadhan 1981] G. C. Papanicolaou and S. R. S. Varadhan, “Boundary value problems with rapidly oscillating random coefficients”, pp. 835–873 in *Random fields* (Esztergom, 1979), edited by J. Fritz et al., Colloq. Math. Soc. János Bolyai **27**, North-Holland, Amsterdam, 1981. MR Zbl
- [Parviainen 2009] M. Parviainen, “Global gradient estimates for degenerate parabolic equations in nonsmooth domains”, *Ann. Mat. Pura Appl.* (4) **188**:2 (2009), 333–358. MR Zbl
- [Temam 1979] R. Temam, *Navier–Stokes equations: theory and numerical analysis*, Studies in Mathematics and its Applications **2**, North-Holland, Amsterdam, 1979. MR Zbl
- [Widder 1961] D. V. Widder, “Series expansions of solutions of the heat equation in  $n$  dimensions”, *Ann. Mat. Pura Appl.* (4) **55** (1961), 389–409. MR Zbl

Received 22 May 2017. Revised 15 Feb 2018. Accepted 9 Apr 2018.

SCOTT ARMSTRONG: [scotta@cims.nyu.edu](mailto:scotta@cims.nyu.edu)

*Courant Institute of Mathematical Sciences, New York University, New York, NY, United States*

ALEXANDRE BORDAS: [alexandre.bordas@ens-lyon.fr](mailto:alexandre.bordas@ens-lyon.fr)

*Ecole normale supérieure de Lyon, Lyon, France*

JEAN-CHRISTOPHE MOURRAT: [jean-christophe.mourrat@ens-lyon.fr](mailto:jean-christophe.mourrat@ens-lyon.fr)

*Ecole normale supérieure de Lyon, CNRS, Lyon, France*

# Analysis & PDE

msp.org/apde

## EDITORS

EDITOR-IN-CHIEF

Patrick Gérard  
patrick.gerard@math.u-psud.fr  
Université Paris Sud XI  
Orsay, France

## BOARD OF EDITORS

Massimiliano Berti	Scuola Intern. Sup. di Studi Avanzati, Italy berti@sissa.it	Clément Mouhot	Cambridge University, UK c.mouhot@dpmms.cam.ac.uk
Sun-Yung Alice Chang	Princeton University, USA chang@math.princeton.edu	Werner Müller	Universität Bonn, Germany mueller@math.uni-bonn.de
Michael Christ	University of California, Berkeley, USA mchrist@math.berkeley.edu	Gilles Pisier	Texas A&M University, and Paris 6 pisier@math.tamu.edu
Alessio Figalli	ETH Zurich, Switzerland alessio.figalli@math.ethz.ch	Tristan Rivière	ETH, Switzerland riviere@math.ethz.ch
Charles Fefferman	Princeton University, USA cf@math.princeton.edu	Igor Rodnianski	Princeton University, USA irod@math.princeton.edu
Ursula Hamenstaedt	Universität Bonn, Germany ursula@math.uni-bonn.de	Sylvia Serfaty	New York University, USA serfaty@cims.nyu.edu
Vaughan Jones	U.C. Berkeley & Vanderbilt University vaughan.f.jones@vanderbilt.edu	Yum-Tong Siu	Harvard University, USA siu@math.harvard.edu
Vadim Kaloshin	University of Maryland, USA vadim.kaloshin@gmail.com	Terence Tao	University of California, Los Angeles, USA tao@math.ucla.edu
Herbert Koch	Universität Bonn, Germany koch@math.uni-bonn.de	Michael E. Taylor	Univ. of North Carolina, Chapel Hill, USA met@math.unc.edu
Izabella Laba	University of British Columbia, Canada ilaba@math.ubc.ca	Gunther Uhlmann	University of Washington, USA gunther@math.washington.edu
Gilles Lebeau	Université de Nice Sophia Antipolis, France lebeau@unice.fr	András Vasy	Stanford University, USA andras@math.stanford.edu
Richard B. Melrose	Massachusetts Inst. of Tech., USA rbm@math.mit.edu	Dan Virgil Voiculescu	University of California, Berkeley, USA dvv@math.berkeley.edu
Frank Merle	Université de Cergy-Pontoise, France Frank.Merle@u-cergy.fr	Steven Zelditch	Northwestern University, USA zelditch@math.northwestern.edu
William Minicozzi II	Johns Hopkins University, USA minicozz@math.jhu.edu	Maciej Zworski	University of California, Berkeley, USA zworski@math.berkeley.edu

## PRODUCTION

production@msp.org  
Silvio Levy, Scientific Editor

---

See inside back cover or [msp.org/apde](http://msp.org/apde) for submission instructions.

---

The subscription price for 2018 is US \$275/year for the electronic version, and \$480/year (+\$55, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscriber address should be sent to MSP.


---

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

---

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

PUBLISHED BY

 **mathematical sciences publishers**  
nonprofit scientific publishing

<http://msp.org/>

© 2018 Mathematical Sciences Publishers

# ANALYSIS & PDE

Volume 11 No. 8 2018

---

Invariant measure and long time behavior of regular solutions of the Benjamin–Ono equation	1841
MOUHAMADOU SY	
Rigidity of minimizers in nonlocal phase transitions	1881
OVIDIU SAVIN	
Propagation and recovery of singularities in the inverse conductivity problem	1901
ALLAN GREENLEAF, MATTI LASSAS, MATTEO SANTACESARIA, SAMULI SILTANEN and GUNTHER UHLMANN	
Quantitative stochastic homogenization and regularity theory of parabolic equations	1945
SCOTT ARMSTRONG, ALEXANDRE BORDAS and JEAN-CHRISTOPHE MOURRAT	
Hopf potentials for the Schrödinger operator	2015
LUIGI ORSINA and AUGUSTO C. PONCE	
Monotonicity of nonpluripolar products and complex Monge–Ampère equations with pre- scribed singularity	2049
TAMÁS DARVAS, ELEONORA DI NEZZA and CHINH H. LU	
On weak weighted estimates of the martingale transform and a dyadic shift	2089
FEDOR NAZAROV, ALEXANDER REZNIKOV, VASILY VASYUNIN and ALEXANDER VOLBERG	
Two-microlocal regularity of quasimodes on the torus	2111
FABRICIO MACIÀ and GABRIEL RIVIÈRE	
Spectral distribution of the free Jacobi process, revisited	2137
TAREK HAMDI	



2157-5045(2018)11:8;1-M