MESOSCOPIC HIGHER REGULARITY AND SUBADDITIVITY IN ELLIPTIC HOMOGENIZATION

SCOTT ARMSTRONG, TUOMO KUUSI, AND JEAN-CHRISTOPHE MOURRAT

ABSTRACT. We introduce a new method for obtaining quantitative results in stochastic homogenization for linear elliptic equations in divergence form. Unlike previous works on the topic, our method does not use concentration inequalities (such as Poincaré or logarithmic Sobolev inequalities in the probability space) and relies instead on a higher (C^k, $k \geq 1$) regularity theory for solutions of the heterogeneous equation, which is valid on length scales larger than a certain specified mesoscopic scale. This regularity theory, which is of independent interest, allows us to, in effect, localize the dependence of the solutions on the coefficients and thereby accelerate the rate of convergence of the expected energy of the cell problem by a bootstrap argument. The fluctuations of the energy are then tightly controlled using subadditivity. The convergence of the energy gives control of the scaling of the spatial averages of gradients and fluxes (that is, it quantifies the weak convergence of these quantities) which yields, by a new "multiscale" Poincaré inequality, quantitative estimates on the sublinearity of the corrector.

1. INTRODUCTION

In this paper, we introduce a new method for obtaining quantitative results in elliptic homogenization. It is based on a new regularity theory for higher derivatives of solutions which is valid on *mesoscopic* scales. The regularity estimates essentially localize the dependence of the solutions on the coefficients, allowing for the mixing assumptions to dictate the rate of homogenization. This idea is formalized by a novel bootstrap argument which uses the regularity estimates to accelerate the convergence of the natural subadditive and superadditive quantities associated to the variational formulation of the equation. As an application, we give explicit estimates on the sublinearity of the correctors and the weak convergence of their rescaled gradients.

1.1. **Motivation and informal summary of results.** We consider the linear elliptic equation

$$(1.1) \qquad\qquad -\nabla \cdot (\mathbf{a}(x)\nabla u) = 0$$

in bounded open subsets of $\mathbb{R}^d$. The coefficient $\mathbf{a}(\cdot)$ is a random field valued in the set of d-by-d symmetric matrices with eigenvalues belonging to the interval $[1, \Lambda]$ for a fixed ellipticity constant $\Lambda \geq 1$. The law of $\mathbf{a}(\cdot)$ is given

by a probability measure $\mathbb{P}$ which is assumed to satisfy a quantitative mixing condition.

We are interested in obtaining quantitative information on the statistical properties of solutions of (1.1) on very large scales, a concern which lies within the realm of homogenization. The qualitative theory of homogenization for such equations was developed by [21, 18, 23, 17] and the first quantitative results are due Yurinskii [24]. Developing a quantitative theory of stochastic homogenization for linear elliptic equations has received a lot of attention since the groundbreaking work of Gloria and Otto [13, 14, 15] and Gloria, Neukamm and Otto [10, 11], who proved an array of optimal estimates under certain very strong mixing assumptions on the coefficients (in particular, that $\mathbb{P}$ satisfies some form of a *spectral gap inequality*). At the core of their results are the use of sensitivity estimates and the spectral gap assumption to derive moment bounds on the gradient of the corrector and the Green's functions, which culminated in the work of Marahrens and Otto [19]. These gradient estimates, following Naddaf and Spencer [20], give control of the sensitivity of the correctors themselves to changes in the coefficient field, and thereby yield optimal estimates after another application of the spectral gap inequality.

A regularity theory for stochastic homogenization was recently introduced by Armstrong and Smart [4]. In particular, [4, Theorem 1.2] provides a gradient bound, of exactly the sort required for the quantitative theory of homogenization, which applies to arbitrary solutions of (1.1) and with much stronger (and essentially optimal) stochastic integrability. The techniques of [4] were further developed by Armstrong and Mourrat [2], for general (nonlinear) divergence-form equations and systems, and, in a similar spirit by Gloria, Neukamm and Otto [12] for linear equations and systems. These works were inspired by the celebrated papers of Avellaneda and Lin [5, 6], who proved in particular uniform Lipschitz estimates for equations with periodic coefficients. As in [5, 6], the philosophy is to show that solutions of the heterogeneous equation (1.1) inherit the regularity of the limiting constant-coefficient equation due to homogenization. The method of [4] is however different from these previous works, which were based on compactness arguments and used strong bounds on the correctors. Instead, the idea is to mimic more closely the proof of the classical Schauder estimates, using quantitative methods to replace compactness: one shows that, at each scale, a solution of (1.1) may be approximated by the solution of the homogenized equation and then iterates the resulting estimate over dyadic scales. If the error in the approximation is small enough (an algebraic or Dini rate of homogenization is needed), then the argument yields a uniform Lipschitz estimate. This argument is more robust than the compactness argument of [5, 6], since it separates the approximation step from the iteration step. Indeed, in addition to its applicability in the stochastic setting, it has yielded new results even in the periodic and almost periodic settings (cf. [3, 22]).

Another idea from [4, 2] is that subadditive arguments are the way to prove quantitative homogenization results that are optimal in stochastic integrability (under essentially any mixing assumption). This leads to a regularity theory which is optimal in terms stochastic integrability. Subadditive arguments are so

effective for this purpose because they reduce questions of stochastic integrability for random variables with complicated dependence on the random environment (such as solutions of the PDE) to estimates of a finite sum of bounded random variables which satisfy the same mixing properties as the coefficients. Under assumptions on the coefficient field which provide a spectral gap inequality, the regularity theory (specifically the Lipschitz estimate) is strong enough to quickly recover the optimal quantitative estimates on the corrector proved by Gloria and Otto. Thus an important consequence of these new regularity estimates on the quantitative theory was to separate the gradient estimates, which can now be proved under very general mixing assumptions, from the rest of the quantitative theory, which until now requires spectral gap-type assumptions.

In the present paper, we propose to take this program one step further by deriving quantitative bounds for the correctors (with explicit exponents) without using the spectral gap or other Poincaré-type concentration inequalities. This is the first step in obtaining a quantitative theory of stochastic homogenization applicable to general coefficient fields. Instead of relying on concentration of measure, we show by a bootstrap argument that the regularity theory itself can be used to improve the rate of homogenization. In addition to providing a new point of view in the quantitative theory of stochastic homogenization and allowing for more general mixing assumptions, our arguments yield estimates which are much stronger in stochastic integrability (under strong mixing assumptions, we get exponential moments rather than just pth moments). This is because, as in [4, 2], our method allows us to use subadditivity to control the stochastic fluctuations. So far, the estimates we can obtain (see Theorem 1.2, below) are unfortunately suboptimal in their scalings compared to what can be proved under spectral gap assumptions. This is due to the presence of boundary layers encountered in the analysis which eventually force the bootstrap to halt before desired. We hope to address this issue in the near future.

In addition to the uniform Lipschitz estimate, our approach requires higher $C^{k,1}$ estimates for $k \geq 1$, which were left essentially implicit in [4, 2]. Of course, unlike the Lipschitz estimate, such higher derivative estimates for $k \geq 1$ cannot hold uniformly on the unit scale, but they are valid on *mesoscopic* scales: the assertion, which is stated precisely in Theorem 2.1 below, is roughly that a solution of (1.1) on the ball B_R, with $R \gg 1$, can be well-appoximated by a kth degree polynomial on any mesoscopic ball $B_r(x) \subseteq B_{R/2}$ with radius $r \geq R^{1-\varepsilon}$, for a specified exponent $\varepsilon > 0$ which depends in particular on k. By "well-approximated", we mean that the quality of the approximation of the solution by the polynomial is, up to a constant, as good as one would have for a harmonic function with the same oscillation. This result is presented below in Theorem 2.1. Fischer and Otto [9] have very recently developed a higher regularity theory along somewhat different lines (see also Remark 2.3).

While we state and prove our results for scalar equations with symmetric coefficients under the assumption of finite range of dependence, we emphasize that these choices are not imposed on us by any limitation of our method. Indeed, the arguments work verbatim for systems with just changes to the notation, the techniques of [2] can be used to get rid of the symmetry assumption,

and it is precisely under weaker mixing conditions that our estimates would be optimal, since the weak mixing stops the bootstrap before it sees boundary layers. Our choice to make these additional assumptions rather reflects a desire to maintain the readability of the paper by presenting the ideas in their simplest setting.

1.2. **Assumptions.** We work in the Euclidean space $\mathbb{R}^d$ in dimension $d \geq 2$ and with a fixed ellipticity constant $\Lambda \geq 1$. We consider the space of coefficient fields $\mathbf{a}(\cdot)$ valued in the symmetric d-by-d matrices satisfying, for all $\xi \in \mathbb{R}^d$,

$$(1.2) \qquad |\xi|^2 \leq \xi \cdot \mathbf{a}(x)\xi \leq \Lambda |\xi|^2 .$$

We define Ω to be the set of all such coefficient fields:

$$\Omega := \left\{ \mathbf{a}(\cdot) : \mathbf{a} : \mathbb{R}^d \to \mathbb{R}^{d \times d} \text{ is Lebesgue measurable, satisfies } (1.2) \text{ and } \mathbf{a}^t = \mathbf{a} \right\}.$$

We endow Ω with the translation group $\{T_y\}_{y \in \mathbb{R}^d}$, which acts on Ω via

$$(T_y \mathbf{a})(x) := \mathbf{a}(x + y),$$

and the family $\{\mathcal{F}(U)\}$ of σ–algebras on Ω, with $\mathcal{F}(U)$ defined for each Borel subset $U \subseteq \mathbb{R}^d$ by

$$\mathcal{F}(U) := \sigma\text{–algebra on } \Omega \text{ generated by the family of maps}$$

$$\mathbf{a} \mapsto \int_U q \cdot \mathbf{a}(x) p\, \varphi(x)\, dx, \quad p, q \in \mathbb{R}^d, \ \varphi \in C_c^\infty(\mathbb{R}^d).$$

Roughly, $\mathcal{F}(U)$ contains the information about the coefficients restricted to U. We denote the largest of these σ–algebras by $\mathcal{F} := \mathcal{F}(\mathbb{R}^d)$. The translation group may be naturally extended to $\mathcal{F}$ itself by defining

$$T_y A := \{ T_y \mathbf{a} : \mathbf{a} \in A \}, \quad A \in \mathcal{F}$$

and to any random element X by setting $(T_z X)(\mathbf{a}) := X(T_z \mathbf{a})$.

Throughout the paper, we consider a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ which is assumed to satisfy the following two conditions:

(P1) $\mathbb{P}$ is stationary with respect to $\mathbb{Z}^d$–translations: for every $z \in \mathbb{Z}^d$ and $A \in \mathcal{F}$,

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

(P2) $\mathbb{P}$ has a unit range of dependence: for every pair of Borel subsets $U, V \subseteq \mathbb{R}^d$ with $\mathrm{dist}(U, V) \geq 1$,

$$\mathcal{F}(U) \text{ and } \mathcal{F}(V) \text{ are } \mathbb{P}\text{–independent}.$$

The expectation of an $\mathcal{F}$-measurable random variable X with respect to $\mathbb{P}$ is denoted by $\mathbb{E}[X]$.

1.3. **Notation.** We continue with some notation used throughout the paper. For a measurable set $E \subseteq \mathbb{R}^d$, we denote the Lebesgue measure of E by $|E|$ unless E is a finite set, in which case $|E|$ denotes the cardinality of E. For a 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

$$(1.3) \qquad \|f\|_{\underline{L}^p(U)} := \left(\fint_U |f(x)|^p \, dx \right)^{\frac{1}{p}}.$$

It is also convenient to denote $\|f\|_{\underline{L}^\infty(U)} := \|f\|_{L^\infty(U)}$. For a vector-valued $F \in L^p(U; \mathbb{R}^d)$, we write $\|F\|_{\underline{L}^p(U)} := \||F|\|_{\underline{L}^p(U)}$. The average of a function $f \in L^1(U)$ on U is denoted by

$$(f)_U := \fint_U f(x) \, dx.$$

For $U \subseteq \mathbb{R}^d$ with $|U| < \infty$ and vector fields $F, G \in L^2(U; \mathbb{R}^d)$, we denote

$$(1.4) \qquad \langle F, G \rangle_U := \fint_U F(x) \cdot G(x) \, dx.$$

Since we also wish to quantify the convergence of various functions in the weak L^p topology, it is natural to work with $W^{-1,p}$ norms. For this purpose we introduce the normalized $W^{-1,p}$ norm of $F \in L^p(U; \mathbb{R}^d)$ by

$$\|F\|_{\underline{W}^{-1,p}(U)} := \sup \left\{ |\langle F, \eta \rangle_U| \, : \, \eta \in W^{1,p'}(U; \mathbb{R}^d), \, (\eta)_U = 0, \, \|\nabla \eta(x)\|_{\underline{L}^{p'}(U)} = 1 \right\}.$$

As usual, p' denotes the Hölder conjugate of an exponent $p \in [1, \infty]$. We also use the shorthand notation

$$(1.5) \qquad \|F\|_{\underline{H}^{-1}(U)} := \|F\|_{\underline{W}^{-1,2}(U)}.$$

Note that $W^{-1,p'}$ and H^{-1} are (somewhat unconventionally) used to denote the duals of $W^{1,p}/\mathbb{R}$ and $H^1/\mathbb{R}$, rather than $W_0^{1,p}$ and H_0^1. We denote cubes of side length $R > 0$ by

$$\Box_R = \Box_R(0) := \left(-\frac{1}{2} R, \frac{1}{2} R \right)^d, \qquad \Box_R(x) := x + \Box_R.$$

The family of cubes of side length at least one is

$$(1.6) \qquad \mathcal{C} := \left\{ \Box_R(x) \, : \, x \in \mathbb{R}^d, \, R \geq 1 \right\}.$$

The set of (real-valued) polynomials on $\mathbb{R}^d$ with degree at most $k \in \mathbb{N}$ is denoted by $\mathcal{P}_k$. The (random) vector space of solutions of (1.1) in $U \subseteq \mathbb{R}^d$ is

$$(1.7) \quad \mathcal{A}(U) := \left\{ u \in H^1_{\mathrm{loc}}(U) \, : \, \forall v \in H^1_0(U), \, \int_U \nabla v(x) \cdot \mathbf{a}(x) \nabla u(x) \, dx = 0 \right\}.$$

Recall that, for each $p \in \mathbb{R}^d$, the *corrector* $\Phi(\cdot, p)$ is defined for $\mathbb{P}$–almost every $\mathbf{a} \in \Omega$ as the unique function (up to a constant) in $H^1_{\mathrm{loc}}(\mathbb{R}^d)$ with a $\mathbb{Z}^d$–stationary, mean-zero gradient and which satisfies the equation

$$-\nabla \cdot (\mathbf{a}(x) (p + \nabla \Phi(\cdot, p))) = 0 \quad \text{in } \mathbb{R}^d.$$

See [17] for details.

1.4. **Subadditive quantities and main results.** We now recall some objects from [4, 2] which play a central role in the paper. Given a bounded Lipschitz domain $U \subseteq \mathbb{R}^d$ and $p, q \in \mathbb{R}^d$, we denote the linear function of slope p by $\ell_p(x) := p \cdot x$ and define

$$\nu(U, p) := \inf_{v \in \ell_p + H_0^1(U)} \fint_U \frac{1}{2} \nabla v(x) \cdot \mathbf{a}(x) \nabla v(x) \, dx,$$

which is the natural subadditive quantity representing the energy of the solution of the "cell problem," introduced (in a more general form) by Dal Maso and Modica [7, 8] in their proof of qualitative homogenization of convex integral functionals. We also define

$$\mu(U, q) = \inf_{u \in H^1(U)} \fint_U \left(\frac{1}{2} \nabla u(x) \cdot \mathbf{a}(x) \nabla u(x) - q \cdot \nabla u(x) \right) dx,$$

which is the natural superadditive quantity introduced in [4] (also in a more general form), and is dual to $\nu(U, p)$.

The quantity $\nu(\cdot, p)$ is *subadditive* in the sense that if $U, U_1, \ldots, U_k$ are bounded domains satisfying

$$U_1, \ldots, U_k \quad \text{are pairwise disjoint and} \quad |U \smallsetminus (U_1 \cup \cdots \cup U_k)| = 0,$$

then

$$(1.8) \qquad \nu(U, p) \le \sum_{i=1}^k \frac{|U_i|}{|U|} \nu(U_i, p).$$

This is immediate from the fact that a candidate minimizer for $\nu(U, p)$ can be obtained by assembling the minimizers of the quantities $\nu(U_i, p)$, as these agree on the boundaries. (We remark that we are using the term "subadditive" in a nonconventional way, due to our normalization, as it would be more standard to say that $U \mapsto |U| \nu(U, p)$ is subadditive.) Conversely, the quantity $\mu(\cdot, q)$ is *superadditive* (that is, $-\mu(\cdot, q)$ is subadditive), since the minimizer for $\mu(U, q)$ provides with a candidate minimizer for each $\mu(U_i, q)$ by restriction.

Note that the minimizer for $\mu(U, q)$ will be the solution of a Neumann problem for (1.1) in U, while of course the minimizer for $\nu(U, p)$ is the solution of a Dirichlet problem. Thus we may think of $\nu(U, p)$ as "the subadditive quantity for gradients" and of $\mu(U, q)$ as "the superadditive quantity for fluxes."

The homogenized matrix $\bar{\mathbf{a}}$ can be defined as the symmetric matrix satisfying

$$(1.9) \qquad \frac{1}{2} p \cdot \bar{\mathbf{a}} p = \lim_{R \to \infty} \mathbb{E}[\nu(\square_R, p)].$$

We view the quantities μ and ν as central to the quantitative theory of homogenization. The main step in the quantitative arguments in [4, 2] is to show roughly that there exists $\alpha(d, \Lambda) > 0$ such that, for every $\square \in \mathcal{C}$,

$$\left| \nu(\square, p) - \tfrac{1}{2} p \cdot \bar{\mathbf{a}} p \right| \le C|p|^2 |\square|^{-\alpha} \quad \text{with overwhelming probability,}$$

and a similar estimate for μ,

$$\left| \mu(\square, q) + \tfrac{1}{2} q \cdot \bar{\mathbf{a}}^{-1} q \right| \le C|q|^2 |\square|^{-\alpha} \quad \text{with overwhelming probability.}$$

The first main result of this paper improves the rate of convergence from the unspecified and tiny $\alpha(d, \Lambda) > 0$ in [4] to any exponent $\alpha < \frac{1}{d}$.

Theorem 1.1 (Rate of convergence of subadditive quantities). *Let $\alpha < \frac{1}{d}$. There exists $C(d, \Lambda, \alpha) < \infty$ such that for every $p, q \in \mathbb{R}^d$, $\square \in \mathcal{C}$ and $\lambda \in \mathbb{R}$,*

$$(1.10) \quad \begin{cases} \log \mathbb{E}\left[\exp\left(\lambda|p|^{-2}|\square|^\alpha \left|\nu(\square, p) - \frac{1}{2}p \cdot \bar{\mathbf{a}}p\right|\right)\right] \le C(1 + \lambda^2), \quad and \\ \log \mathbb{E}\left[\exp\left(\lambda|q|^{-2}|\square|^\alpha \left|\mu(\square, q) + \frac{1}{2}q \cdot \bar{\mathbf{a}}^{-1}q\right|\right)\right] \le C(1 + \lambda^2). \end{cases}$$

Testing the definition of $\mu(U, q)$ with the minimizer of $\nu(U, p)$ yields

$$(1.11) \qquad\qquad J(U, p, q) := \nu(U, p) - \mu(U, q) - p \cdot q \ge 0.$$

Moreover, as we will see, $J(U, p, q)$ can be expressed in a variational form, and the optimizer is precisely the difference of the minimizers of $\mu(U, q)$ and $\nu(U, p)$. In other words, the difference of the subadditive energies is an energy of the differences of the minimizers. The proof of Theorem 1.1 is based on a bootstrap argument, using the higher regularity estimates, applied to this difference, to accelerate the convergence of $J(\square, p, \bar{\mathbf{a}}p)$ to zero. See Section 4.1 for a sketch of the bootstrap argument.

The estimate (1.10) is optimal in the sense that it is false for any $\alpha > \frac{1}{d}$. To see this, we note that minimizers for μ satisfy (oscillating) Neumann conditions while those of ν satisfy Dirichlet conditions. Therefore, the gradient of the difference of the minimizers of $\mu(\square, \bar{\mathbf{a}}p)$ and $\nu(\square, p)$ will in general be $O(1)$ in a boundary layer of at least unit thickness. The proportion of volume of such a boundary layer relative to the whole cube $\square$ is of order $|\square|^{-1/d}$. Therefore $J(\square, p, \bar{\mathbf{a}}p)$ should be at least $c|\square|^{-1/d}$ in general.

It is the presence of this boundary layer that so far limits our quantitative results to consequences of (1.10). On the other hand, one expects the minimizers of $\nu(U, p)$ and $\mu(U, \bar{\mathbf{a}}p)$ to be much closer in the interior of U than near the boundary, and so if the boundary layer could be neglected, there is hope to prove much more precise results.

Theorem 1.1 can be used to prove an array of quantitative estimates in homogenization with exponents which, under the strongest mixing assumptions, will typically differ from the optimal one by a square root. For example, one can show that the L^2 error in homogenization is $O(\varepsilon^{1/2-})$, where the microscopic length scale is ε, or that the H^1 norm of the two-scale expansion is $O(\varepsilon^{1/4-})$. The application we present here is an estimate of the sublinearity of the corrector and of the H^{-1} norm of its gradient (recall that the H^{-1} norm measures weak convergence in L^2). It roughly states that

$$\left\|\Phi(\cdot, p) - (\Phi(\cdot, p))_{B_R}\right\|_{\underline{L}^2(B_R)} + \|\nabla\Phi(\cdot, p)\|_{\underline{H}^{-1}(B_R)} \lesssim R^{\frac{1}{2}+},$$

with very strong stochastic integrability. Note that estimates on the sublinearity of the corrector are intimately connected to estimates for the error in homogenization.

Theorem 1.2 (Sublinear growth of the corrector). *Let $\beta \in \left(0, \frac{1}{2}\right)$. There exists a constant $C(d, \Lambda, \beta) < \infty$ such that, for every $p \in \mathbb{R}^d$, $R \ge 1$ and $\lambda \in \mathbb{R}$,*

$$\log \mathbb{E}\left[\exp\left(\lambda|p|^{-2}R^{-2+2\beta}\left(\left\|\Phi(\cdot, p) - (\Phi(\cdot, p))_{B_R}\right\|_{\underline{L}^2(B_R)}^2 + \|\nabla\Phi(\cdot, p)\|_{\underline{H}^{-1}(B_R)}^2\right)\right)\right]$$
$$\le C(1 + \lambda^2).$$

Theorem 1.2 is new, even though it is almost certainly suboptimal and does not compare favorably with the results of Gloria and Otto [15] under spectral gap assumptions, who proved that correctors are almost bounded (see also Gloria, Neukamm and Otto [12] for more results under spectral gap assumptions). Indeed, it seems to be an open question (and there is some doubt) whether finite range of dependence implies a spectral gap-type inequality in $d > 1$. The best previous result for finite range of dependence was for $\beta(d, \Lambda) > 0$ very small (cf. [4]). Moreover, as mentioned above, our methods are applicable under essentially any mixing condition and we believe yield essentially optimal estimates under weaker mixing conditions (such as mixing conditions so slow that spatial averages of the coefficients on scale R converge slower than R^{-1} to their mean). This will be explained in more details in future papers.

The novelty of Theorem 1.2 is not however in its statement, but in its proof.

1.5. **Outline of the paper.** Our results rely crucially on $C^{k,1}$ estimates for solutions of (1.1) on *mesoscopic* scales. This was essentially proved in [4], where the higher regularity theory in stochastic homogenization was introduced; the precise statements we require are proved in Section 2. In Section 3, we make the fundamental observation that the quantity J in (1.11) can be expressed as a "modulated energy" of the difference of the minimizers of μ and ν.

We break the proof of Theorem 1.1 into two main steps. First, in Section 4, we use an induction argument on the exponent α to show that, for every $\alpha < \frac{1}{d}$, $p \in \mathbb{R}^d$ and cube $\square \in \mathcal{C}$,

$$(1.12) \qquad \left| \mathbb{E}[\nu(\square, p)] - \frac{1}{2} p \cdot \bar{\mathbf{a}} p \right| \le C |p|^2 |\square|^{-\alpha},$$

and a similar estimate with μ in place of ν. This statement is *a priori* much weaker than the conclusion of Theorem 1.1, but we show in Section 5 that the subadditivity of $-\mu$ and ν can be used to upgrade the stochastic integrability from L^1 to exponential moments without sacrificing any of the exponent.

Finally, in Section 6, we present a functional inequality, which appears to be new and which we call *multiscale Poincaré inequality.* It has a wavelet flavor and converts control over the spatial averages of the gradient of a function in triadic subcubes into an estimate on the oscillation of the function itself. This, together with the regularity theory, reduces Theorem 1.2 to Theorem 1.1.

2. Higher regularity on mesoscopic scales

A cornerstone of the methods in this paper is a quenched $C^{k,1}$ estimate for solutions of (1.1) on *mesoscopic* scales. This was essentially proved in [4], where the higher regularity theory in stochastic homogenization was introduced.

The rough statement, which is given precisely in the following theorem, asserts that a solution of (1.1) can be approximated, just as well as a harmonic function, by kth degree harmonic polynomials – on length scales larger than a fixed *mesoscopic* scale (or the microscopic scale, if $k = 0$). We let $\overline{\mathcal{A}}_k$ denote the set of polynomials of degree at most k which are $\bar{\mathbf{a}}$-harmonic.

Theorem 2.1 ($C^{k,1}$ regularity on mesoscopic scales)**.** *Fix $s \in (0, d)$ and $k \in \mathbb{N}$. There exist an exponent $\delta(s, d, \Lambda) > 0$, a constant $C(s, k, d, \Lambda) < \infty$ and an $\mathcal{F}$-measurable random variable $\mathcal{X} : \Omega \to [1, \infty)$ satisfying the estimate*

$$(2.1) \qquad \qquad \mathbb{E}\left[\exp\left(\mathcal{X}^s\right)\right] < \infty$$

and such that, for every $R \geq 2\mathcal{X}$, $v \in \mathcal{A}(B_R)$ and $r \in \left[\mathcal{X} \vee R^{\frac{k}{k+\delta}}, \frac{1}{2}R\right]$,

$$(2.2) \qquad \qquad \inf_{w \in \overline{\mathcal{A}}_k} \|v - w\|_{\underline{L}^2(B_r)} \leq C\left(\frac{r}{R}\right)^{k+1} \|v\|_{\underline{L}^2(B_R)}.$$

Theorem 2.1 is essentially proved in [4]. That paper only stated the (uniform) $C^{0,1}$ estimate and left (mesoscopic) higher regularity statements for $k \geq 1$ implicit in the proof. To prove the proposition, we just need to expound the argument from [4].

A second ingredient in the proof of Theorem 2.1, which also plays a key role in the bootstrap argument, is the following quenched error estimate for the Dirichlet problems proved in [4]. It gives a quenched, deterministic estimate for the error in homogenization for the Dirichlet problem on length scales larger than a certain random scale of characteristic size $O(1)$.

Proposition 2.2 ([4, Theorem 1.1])**.** *Fix $s \in (0, d)$, $\varepsilon > 0$, and let $U \subseteq B_1$ be a Lipschitz domain. There exists an exponent $\delta(d, \Lambda, s, \varepsilon) > 0$, a constant $C(s, d, \Lambda, U) < \infty$ and an $\mathcal{F}$-measurable random variable $\mathcal{R} : \Omega \to [1, \infty)$, which depends only on $(d, \Lambda, s, \varepsilon)$ and satisfies the estimate*

$$(2.3) \qquad \qquad \mathbb{E}\left[\exp\left(\mathcal{R}^s\right)\right] < \infty,$$

such that, for every $r \geq \mathcal{R}$, $f \in W^{1,2+\varepsilon}(rU)$ and solutions $u, \overline{u} \in f + H_0^1(rU)$ of

$$-\nabla \cdot (\mathbf{a}\nabla u) = 0 \quad and \quad -\nabla \cdot (\overline{\mathbf{a}}\nabla \overline{u}) = 0 \quad in \ rU,$$

we have the estimate

$$(2.4) \qquad \qquad \frac{1}{r}\|u - \overline{u}\|_{\underline{L}^2(rU)} \leq r^{-\delta} \|\nabla f\|_{\underline{L}^{2+\varepsilon}(rU)}.$$

Remark 2.3. The mesoscopic regularity estimate in Theorem 2.1 implies Liouville theorems of all orders. These state roughly that, for each $k \in \mathbb{N}$, the subspace of $\mathcal{A}(\mathbb{R}^d)$ consisting of functions which grow at most like $o\left(|x|^{k+1}\right)$ has the same dimension as $\overline{\mathcal{A}}_k$. The latter functions correspond precisely to the *kth order correctors.* Note that, in view of the results of [2], Theorem 2.1 and therefore these Liouville theorems hold in much greater generality than we present here (e.g., for nonlinear equations and with weaker mixing conditions). A similar result was recently proved by Fischer and Otto [9], who developed a similar higher regularity theory in the general stationary ergodic setting. Similar to [11], they proceed a bit differently: rather than measure the distance of an element of $\mathcal{A}(B_R)$ to $\overline{\mathcal{A}}_k$, they measure its distance to the kth order correctors. We believe that while both approaches are of interest, the one presented here is more faithful to what a $C^{k,1}$ estimate should be in this context (we would like, for example, to measure the $C^{k,1}$ seminorm of a jth order corrector, for $j \leq k$, and not get zero, otherwise we gain no information on the correctors themselves). See below Remark 4.5 in [3] for some similar comments.

That a quantitative estimate for the error in homogenization for the Dirichlet problem implies higher regularity estimates via a Campanato iteration was an idea introduced in [4]. Here we formulate a version in the following lemma, which is a variation of [4, Lemma 5.1].

Lemma 2.4. *Fix $R \geq 2$, $\alpha > 0$, $p \in [1, \infty]$ and $u \in L^p(B_R)$. For each $k \in \mathbb{N}$ and $s \in (0, R]$, denote*

$$D_k(s) := \inf_{w \in \overline{\mathcal{A}}_k} \|u - w\|_{\underline{L}^p(B_s)}$$

Assume that $h \in \left[1, \frac{1}{2}R\right]$ and have the property that, for every $r \in \left[h, \frac{1}{2}R\right]$, there exists an $\bar{\mathbf{a}}$-harmonic function $v \in C^\infty(B_r)$ such that

$$(2.5) \qquad\qquad \|u - v\|_{\underline{L}^p(B_r)} \leq r^{-\alpha} D_0(2r).$$

Then, for each $k \in \mathbb{N}$, there exists a constant $C(d, \Lambda, k, \alpha) < \infty$ such that, for every $r \in \left[h, \frac{1}{2}R\right]$,

$$(2.6) \qquad\qquad D_k(r) \leq C\left(\frac{r}{R}\right)^{k+1} D_k(R) + Cr^{-\alpha}\left(\frac{r}{R}\right) D_0(R).$$

Proof. Fix $k \in \mathbb{N}$. Throughout, we denote by C and c positive constants depending only on (k, α, Λ) which may vary in each occurrence. In the first few steps, we make some preliminary observations and introduce the notation we need in the main part of the argument, which begins in Step 3.

Step 1. We first observe that the triangle inequality and the hypothesis (2.5) imply that, for every $r \in \left[h, \frac{1}{2}R\right]$ and $s \in \left(0, \frac{1}{2}r\right]$,

$$(2.7) \qquad\qquad D_k(s) \leq C\left(\frac{s}{r}\right)^{k+1} D_k(r) + C\left(\frac{r}{s}\right)^{\frac{d}{p}} r^{-\alpha} D_0(r).$$

Selecting a harmonic function $v \in C^\infty(B_r)$ to satisfy (2.5), we find that

$$\begin{aligned}
D_k(s) &= \inf_{w \in \overline{\mathcal{A}}_k} \|u - w\|_{\underline{L}^p(B_s)} \\
&\leq \inf_{w \in \overline{\mathcal{A}}_k} \|v - w\|_{\underline{L}^p(B_s)} + \|u - v\|_{\underline{L}^p(B_s)} \\
&\leq C\left(\frac{s}{r}\right)^{k+1} \inf_{w \in \overline{\mathcal{A}}_k} \|v - w\|_{\underline{L}^p(B_r)} + Cr^{-\alpha} D_0(r)\left(\frac{r}{s}\right)^{\frac{d}{p}}.
\end{aligned}$$

In the last line, we used the fact that any $\bar{\mathbf{a}}$-harmonic function v satisfies, for every $0 < s \leq \frac{1}{2}r$,

$$\inf_{w \in \overline{\mathcal{A}}_k} \|v - w\|_{L^\infty(B_s)} \leq C\left(\frac{s}{r}\right)^{k+1} \inf_{w \in \overline{\mathcal{A}}_k} \|v - w\|_{\underline{L}^p(B_r)}.$$

Next we use the triangle inequality and (2.5) a second time to get

$$\inf_{w \in \overline{\mathcal{A}}_k} \|v - w\|_{\underline{L}^p(B_r)} \leq D_k(r) + Cr^{-\alpha} D_0(r).$$

Substituting into the inequality above, we get

$$D_k(s) \le C\left(\frac{s}{r}\right)^{k+1}\left(D_k(r) + r^{-\alpha}D_0(r)\right) + C\left(\frac{r}{s}\right)^{\frac{d}{p}} r^{-\alpha}D_0(r)$$

$$\le C\left(\frac{s}{r}\right)^{k+1} D_k(r) + C\left(\frac{r}{s}\right)^{\frac{d}{p}} r^{-\alpha}D_0(r).$$

Step 2. We set up the rest of the argument. Fix $\theta = \theta(k) \ge c$ so small that $C\theta = \frac{1}{4}$, where C is the constant in (2.7), so that the latter inequality implies

$$\widetilde{D}_k(\theta r) \le \frac{1}{2}\widetilde{D}_k(r) + Cr^{-k-\alpha}D_0(r),$$

where here and in what follows we set $\widetilde{D}_k(s) := s^{-k}D_k(s)$. An iteration of the previous inequality gives

$$(2.8) \qquad \widetilde{D}_k(\theta^m r) \le 2^{-m}\widetilde{D}_k(r) + C\sum_{j=0}^{m-1} 2^{j-m}\left(\theta^j r\right)^{-k-\alpha} D_0\left(\theta^j r\right)$$

provided $\theta^{m-1}r \ge h$. To shorten the notation, we denote $r_j := \theta^j R$, $B^j := B_{r_j}$ and by $w_{k,j}$ the best kth degree polynomial approximation of u in B^j, that is, $w_{k,j}$ satisfies

$$\|u - w_{k,j}\|_{\underline{L}^p(B^j)} = \inf_{w\in\mathcal{P}_k} \|u - w\|_{\underline{L}^p(B^j)}$$

Step 3. We now complete the proof of the proposition in the case $k = 0$. In fact, this has been already proved in [4, Lemma 5.1], but we give the argument for the sake of completeness.

First, since we may add constants to both u and v, we may assume without loss of generality that $w_{0,0} = 0$. It follows that

$$\|w_{1,0}\|_{\underline{L}^p(B_R)} \le 2D_0(R),$$

and hence we easily get

$$(2.9) \qquad \|w_{1,0}\|_{L^\infty(B_R)} + R\|\nabla w_{1,0}\|_{L^\infty(B_R)} \le CD_0(R).$$

Using (2.8) we obtain

$$\frac{1}{r_{m+1}}\|w_{1,m+1} - w_{1,m}\|_{\underline{L}^p(B^{m+1})} \le \widetilde{D}_1(r_{m+1}) + \theta^{-1-\frac{d}{p}}\widetilde{D}_1(r_m)$$

$$\le 2^{-m}C\widetilde{D}_1(R) + C\sum_{j=0}^{m} 2^{j-m} r_j^{-1-\alpha} D_0\left(r_j\right).$$

We thus deduce that

$$\left|\nabla w_{1,m+1} - \nabla w_{1,m}\right| \le 2^{-m}C\widetilde{D}_1(R) + C\sum_{j=0}^{m} 2^{j-m} r_j^{-\alpha}\frac{D_0\left(r_j\right)}{r_j},$$

and it follows by summation that

$$\left|\nabla w_{1,n} - \nabla w_{1,0}\right| \le C\widetilde{D}_1(R) + C\sum_{j=0}^{n} r_j^{-\alpha}\frac{D_0\left(r_j\right)}{r_j}$$

provided that $r_n \geq h$. Combining this with (2.8) and (2.9) leads to

$$\frac{D_0(r_n)}{r_n} \leq \widetilde{D}_1(r_n) + |\nabla w_{1,n}| \leq C\frac{D_0(R)}{R} + C\sum_{j=0}^{n-1} r_j^{-\alpha}\frac{D_0(r_j)}{r_j}.$$

Taking supremum then gives

$$\sup_{m\in\{0,\dots,n\}} \frac{D_0(r_m)}{r_m} \leq C\frac{D_0(R)}{R} + C\sum_{j=0}^{n} r_j^{-\alpha} \sup_{m\in\{0,\dots,n\}} \frac{D_0(r_m)}{r_m}$$

provided that $r_n \geq h$. Letting now $n^* = n^*(d,\alpha)$ be the largest integer such that $C\sum_{j=0}^{n^*} r_j^{-\alpha} \leq \frac{1}{2}$ and $r_{n^*} \geq h$ we obtain after straightforward manipulations that

$$(2.10) \qquad\qquad \sup_{h\leq r\leq R} \frac{D_0(r)}{r} \leq C\frac{D_0(R)}{R}.$$

Step 4. We complete the argument in the case of general $k \in \mathbb{N}$. Following the reasoning of the previous step, letting $\widetilde{w}_{k+1,j}$ stand for the kth order polynomial part of $w_{k+1,j}$, we get, by (2.8) and (2.10),

$$(2.11) \qquad \widetilde{D}_k(r_j) \leq \frac{1}{r_j^k}\|u - \widetilde{w}_{k+1,j}\|_{\underline{L}^p(B^j)}$$

$$\leq r_j\widetilde{D}_{k+1}(r_j) + \frac{1}{r_j^k}\|w_{k+1,j} - \widetilde{w}_{k+1,j}\|_{\underline{L}^p(B^j)}$$

$$\leq r_j\left(2^{-j}\widetilde{D}_{k+1}(R) + Cr_j^{-k-\alpha}\frac{D_0(R)}{R} + |\nabla^{k+1}w_{k+1,j}|\right).$$

Thus we are left to estimate $|\nabla^{k+1}w_{k+1,j}|$. Since

$$\frac{1}{r_{j+1}^{k+1}}\|w_{k+1,j+1} - w_{k+1,j+1}\|_{\underline{L}^p(B^{j+1})} \leq C\left(\widetilde{D}_{k+1}(r_{j+1}) + \widetilde{D}_{k+1}(r_j)\right)$$

$$\leq C2^{-j}\widetilde{D}_{k+1}(R) + Cr_j^{-k-\alpha}\frac{D_0(R)}{R},$$

we deduce that

$$(2.12) \qquad \left|\nabla^{k+1}w_{k+1,j+1} - \nabla^{k+1}w_{k+1,j}\right| \leq C2^{-j}\widetilde{D}_{k+1}(R) + Cr_j^{-k-\alpha}\frac{D_0(R)}{R}.$$

Indeed, for any $n \in \mathbb{N}$ and polynomial $\phi \in \mathcal{P}_n$, we have that

$$(2.13) \qquad \left(\fint_{B_r} |\phi(x)|^p\,dx\right)^{\frac{1}{p}} = \left(\fint_{B_1} |\phi(rx)|^p\,dx\right)^{\frac{1}{p}} \geq c\sup_{x\in B_1}|\phi(rx)| \geq cr^n|\nabla^n\phi|$$

for a constant $c = c(d,n) > 0$. This is due to scaling and Lemma 2.6, which is stated and proved below.

Now (2.12) implies after summation that

$$\left|\nabla^{k+1}w_{k+1,j} - \nabla^{k+1}w_{k+1,0}\right| \leq C\widetilde{D}_{k+1}(R) + Cr_j^{-k-\alpha}\frac{D_0(R)}{R}.$$

Furthemore, since

$$\|w_{k,0} - w_{k+1,0}\|_{\underline{L}^p(B_R)} \leq D_{k+1}(R) + D_k(R) \leq 2D_k(R),$$

we obtain

$$\left|\nabla^{k+1}w_{k+1,0}\right| \leq CR^{-k-1}D_k(R),$$

and consequently

$$\left|\nabla^{k+1}w_{k+1,j}\right| \le \left|\nabla^{k+1}w_{k+1,j} - \nabla^{k+1}w_{k+1,0}\right| + \left|\nabla^{k+1}w_{k+1,0}\right|$$
$$\le CR^{-k-1}D_k(R) + Cr_j^{-k-\alpha}\frac{D_0(R)}{R}.$$

Combining this finally with (2.11) allows us to conclude with

$$D_k(r_j) \le C\left(\frac{r_j}{R}\right)^{k+1} D_k(R) + Cr_j^{-\alpha}\left(\frac{r_j}{R}\right)\frac{D_0(R)}{R},$$

from which the statement (2.6) can be easily deduced. □

Remark 2.5. Notice that since $D_k(R) \le D_0(R)$, the estimate (2.6) implies

$$D_k(r) \le C\left(\frac{r}{R}\right)^{k+1} D_0(R) + Cr^{-\delta}\left(\frac{r}{R}\right)D_0(R).$$

This is a $C^{k,1}$ estimate on scales r for which the second term on the right side is smaller than the first term, that is, for r satisfying

$$r \gtrsim R^{\frac{k}{k+\delta}}.$$

In the proof of Lemma 2.4, we used the following fact, which is a simple consequence of the equivalence of norms in finite-dimensional vector spaces.

Lemma 2.6. *Fix $k \in \mathbb{N}$. There exists a constant $C(d,k) < \infty$ such that, for every $m \le k$ and $w \in \mathcal{P}_k$,*

$$\|\nabla^m w\|_{L^\infty(B_1)} \le C \|w\|_{L^1(B_1)}.$$

Proof of Theorem 2.1. The proof is essentially the same as – indeed, even considerably simpler than (due to homogeneity) – the argument given in the general nonlinear case considered in [4, 2]. If we take $U = B_1$ and $\varepsilon(d,\Lambda) > 0$ to be the exponent in the interior Meyers estimate (cf. [2, Proposition B.6] for instance), then the solvability of the Dirichlet problem and the conclusion of Proposition 2.2, with $\mathcal{R}$ given there, ensures that the hypothesis of Lemma 2.4 is satisfied for $h = \mathcal{R}$. The lemma then yields the result for $\mathcal{X} = \mathcal{R}$. □

3. Properties of the modulated energy J

We begin this section with the simple but key observation that the difference between ν and μ can be expressed as a "modulated energy" of the difference of the minimizers. We define $\mathcal{J}(\cdot, U, p, q)$ for functions $w \in H^1(U)$ by

$$\mathcal{J}(w,U,p,q) := \fint_U \left(-\frac{1}{2}\nabla w(x) \cdot \mathbf{a}(x)\nabla w(x) - p \cdot \mathbf{a}(x)\nabla w(x) + q \cdot \nabla w(x)\right) dx.$$

We denote the maximum of $\mathcal{J}(\cdot, U, p, q)$ among solutions of the PDE by

$$J(U,p,q) := \max_{w \in \mathcal{A}(U)} \mathcal{J}(w,U,p,q).$$

To motivate the definition of $J(U,p,q)$, we show that it is actually a familiar object: it can be decomposed into the quantities studied in [4], which are the focus of the analysis in that paper.

Lemma 3.1. *For every bounded Lipschitz domain $U \subseteq \mathbb{R}^d$ and $p, q \in \mathbb{R}^d$,*

$$(3.1) \qquad\qquad J(U, p, q) = \nu(U, p) - \mu(U, q) - p \cdot q.$$

Proof. Fix $p, q \in \mathbb{R}^d$, $u \in \mathcal{A}(U)$ and let $v \in \ell_p + H_0^1(U)$ be the minimizer in the definition of $\nu(U, p)$. Note that $v \in \mathcal{A}(U)$ and compute

$$\nu(U, p) - \fint_U \frac{1}{2} \left(\nabla u(x) \cdot \mathbf{a}(x) \nabla u(x) - q \cdot \nabla u(x) \right) dx - p \cdot q$$

$$= \fint_U \left(\frac{1}{2} \nabla v(x) \cdot \mathbf{a}(x) \nabla v(x) - \frac{1}{2} \nabla u(x) \cdot \mathbf{a}(x) \nabla u(x) + q \cdot \nabla u(x) \right) dx - p \cdot q$$

$$= \fint_U \Bigg(-\frac{1}{2} \left(\nabla u(x) - \nabla v(x) \right) \cdot \mathbf{a}(x) \left(\nabla u(x) - \nabla v(x) \right)$$

$$\qquad\qquad - p \cdot \mathbf{a}(x) \left(\nabla u(x) - \nabla v(x) \right) + q \cdot \left(\nabla u(x) - \nabla v(x) \right) \Bigg) dx$$

$$= \mathcal{J}(u - v, U, p, q).$$

Here we used integration by parts twice, both taking advantage of the affine boundary condition for v to get

$$p = \fint_U \nabla v(x) \, dx$$

and another, which also uses that $u - v \in \mathcal{A}(U)$ to get

$$\fint_U p \cdot \mathbf{a}(x) \left(\nabla u(x) - \nabla v(x) \right) dx = \fint_U \nabla v(x) \cdot \mathbf{a}(x) \left(\nabla u(x) - \nabla v(x) \right) dx.$$

We deduce that

$$J(U, p, q) = \max_{w \in \mathcal{A}(U)} \mathcal{J}(w, U, p, q)$$

$$= \max_{u \in \mathcal{A}(U)} \mathcal{J}(u - v, U, p, q)$$

$$= \max_{u \in \mathcal{A}(U)} \left(\nu(U, p) - \fint_U \frac{1}{2} \left(\nabla u(x) \cdot \mathbf{a}(x) \nabla u(x) - q \cdot \nabla u(x) \right) dx - p \cdot q \right)$$

$$= \nu(U, p) - \mu(U, q) - p \cdot q. \qquad\qquad\qquad \square$$

Notice that the proof of Lemma 3.1 gave more than its statement: namely, the maximizer in the definition of $J(U, p, q)$ is precisely the difference of the minimizers of $\mu(U, q)$ and $\nu(U, p)$.

In the rest of this section, we present some basic properties of J which are needed in the bootstrap argument in the next section, and we fix a bounded Lipschitz domain $U \subseteq \mathbb{R}^d$ throughout. We denote the unique (up to additive constants) maximizer in the definition of $J(U, p, q)$ by

$$u(\cdot, U, p, q) := \text{maximizer of } \mathcal{J}(\cdot, U, p, q) \text{ among functions in } \mathcal{A}(U).$$

Notice that $u(\cdot, U, 0, q)$ is the minimizer for $\mu(U, q)$ and $-u(\cdot, U, p, 0)$ is the minimizer for $\nu(U, p)$.

We next record the first and second variations of the optimization problem implicit in the definition of $J(U, p, q)$. Henceforth, we make use of the notation (1.4).

Lemma 3.2. *For every* $p, q \in \mathbb{R}^d$ *and* $v \in \mathcal{A}(U)$,

$$(3.2) \qquad \langle \mathbf{a}\nabla u(\cdot, U, p, q), \nabla v \rangle_U = \langle -\mathbf{a}p + q, \nabla v \rangle_U$$

and

$$(3.3) \qquad J(U, p, q) - \mathcal{J}(u(\cdot, U, p, q) + v, U, p, q) = \frac{1}{2} \langle \nabla v, \mathbf{a}\nabla v \rangle_U .$$

Proof. For convenience, for $t \geq 0$, set $u_t := u(\cdot, U, p, q) + tv$. Compute

$$0 \leq \mathcal{J}(u_0, U, p, q) - \mathcal{J}(u_t, U, p, q)$$
$$= t^2 \fint_U \left(\frac{1}{2} \nabla v(x) \cdot \mathbf{a}(x) \nabla v(x) \right) dx$$
$$+ t \fint_U (\mathbf{a}(x)\nabla u_0(x) + \mathbf{a}(x)p - q) \cdot \nabla v(x) \, dx.$$

Dividing by t and sending $t \to 0$ gives

$$\fint_U (\mathbf{a}(x)\nabla u_0(x) + \mathbf{a}(x)p - q) \cdot \nabla v(x) \, dx \geq 0.$$

Repeating the argument with $-v$ in place of v yields (3.2). Returning to the previous identity and taking $t = 1$ gives (3.3). $\qquad\square$

For reference, we observe that by taking $v = u(\cdot, U, p, q)$ in (3.2) we obtain the identities

$$(3.4) \qquad J(U, p, q) = \frac{1}{2} \langle \mathbf{a}\nabla u(\cdot, U, p, q), \nabla u(\cdot, U, p, q) \rangle_U$$
$$= \frac{1}{2} \langle -\mathbf{a}p + q, \nabla u(\cdot, U, p, q) \rangle_U .$$

It is easy to see from (3.2) that $(p, q) \mapsto u(\cdot, U, p, q)$ is a linear map from $\mathbb{R}^d \times \mathbb{R}^d$ into $\mathcal{A}(U)$; that is, for every $p_1, p_2, q_1, q_2 \in \mathbb{R}^d$ and $s, t \in \mathbb{R}$,

$$(3.5) \qquad u(\cdot, U, tp_1 + sp_2, tq_1 + sq_2) = tu(\cdot, U, p_1, q_1) + su(\cdot, U, p_2, q_2)$$

Likewise, J is quadratic: for every $p, q \in \mathbb{R}^d$ and $t > 0$,

$$(3.6) \qquad J(U, tp, tq) = t^2 J(U, p, q).$$

We next show that $\mathcal{J}(\cdot, U, p, q)$ responds quadratically to perturbations near its maximum.

Lemma 3.3. *For every* $p, q \in \mathbb{R}^d$ *and* $v, w \in \mathcal{A}(U)$,

$$(3.7) \qquad \frac{1}{4} \|\nabla v - \nabla w\|_{\underline{L}^2(U)}^2 \leq 2J(U, p, q) - \mathcal{J}(w, U, p, q) - \mathcal{J}(v, U, p, q)$$

and

$$(3.8) \qquad 2\mathcal{J}(v, U, p, q) - \mathcal{J}(w, U, p, q) - J(U, p, q) \leq \frac{\Lambda}{4} \|\nabla v - \nabla w\|_{\underline{L}^2(U)}^2 .$$

Proof. For any $v_1, v_2 \in \mathcal{A}(U)$,

$$2\mathcal{J}\left(\frac{v_1 + v_2}{2}, U, p, q \right) - \mathcal{J}(v_1, U, p, q) - \mathcal{J}(v_2, U, p, q)$$
$$= \frac{1}{4} \langle \mathbf{a}(\nabla v_1 - \nabla v_2), \nabla v_1 - \nabla v_2 \rangle_U .$$

Then (3.7) follows by choosing $v_1 = v$ and $v_2 = w$ and using the maximality of $J(U,p,q)$, and (3.8) follows similarly by choosing $v_1 = w$ and $v_2 = 2v - w$. $\quad\square$

Lemma 3.4. *For every $p, q, p', q' \in \mathbb{R}^d$,*

$$(3.9) \quad J(U, p + p', q + q') - J(U, p, q) - J(U, p', q') = \langle -\mathbf{a}p' + q', \nabla u(\cdot, U, p, q) \rangle_U.$$

Proof. By the first variation (3.2),

$$\langle \mathbf{a}\nabla u(\cdot, U, p, q), \nabla u(\cdot, U, p + p', q + q') \rangle_U = \langle -\mathbf{a}(p + p') + q + q', \nabla u(\cdot, U, p, q) \rangle_U$$
$$= 2J(U, p, q) + \langle -\mathbf{a}p' + q', \nabla u(\cdot, U, p, q) \rangle_U$$

and, by (3.5),

$$\langle \mathbf{a}\nabla u(\cdot, U, p, q), \nabla u(\cdot, U, p + p', q + q') \rangle_U$$
$$= \langle -\mathbf{a}p + q, \nabla u(\cdot, U, p + p', q + q') \rangle_U$$
$$= 2J(U, p + p', q + q') - \langle -\mathbf{a}p' + q', \nabla u(\cdot, U, p + p', q + q') \rangle_U$$
$$= 2J(U, p + p', q + q') - 2J(U, p', q') - \langle -\mathbf{a}p' + q', \nabla u(\cdot, U, p, q) \rangle_U.$$

Combining the two displays above gives (3.9). $\quad\square$

We denote by $\nabla J(U, p, q)$ the gradient of $J(U, \cdot)$ at (p, q). By (3.6), this is a linear mapping from $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ which, in view of the previous lemma, can be expressed by

$$(3.10) \qquad \nabla J(U, p, q)(p', q') = \langle -\mathbf{a}p' + q', \nabla u(\cdot, U, p, q) \rangle_U.$$

This relation between $\nabla J(U, p, q)$ and the spatial averages of the gradient and flux of $u(\cdot, U, p, q)$ will play an important role in the proof of Proposition 4.1. For simplicity, the maps $p' \mapsto \nabla J(U, p, q)(p', 0)$ and $q' \mapsto \nabla J(U, p, q)(0, q')$ are sometimes denoted by $\nabla_p J(U, p, q)$ and $\nabla_q J(U, p, q)$, respectively.

Lemma 3.4 gives the identity

$$(3.11) \quad \frac{1}{2}J(U, p_1, q_1) + \frac{1}{2}J(U, p_2, q_2) - J\left(U, \frac{p_1 + p_2}{2}, \frac{q_1 + q_2}{2}\right)$$
$$= \frac{1}{4}J(U, p_1 - p_2, q_1 - q_2).$$

This readily implies the following upper convexity estimate for J in (p, q) and the lower convexity estimates for J in the variables p and q separately.

Lemma 3.5. *For every $p, p_1, p_2, q, q_1, q_2 \in \mathbb{R}^d$,*

$$(3.12) \quad \frac{1}{2}J(U, p_1, q_1) + \frac{1}{2}J(U, p_2, q_2) - J\left(U, \frac{1}{2}(p_1 + p_2), \frac{1}{2}(q_1 + q_2)\right)$$
$$\leq \Lambda\left(|p_1 - p_2|^2 + |q_1 - q_2|^2\right),$$

$$(3.13) \qquad \frac{1}{2}J(U, p_1, q) + \frac{1}{2}J(U, p_2, q) - J\left(U, \frac{1}{2}(p_1 + p_2), q\right) \geq \frac{1}{2}|p_1 - p_2|^2$$

and

$$(3.14) \qquad \frac{1}{2}J(U, p, q_1) + \frac{1}{2}J(U, p, q_2) - J\left(U, p, \frac{1}{2}(q_1 + q_2)\right) \geq \frac{1}{2\Lambda}|q_1 - q_2|^2.$$

Proof. We apply (3.11) together with (3.4) to obtain an upper bound for $J(U, p_1 - p_2, q_1 - q_2)$, and rearrange to obtain (3.12). We obtain (3.13) and (3.14) in a similar way, using the bounds

$$J(U, p, 0) = \nu(U, p) \geq \frac{1}{2}|p|^2 \quad \text{and} \quad J(U, 0, q) = -\mu(U, q) \geq \frac{1}{2\Lambda}|q|^2. \qquad \square$$

By subadditivity, stationarity and (1.9), for every $\square \in \mathcal{C}$, we have

$$(3.15) \qquad\qquad \frac{1}{2}p \cdot \bar{\mathbf{a}}p \leq \mathbb{E}[\nu(\square, p)].$$

By (1.11), this implies

$$(3.16) \qquad\qquad \frac{1}{2}q \cdot \bar{\mathbf{a}}^{-1}q \leq -\mathbb{E}[\mu(\square, q)],$$

so that

$$(3.17) \qquad\qquad \mathbb{E}[J(\square, p, q)] \geq \frac{1}{2}p \cdot \bar{\mathbf{a}}p + \frac{1}{2}q \cdot \bar{\mathbf{a}}^{-1}q - p \cdot q.$$

Due to Lemma 3.5, there exists a unique, deterministic matrix $Q = Q(U)$ such that, for every $p \in \mathbb{R}^d$,

$$p = \nabla_q \mathbb{E}\left[J(U, 0, Qp)\right] = \mathbb{E}\left[\nabla_q J(U, 0, Qp)\right].$$

Indeed, for each fixed $p \in \mathbb{R}^d$, we can consider the minimum of the uniformly convex, quadratic function

$$q \mapsto \mathbb{E}\left[J(U, 0, q)\right] - p \cdot q.$$

This defines a linear map from p to the minimum point $q(p)$. The matrix Q is defined by $q = Qp$. It is easy to check that Q is symmetric, and the upper and lower uniform convexity of $J(U, 0, \cdot)$ ensures that Q is positive and in particular invertible. In fact, we have

$$(3.18) \qquad\qquad I_d \leq Q \leq \Lambda I_d.$$

Notice that, for every $\square \in \mathcal{C}$, we have

$$(3.19) \qquad\qquad Q(\square) \leq \bar{\mathbf{a}}.$$

Indeed, by (3.17),

$$\begin{aligned}
\frac{1}{2}p \cdot Q(\square)p &= \frac{1}{2}Q(\square)p \cdot \mathbb{E}\left[\nabla_q J(U, 0, Q(\square)p)\right] \\
&= \mathbb{E}\left[J(U, 0, Q(\square)p)\right] \\
&\geq \frac{1}{2}Q(\square)p \cdot \bar{\mathbf{a}}^{-1}Q(\square)p.
\end{aligned}$$

Putting $p := Q(\square)^{-1}q$ gives $Q(\square)^{-1} \geq \bar{\mathbf{a}}^{-1}$, which is equivalent to the claim. For future reference, we notice that since $u(\cdot, U, p, Qp) = u(\cdot, U, p, 0) + u(\cdot, U, 0, Qp)$ and $-u(\cdot, U, p, 0)$ is the minimizer of $\nu(U, p)$, we have that $-u(\cdot, U, p, 0) \in \ell_{-p} + H_0^1(U)$ and

$$\fint_U \nabla u(x, U, p, 0)\, dx = -p$$

and thus by (3.10) that

$$\nabla_q \mathbb{E}\left[J(U, p, Qp)\right] = \nabla_q \mathbb{E}\left[J(U, 0, Qp)\right] + \nabla_q \mathbb{E}\left[J(U, p, 0)\right] = p - p = 0.$$

Similarly, we denote by $P = P(U)$ the deterministic, symmetric matrix defined via the relation

$$q = \nabla_p \mathbb{E}\left[J(U, Pq, 0)\right].$$

As above, we have the estimates

$$\frac{1}{\Lambda} I_d \leq P(U) \leq I_d,$$

for every $\square \in \mathcal{C}$,

$$P(\square) \leq \bar{\mathbf{a}}^{-1}$$

and

(3.20) $$\nabla_q \mathbb{E}\left[J(U, Pq, q)\right] = 0.$$

We next give an estimate for the difference between $Q(\square)$ and the homogenized coefficients $\bar{\mathbf{a}}$ in terms of the expected size of $J(\square, p, Q(\square)p)$.

Lemma 3.6. *There exists $C(d, \Lambda) < \infty$ such that, for any $\square \in \mathcal{C}$,*

$$\left|\bar{\mathbf{a}} - Q(\square)\right| + \left|\bar{\mathbf{a}}^{-1} - P(\square)\right| \leq C \sup_{p \in \mathbb{R}^d} |p|^{-2} \mathbb{E}\left[J(\square, p, Q(\square)p)\right].$$

Proof. We prove only the estimate for $|\bar{\mathbf{a}} - Q(\square)|$. The argument for the bound on $|\bar{\mathbf{a}}^{-1} - P(\square)|$ is similar. We drop the dependence on $\square$ and write $Q = Q(\square)$ and $J(p, q) = J(\square, p, q)$. Set

$$\eta := \sup_{p \in \mathbb{R}^d} |p|^{-2} \mathbb{E}\left[J(p, Qp)\right].$$

By (3.17), we have, for every $p \in \mathbb{R}^d$,

$$0 \leq \frac{1}{2} p \cdot \bar{\mathbf{a}} p + \frac{1}{2} Qp \cdot \bar{\mathbf{a}}^{-1} Qp - p \cdot Qp \leq \mathbb{E}\left[J(p, Qp)\right] \leq \eta |p|^2.$$

It follows by uniform convexity that, for every $p \in \mathbb{R}^d$,

$$\left|Qp - \bar{\mathbf{a}}p\right|^2 \leq C\eta |p|^2$$

and thus, by upper uniform convexity of $q \mapsto \mathbb{E}\left[J(p, q)\right]$ and the fact that this map achieves its minimum at $q = Qp$, we deduce

$$\mathbb{E}\left[J(p, \bar{\mathbf{a}}p)\right] \leq C\eta |p|^2.$$

Using (3.1), (3.15) and (3.16), we get

$$p \cdot \bar{\mathbf{a}}p \leq \mathbb{E}\left[J(p, 0)\right] + \mathbb{E}\left[J(0, \bar{\mathbf{a}}p)\right] \leq p \cdot \bar{\mathbf{a}}p + C\eta |p|^2.$$

In view of (3.15) and (3.16), this implies that, for every $p, q \in \mathbb{R}^d$,

$$\left|\mathbb{E}\left[J(p, 0)\right] - \frac{1}{2} p \cdot \bar{\mathbf{a}}p\right| \leq C\eta |p|^2 \quad \text{and} \quad \left|\mathbb{E}\left[J(0, q)\right] - \frac{1}{2} q \cdot \bar{\mathbf{a}}^{-1} q\right| \leq C\eta |q|^2.$$

From this, the definition of Q and the fact that J is quadratic, we get

$$\frac{1}{2} p \cdot Qp = \frac{1}{2} Qp \cdot \mathbb{E}\left[\nabla_q J(0, Qp)\right] = \mathbb{E}\left[J(0, Qp)\right] \leq \frac{1}{2} Qp \cdot \bar{\mathbf{a}}^{-1} Qp + C\eta |p|^2.$$

Thus $Q\bar{\mathbf{a}}^{-1}Q \geq Q - C\eta I_d$, that is, $Q(\bar{\mathbf{a}}^{-1} - Q^{-1})Q \geq -C\eta I_d$. By (3.18), this implies $Q \geq \bar{\mathbf{a}} - C\eta I_d$. In view of (3.19), the proof is now complete. $\square$

4. The bootstrap argument

This section is devoted to the proof of the following estimate for J.

Proposition 4.1. *For every* $\alpha \in \left(0, \frac{1}{d}\right)$, *there exist* $C(d, \Lambda, \alpha) < \infty$ *such that, for every* $p \in \mathbb{R}^d$ *and* $\square \in \mathcal{C}$,

$$(4.1) \qquad \mathbb{E}\left[J\left(\square, p, \bar{\mathbf{a}}p\right)\right] \le C|p|^2 |\square|^{-\alpha}.$$

An immediate consequence of Proposition 4.1 is

Corollary 4.2. *For every* $\alpha \in \left(0, \frac{1}{d}\right)$, *there exist* $C(d, \Lambda, \alpha) < \infty$ *such that, for every* $p, q \in \mathbb{R}^d$ *and* $\square \in \mathcal{C}$,

$$(4.2) \qquad \begin{cases} \left| \mathbb{E}\left[\nu(\square, p)\right] - \dfrac{1}{2}p \cdot \bar{\mathbf{a}}p \right| \le C|p|^2 |\square|^{-\alpha}, \qquad and \\[2mm] \left| \mathbb{E}\left[\mu(\square, q)\right] + \dfrac{1}{2}q \cdot \bar{\mathbf{a}}^{-1}q \right| \le C|q|^2 |\square|^{-\alpha}. \end{cases}$$

Proof of Corollary 4.2. By (3.1) and Proposition 4.1,

$$(4.3) \qquad \mathbb{E}\left[\nu(\square, p)\right] - \mathbb{E}\left[\mu(\square, \bar{\mathbf{a}}p)\right] - p \cdot \bar{\mathbf{a}}p \le C|p|^2 |\square|^{-\alpha}.$$

Combining this with (3.15) and (3.16) yields the result. $\qquad \square$

We begin the proof of Proposition 4.1 with some reductions. First, observe that it suffices to demonstrate (4.1) for cubes of the form $\square_R$, $R \ge 1$, since we can then apply the result to translations of the law $\mathbb{P}$ (recall that we only assume $\mathbb{Z}^d$-stationarity). Second, in view of Lemma 3.6, in order to prove (4.1), it suffices to show that, for each $\alpha < \frac{1}{d}$, there exists $C(d, \Lambda, \alpha) < \infty$ such that, for every $R \ge 1$ and $p \in \mathbb{R}^d$,

$$(4.4) \qquad \mathbb{E}\left[J(\square_R, p, Q(\square_R)p)\right] \le C|p|^2 R^{-d\alpha}.$$

The proof of (4.4) is by induction. For each $\alpha \in (0, 1)$ and $K \ge 1$, we let $\mathcal{S}(\alpha, K)$ be the assertion that, for every $R \ge 1$ and $p \in \mathbb{R}^d$,

$$\mathbb{E}\left[J(\square_R, p, Q(\square_R)p)\right] \le K|p|^2 R^{-d\alpha}.$$

In view of Lemma 3.1, the base case of our bootstrap was proved in [4].

Proposition 4.3 ([4, Theorem 3.1]). *There exists* $\alpha_0(d, \Lambda) > 0$ *and* $K_0(d, \Lambda) < \infty$ *such that*

$$(4.5) \qquad \mathcal{S}(\alpha_0, K_0) \quad holds.$$

The previous proposition was proved in [4, Section 3] by showing that minimizers of $\mu(\square_n, q)$ for large n are expected to be flat– that is, close to a deterministic affine function– and this allows comparison to $\nu(\square_n, p)$ for an appropriate p (the slope of the affine function). Note that the result was proved with $\bar{\mathbf{a}}$ in place of $Q(\square_R)$, but since $Q(\square_R)$ is the minimum of the map $q \mapsto \mathbb{E}\left[J(\square_R, p, q)\right]$, we can make this replacement.

In view of Proposition 4.3, it therefore suffices to show that there exist $\varepsilon(d, \Lambda) > 0$ and $C(d, \Lambda) < \infty$ such that, for every $\alpha \in \left[\alpha_0, \frac{1}{d}\right)$ and $K \ge 1$,

$$(4.6) \qquad \mathcal{S}(\alpha, K) \implies \mathcal{S}\left(\alpha + \varepsilon(1 - d\alpha), CK^{\frac{3}{2}}\right).$$

Indeed, an iteration of (4.6), starting from $\mathcal{S}(\alpha_0, K_0)$, yields (4.4).

The rest of this section is focused on the proof of (4.6). Throughout, we fix $\alpha \in \left[\alpha_0, \frac{1}{d}\right)$ and $K \geq 1$, and assume that $\mathcal{S}(\alpha, K)$ holds. We fix $R \geq 1$ and $p \in \mathbb{R}^d$. By homogeneity, we may assume that $|p| \leq 1$. We set

$$(4.7) \qquad q := Q(\square_R)p,$$

where we recall that $Q(\square_R)$ is defined before Lemma 3.6. Throughout,

$$v := u\left(\cdot, \square_R, p, q\right)$$

denotes the maximizer of $J(\square_R, p, q)$. Notice that Lemma 3.3 gives

$$(4.8) \qquad \|\nabla v\|^2_{\underline{L}^2(\square_R)} \leq 4J\left(\square_R, p, q\right).$$

In particular, the induction hypothesis yields

$$(4.9) \qquad \mathbb{E}\left[\|\nabla v\|^2_{\underline{L}^2(\square_R)}\right] \leq 4KR^{-d\alpha}.$$

4.1. **Rough sketch of the argument for** (4.6). Before giving the complete details, we present an informal summary of the proof of (4.6). Choose a mesoscopic scale $r \in (1, R)$ and partition the cube $\square_R$ into smaller cubes of the form $\square_r(y)$, $y \in \square_R$. Observe that, using (3.2) in the last step,

$$\begin{aligned} J(\square_R, p, q) &= \left(\frac{R}{r}\right)^{-d} \sum_y \mathcal{J}(v, \square_r(y), p, q) \\ &\leq \left(\frac{R}{r}\right)^{-d} \sum_y \langle -\mathbf{a}p + q, \nabla v \rangle_{\square_r(y)} \\ &= \left(\frac{R}{r}\right)^{-d} \sum_y \langle \mathbf{a}\nabla \widetilde{u}_y, \nabla v \rangle_{\square_r(y)}, \end{aligned}$$

where $\widetilde{u}_y := u(\square_r(y), p, q)$. If the induction hypothesis $\mathcal{S}(\alpha, K)$ holds, then, using also (4.9) and the Lipschitz estimate, we have

$$(4.10) \qquad \mathbb{E}[\|\nabla \widetilde{u}_y\|^2_{\underline{L}^2(\square_r(y))}] \leq Cr^{-d\alpha} \quad \text{and} \quad \mathbb{E}[\|\nabla v\|^2_{\underline{L}^2(\square_r(y))}] \leq CR^{-d\alpha}.$$

No improvement on the estimate of $\mathbb{E}[J(\square_R, p, q)]$ can be obtained from these observations alone, as we have not yet used the mixing properties of the coefficients.

If v is sufficiently close (in say L^2) to an affine function of slope p_y in the mesoscopic cube $\square_r(y)$, then, since $\widetilde{u}_y$ is a solution of (1.1), we may integrate by parts to get

$$\langle \mathbf{a}\nabla \widetilde{u}_y, \nabla v \rangle_{\square_r(y)} = \langle \mathbf{a}\nabla \widetilde{u}_y, p_y \rangle_{\square_r(y)} + \text{a small error}.$$

Recall that the dual vector q was chosen in (4.7) so that

$$\mathbb{E}\left[\fint_{\square_R} \nabla v(x)\, dx\right] = 0, \quad \text{that is,} \quad \sum_y \mathbb{E}\left[p_y\right] = 0.$$

Hence, the sum

$$(4.11) \qquad \left(\frac{R}{r}\right)^{-d} \sum_y \mathbb{E}[\langle \mathbf{a}\nabla \widetilde{u}_y, p_y \rangle_{\square_r(y)}] = \left(\frac{R}{r}\right)^{-d} \sum_y \mathbb{E}\left[p_y \cdot \fint_{\square_y(r)} \mathbf{a}\nabla \widetilde{u}_y(x)\, dx\right]$$

can be seen as a measure of the correlations between the (discrete) random fields $\left(\fint_{\square_r(y)} \mathbf{a}\nabla\widetilde{u}_y\right)$ and (p_y).

The main insight is that the higher regularity theory guarantees that the slopes p_y exist, that the approximation to v has a sufficiently small error, and that p_y *changes very slowly in y.* Meanwhile, the spatial averages of the fluxes $\left(\fint_{\square_r(y)} \mathbf{a}\nabla\widetilde{u}_y\right)$ are *essentially independent as y varies.* If we denote by s a mesoscale with $r \ll s \ll R$ on which the vectors p_y are approximately constant, then we can expect to bound (4.11) by

$$(4.12) \qquad R^{-d\alpha/2} r^{-d\alpha/2} \left(\frac{s}{r}\right)^{-d/2}.$$

Indeed, by (4.10), $R^{-d\alpha/2}$ is the expected size of each p_y and $r^{-d\alpha/2}$ is the expected size of the spatially averaged flux in a mesocube. The factor of $(s/r)^{-d/2}$ comes from the CLT scaling, as $(s/r)^d$ is the number of smaller mesocubes of size r in each larger mesocube of size s.

The expression in (4.12) can be made smaller than $R^{-d\alpha}$ by choosing the mesoscales r and s appropriately, provided that $\alpha < 1$. This suggests that the correct estimate for $\mathbb{E}\left[J(\square_R, p, q)\right]$ should be (up to possible logarithmic factors) R^{-d}. The argument we present below does not perform so well, and saturates at R^{-1} due to a boundary layer we have neglected in this sketch.

4.2. **Declaration of parameters and mesoscales.** We now proceed with the rigorous argument for (4.6). We take mesoscopic scales $r, s, l \geq 1$ such that

$$(4.13) \qquad \frac{R}{l}, \frac{l}{3s}, \frac{s}{r} \in \mathbb{N}$$

and

$$(4.14) \qquad r = CR^{1-(m+2)\varepsilon}, \quad s = CR^{1-2\varepsilon}, \qquad l = CR^{1-\varepsilon},$$

for fixed parameters $\varepsilon \in \left(0, \frac{1}{4m}\right]$ and $m \geq \max\{d, 2\}$ to be selected below. Here the constants C in (4.14) are very close to 1 and are chosen so that the constraints (4.13) are satisfied.

The largest mesoscale l denotes the thickness of a boundary layer we remove from $\square_R$ in the first part of the argument; it also serves as a reference scale for the regularity estimates we will apply at the smaller scale s, where we apply Theorem 2.1 to obtain polynomial approximations to v. Finally, r denotes the size of the smallest mesoscale cells in which we compare v to local solutions (denoted by v_z below) chosen to match the polynomial approximations made on the larger scale. Each of these mesoscales, even the smallest, will be chosen to be very close to the macroscale R.

The degree of the polynomials used in the mesoscopic approximation will be $k \in \mathbb{N}$, also to be chosen. We let the exponent $\delta = \delta(d, \Lambda) > 0$ be the minimum of the exponent given in the statement of Theorem 2.1 for the choice $s = 1$ and the one given in the statement of Proposition 2.2 for the choices $s = 1$, $U = \square_{1/\sqrt{d}}$ and $\varepsilon = 1$. We may assume $\delta \leq 1$. We let $\mathcal{X}$ denote the maximum of the random variables $\mathcal{X}$ and $\mathcal{R}$ from Theorem 2.1 and Proposition 2.2, respectively, with the same choices of parameters.

In order to apply Theorem 2.1 from scale l to scale s, we must have that

$$(4.15) \qquad s \geq l^{\frac{k}{k+\delta}}.$$

For this, it suffices to impose the restriction

$$(4.16) \qquad \varepsilon \leq \frac{\delta}{2(k+1)}.$$

In the course of the proof, we will also find it convenient to fix the parameters m, k, ε as follows:

$$(4.17) \qquad m := \frac{2(1+\alpha)}{1-\alpha} \vee \frac{2d(1+\alpha)}{2-d\alpha} \vee d$$

and

$$(4.18) \qquad k := \left\lceil \frac{10d(1+\theta)m}{\theta} \right\rceil \quad \text{and} \quad \varepsilon := \frac{\alpha}{2(m+3)} \wedge \frac{\delta}{2(k+1)} \wedge \frac{\delta\theta}{20d(1+\theta)m}.$$

Here the positive parameter $\theta(d,\Lambda)$ is the exponent related to the local and global Meyers' estimates, which in particular give us that, for every $\psi \in W^{1,2(1+\theta)}(\square_r(y))$, $w \in \mathcal{A}(\square_{2r}(y))$ and $\widetilde{w} \in (\psi + H_0^1(\square_r(y))) \cap \mathcal{A}(\square_r(y))$,

$$(4.19) \qquad \|\nabla w\|_{\underline{L}^{2(1+\theta)}(\square_r(y))} \leq C \|\nabla w\|_{\underline{L}^2(\square_{2r}(y))}$$

and

$$(4.20) \qquad \|\nabla \widetilde{w}\|_{\underline{L}^{2(1+\theta)}(\square_r(y))} \leq C \|\nabla \psi\|_{\underline{L}^{2(1+\theta)}(\square_r(y))}.$$

Notice that, since $\alpha \in \left[\alpha_0, \frac{1}{d}\right)$, we have $\varepsilon \geq c(d, \Lambda) > 0$. In fact, each of k, m, and ε are bounded above and below by constants depending only on (d, Λ). This implies that each of the constants C and c in the estimates below will depend only on (d, Λ) instead of $(d, \Lambda, k, m, \varepsilon)$.

The improvement in the exponent α will be the result of a CLT scaling arising due to the fact that the v_z's inside each s sized cell are independent of each other. This is seen in the proof in Step 3, below. In the final step of the proof, we will define each of the mesoscales and other parameters in such a way that the error terms encountered in the argument will be at most $CKR^{-d\alpha-\varepsilon(1-d\alpha)}$.

4.3. **Removal of boundary layer.** The improvement in the exponent α is based on the application of regularity estimates to v. Since we do not have a boundary condition for v, these estimates are inapplicable near the boundary of the macroscopic cube $\square_R$. Therefore we must remove a boundary layer, which is accomplished very simply by subadditivity. It is the error we make here that forces our bootstrap argument to halt before $d\alpha = 1$.

As R/l is an integer, the cubes $\{\square_l(y)\}_{y \in l\mathbb{Z}^d \cap \square_R}$ form a partition of $\square_R$ (up to a set of Lebesgue measure zero). We let $\square_R^\circ$ denote the cube obtained from $\square_R$ after removing a mesoscopic boundary layer of thickness l:

$$\square_R^\circ := \square_R \smallsetminus \bigcup \left\{ \square_l(y) : y \in l\mathbb{Z}^d, \partial\square_l \cap \partial\square_R \neq \varnothing \right\}.$$

Notice that

$$\frac{|\square_R \smallsetminus \square_R^\circ|}{|\square_R|} \leq C \left(\frac{R}{l}\right)^{-1}.$$

Then by subadditivity, stationarity, Lemma 3.6 and the induction hypothesis, we obtain

$$(4.21) \qquad \mathbb{E}\left[J(\square_R, p, q)\right]$$

$$\leq \frac{|\square_R^\circ|}{|\square_R|}\mathbb{E}\left[J(\square_R^\circ, p, q)\right] + \frac{|\square_R \smallsetminus \square_R^\circ|}{|\square_R|}\mathbb{E}\left[J(\square_R \smallsetminus \square_R^\circ, p, q)\right]$$

$$\leq \frac{|\square_R^\circ|}{|\square_R|}\mathbb{E}\left[J(\square_R^\circ, p, q)\right] + CK\left(\frac{R}{l}\right)^{-1} l^{-d\alpha}.$$

In the second inequality in the display above, we used the fact that $\square_R \smallsetminus \square_R^\circ$ is a union of cubes of the form $y + \square_l$; the induction hypothesis and Lemma 3.6 imply that

$$(4.22) \qquad \mathbb{E}\left[|Q(\square_R) - Q(y + \square_l)|\right] \leq CKl^{-d\alpha},$$

so that another use of the induction hypothesis gives the desired estimate.

Throughout the rest of the argument we denote

$$\mathcal{Z} := r\mathbb{Z}^d \cap \square_R^\circ.$$

4.4. Local mesoscopic approximations of v.

The next step in the argument is to introduce local approximations of v in each mesoscopic cube $\square_{3r}(y)$, with $y \in \mathcal{Z}$. These are also solutions of (1.1), that is, members of $\mathcal{A}(\square_{3r}(y))$. The advantage will be that the approximations, which we denote by v_y, serve to localize the dependence on the environment and allow us to exploit the independence assumption.

We first introduce polynomial approximations of v in the larger mesoscopic cubes of the form $\square_{3s}(z)$, $z \in s\mathbb{Z}^d \cap \square_R^\circ$. With $k \in \mathbb{N}$ to be defined below, we select $w_z \in \overline{\mathcal{A}}_k$ such that

$$(4.23) \qquad \|v - w_z\|_{\underline{L}^2(\square_{3s}(z))} = \inf_{w \in \overline{\mathcal{A}}_k} \|v - w\|_{\underline{L}^2(\square_{3s}(z))}.$$

That is, w_z is the best approximation to v in $L^2(\square_{3s}(z))$ among $\bar{\mathbf{a}}$-harmonic polynomials of degree at most k.

We next present some basic estimates concerning the expected size of w_z and the quality of the approximation to u. This is where we use the higher regularity estimates in Theorem 2.1.

Lemma 4.4. *Assume that* (4.16) *holds. Then there exists* $C(d, \Lambda, k) < \infty$ *such that, for each* $z \in s\mathbb{Z}^d \cap \square_R^\circ$,

$$(4.24) \qquad \mathbb{E}\left[\|v - w_z\|_{\underline{L}^2(\square_{3s}(z))}^2\right] \leq Cl^2\left(\frac{s}{l}\right)^{2k+2}\mathbb{E}\left[\|\nabla v\|_{\underline{L}^2(\square_l(z))}^2\right] + CR^{-2d-1},$$

$$(4.25) \qquad \mathbb{E}\left[\|\nabla w_z\|_{L^\infty(\square_{3s}(z))}^2\right] \leq C\mathbb{E}\left[\|\nabla v\|_{\underline{L}^2(\square_l(z))}^2\right] + CR^{-2d-1},$$

and, for every $m \in \{2, \ldots, k\}$,

$$(4.26) \qquad \mathbb{E}\left[s^{2(m-2)}\|\nabla^m w_z\|_{L^\infty(\square_{3s}(z))}^2\right] \leq Cl^{-2}\mathbb{E}\left[\|\nabla v\|_{\underline{L}^2(\square_l(z))}^2\right] + Cs^{-4}R^{-2d-1}.$$

Proof. We break the argument into three steps.

Step 1. The proof of (4.25). Since w_z is an $\bar{\mathbf{a}}$-harmonic polynomial, we have

$$(4.27) \qquad \|\nabla w_z\|_{L^\infty(\square_{3s}(z))} \leq \frac{C}{s} \left\| w_z - (v)_{\square_{3s}(z)} \right\|_{\underline{L}^2(\square_{3s}(z))}$$

$$\leq \frac{C}{s} \left\| v - (v)_{\square_{3s}(z)} \right\|_{\underline{L}^2(\square_{3s}(z))}$$

$$\leq C \left\| \nabla v \right\|_{\underline{L}^2(\square_{3s}(z))},$$

where, in the above display, we used Lemma 2.6 in the first line, the optimality of w_z in (4.23) tested against the constant function $(v)_{\square_{3s}(z)}$ and the triangle inequality to get the second line, and finally the Poincaré inequality to get the last line. Squaring and taking expectations, we get

$$\mathbb{E}\left[\|\nabla w_z\|_{L^\infty(\square_{3s}(z))}^2\right]$$
$$\leq C\mathbb{E}\left[\|\nabla v\|_{\underline{L}^2(\square_{3s}(z))}^2 \mathbb{1}_{\{T_z\mathcal{X}\leq Cs\}}\right] + C\mathbb{E}\left[\|\nabla v\|_{\underline{L}^2(\square_{3s}(z))}^2 \mathbb{1}_{\{T_z\mathcal{X}>Cs\}}\right]$$
$$\leq C\mathbb{E}\left[\|\nabla v\|_{\underline{L}^2(\square_l(z))}^2\right] + C\mathbb{E}\left[\|\nabla v\|_{\underline{L}^2(\square_{3s}(z))}^2 \mathbb{1}_{\{T_z\mathcal{X}>Cs\}}\right].$$

To bound the second term on the right side, we use (4.8) and the (deterministic) estimate $J(\square_R, p, q) \leq C$ to very crudely bound

$$\|\nabla v\|_{\underline{L}^2(\square_{3s}(z))}^2 \leq Cs^{-d}R^d \leq CR^d$$

and then combine this with strong integrability of $\mathcal{X}$, which give the following estimate:

$$(4.28) \quad \mathbb{E}\left[\mathbb{1}_{\{T_z\mathcal{X}>s\}} \|\nabla v\|_{\underline{L}^2(\square_{3s}(z))}^2\right]$$
$$\leq CR^{d+2}\mathbb{P}\left[\mathcal{X}>s\right] \leq CR^{d+2}\exp\left(-cs\right) \leq CR^{-2d-1}.$$

In the last line we used $s \geq R^{\frac{1}{2}}$ from (4.14). This completes the proof of (4.25).

Step 2. The proof of (4.24). According to Theorem 2.1 (recall that the assumption (4.16) implies (4.15)), we have

$$\|v - w_z\|_{\underline{L}^2(\square_{3s}(z))} \mathbb{1}_{\{T_z\mathcal{X}\leq Cs\}} \leq C\left(\frac{s}{l}\right)^{k+1} l \|\nabla v\|_{\underline{L}^2(\square_l(z))} \mathbb{1}_{\{T_z\mathcal{X}\leq Cs\}}.$$

Recall that $\{T_y\}_{y\in\mathbb{R}^d}$ is the translation group acting on the probability space Ω. In the event that $T_z\mathcal{X}$ is too large compared to s, we proceed differently. Applying (4.23) and the Poincaré inequality, we get

$$\|v - w_z\|_{\underline{L}^2(\square_{3s}(z))} \mathbb{1}_{\{T_z\mathcal{X}>Cs\}} \leq \left\| v - (v)_{\square_l(z)} \right\|_{\underline{L}^2(\square_{3s}(z))} \mathbb{1}_{\{T_z\mathcal{X}>Cs\}}$$

$$\leq C\left(\frac{l}{s}\right)^{\frac{d}{2}} \left\| v - (v)_{\square_l(z)} \right\|_{\underline{L}^2(\square_l(z))} \mathbb{1}_{\{T_z\mathcal{X}>Cs\}}$$

$$\leq Cl\left(\frac{l}{s}\right)^{\frac{d}{2}} \|\nabla v\|_{\underline{L}^2(\square_l(z))} \mathbb{1}_{\{T_z\mathcal{X}>Cs\}}.$$

Combining the previous two displays yields

$$\|v - w_z\|_{\underline{L}^2(\square_{3s}(z))}$$
$$\leq C\left(\frac{s}{l}\right)^{k+1} l \|\nabla v\|_{\underline{L}^2(\square_l(z))} \mathbb{1}_{\{T_z \mathcal{X} \leq Cs\}} + Cl\left(\frac{l}{s}\right)^{\frac{d}{2}} \|\nabla v\|_{\underline{L}^2(\square_l(z))} \mathbb{1}_{\{T_z \mathcal{X} > Cs\}}.$$

Squaring and taking expectations, we obtain

$$(4.29) \quad \mathbb{E}\left[\|v - w_z\|_{L^\infty(\square_{3s}(z))}^2\right] \leq C\left(\frac{s}{l}\right)^{2k+2} l^2 \mathbb{E}\left[\|\nabla v\|_{\underline{L}^2(\square_l(z))}^2\right]$$
$$+ Cl^2\left(\frac{l}{s}\right)^d \mathbb{E}\left[\mathbb{1}_{\{T_z \mathcal{X} > s\}} \|\nabla v\|_{\underline{L}^2(\square_l(z))}^2\right].$$

To estimate the second term on the right side, we proceed in the same way as the end of Step 1, above, to get

$$l^2\left(\frac{l}{s}\right)^d \mathbb{E}\left[\mathbb{1}_{\{T_{[y]} \mathcal{X} > s\}} \|\nabla v\|_{\underline{L}^2(\square_l(z))}^2\right] \leq CR^{-2d-1}.$$

This completes the proof of (4.24).

Step 3. The proof of (4.26). We take $\ell_z \in \overline{\mathcal{A}}_1$ to be the best affine approximation to v:

$$\|v - \ell_z\|_{\underline{L}^2(\square_{3s}(z))} = \inf_{\ell \in \overline{\mathcal{A}}_1} \|v - \ell\|_{\underline{L}^2(\square_{3s}(z))}.$$

Using the triangle inequality and (4.24) twice, once with k as above and once with $k = 1$, we get

$$\mathbb{E}\left[\|w_z - \ell_z\|_{\underline{L}^2(\square_{3s}(z))}^2\right] \leq Cl^2\left(\frac{s}{l}\right)^4 \mathbb{E}\left[\|\nabla v\|_{\underline{L}^2(\square_l(z))}^2\right] + CR^{-2d-1}.$$

Since $w_z - \ell_z$ is an $\bar{\mathbf{a}}$-harmonic polynomial, this yields, by Lemma 2.6,

$$\mathbb{E}\left[\sup_{m \in \{2,\ldots,k\}} s^{2(m-2)} \|\nabla^m w_z\|_{L^\infty(\square_{3s}(z))}^2\right] \leq C\mathbb{E}\left[s^{-4} \|w_z - \ell_z\|_{L^\infty(\square_{3s}(z))}^2\right]$$
$$\leq Cl^{-2}\mathbb{E}\left[\|\nabla v\|_{\underline{L}^2(\square_l(z))}^2\right] + Cs^{-4}R^{-2d-1},$$

which is (4.26). □

We next introduce the local mesoscopic approximations to v, which are based on w_z. For each $y \in \mathcal{Z}$, we denote by $[y]$ the unique element of $s\mathbb{Z}^d \cap \square_R^\circ$ such that $y \in \square_s([y])$. For each $y \in \mathcal{Z}$, we let v_y denote the solution of the Dirichlet problem in $\square_r(y)$ with boundary condition $w_{[y]}$. That is, v_y is the unique element of $\mathcal{A}(\square_r(y)) \cap \left(w_{[y]} + H_0^1(\square_r(y))\right)$. Note that, for every $\phi \in \mathcal{A}(\square_r(y))$,

$$(4.30) \qquad \left\langle \mathbf{a}\left(\nabla v_y - \nabla w_{[y]}\right), \nabla \phi \right\rangle_{\square_r(y)} = 0.$$

We next give the estimate for the expected difference between the gradients of v and v_y, using the previous lemma.

Lemma 4.5. *Assume that* (4.16) *and* (4.18) *hold. Then there exists a constant* $C(d, \Lambda, k) < \infty$ *such that*

$$(4.31) \qquad |\mathcal{Z}|^{-1} \sum_{y \in \mathcal{Z}} \mathbb{E}\left[\|\nabla v - \nabla v_y\|_{\underline{L}^2(\square_r(y))}^2\right] \leq CKR^{-d\alpha - 4dm\varepsilon}.$$

Proof. Set $\eta_y := v - v_y$ and let

$$(4.32) \qquad t := r - r^{1-\widetilde{\varepsilon}}, \qquad \widetilde{\varepsilon} := \frac{6dm\varepsilon(1+\theta)}{\theta},$$

where $\theta(d, \Lambda)$ is as in Meyers' estimates (4.19) and (4.20). Observe that the choice of ε in (4.18) guarantees that $\widetilde{\varepsilon} \le \frac{1}{2}$. We first split each of the summands into two pieces as follows:

$$(4.33) \qquad \int_{\square_r(y)} |\nabla\eta_y(x)|^2 \, dx = \int_{\square_t(y)} |\nabla\eta_y(x)|^2 \, dx + \int_{y+\square_r \smallsetminus \square_t} |\nabla\eta_y(x)|^2 \, dx.$$

The first term we will estimate with the aid of the Caccioppoli estimate and Lemma 4.4, and the second one using the Meyers' estimates. Indeed, we have by the Caccioppoli estimate that

$$\frac{1}{|\square_r|} \int_{\square_t(y)} |\nabla\eta_y(x)|^2 \, dx \le \frac{C}{(r-t)^2} \fint_{\square_r(y)} |\eta_y(x)|^2 \, dx.$$

To estimate the term on the right, we have by the triangle inequality that

$$\fint_{\square_r(y)} |\eta_y(x)|^2 \, dx \le 2 \left\| v - w_{[y]} \right\|^2_{\underline{L}^2(\square_r(y))} + 2 \left\| v_y - w_{[y]} \right\|^2_{\underline{L}^2(\square_r(y))}.$$

Now, for the first term we have

$$|\mathcal{Z}|^{-1} \sum_{y \in \mathcal{Z}} \left\| v - w_{[y]} \right\|^2_{\underline{L}^2(\square_r(y))} = |\mathcal{Z}|^{-1} \sum_{z \in s\mathbb{Z}^d \cap \square_R^\circ} \left(\frac{s}{r}\right)^d \left\| v - w_z \right\|^2_{\underline{L}^2(\square_s(z))}$$

$$\le C\left(\frac{s}{R}\right)^d \sum_{z \in s\mathbb{Z}^d \cap \square_R^\circ} \left\| v - w_z \right\|^2_{\underline{L}^2(\square_s(z))},$$

and thus we obtain by (4.24) that

$$|\mathcal{Z}|^{-1} \sum_{y \in \mathcal{Z}} \left\| v - w_{[y]} \right\|^2_{\underline{L}^2(\square_r(y))}$$

$$\le C\left(\frac{s}{R}\right)^d \sum_{z \in s\mathbb{Z}^d \cap \square_R^\circ} \left(l^2 \left(\frac{s}{l}\right)^{2k+2} \mathbb{E}\left[\|\nabla v\|^2_{\underline{L}^2(\square_l(z))}\right] + CR^{-2d-1} \right)$$

$$\le C\left(l^2 \left(\frac{s}{l}\right)^{2k+2} \mathbb{E}\left[\|\nabla v\|^2_{\underline{L}^2(\square_R)}\right] + R^{-2d-1} \right).$$

On the other hand, using Proposition 2.2 and (4.25) we get

$$\mathbb{E}\left[\fint_{\square_r(y)} |v_y(x) - w_{[y]}(x)|^2 \, dx \right] \le Cr^{2-2\delta} \mathbb{E}\left[\left\| \nabla w_{[y]} \right\|^2_{L^\infty(\square_r(y))} \right]$$

$$\le Cr^{2-2\delta} \left(\mathbb{E}\left[\|\nabla v\|^2_{\underline{L}^2(\square_l(y))} \right] + CR^{-2d-1} \right).$$

Summing this over $\mathcal{Z}$ implies

$$|\mathcal{Z}|^{-1} \sum_{y \in \mathcal{Z}} \mathbb{E}\left[\fint_{\square_r(y)} |v_y(x) - w_{[y]}(x)|^2 \, dx \right] \le Cr^{2-2\delta} \left(E\left[\|\nabla v\|^2_{\underline{L}^2(\square_R)} \right] + R^{-d-1} \right).$$

Connecting above estimates and using the fact, from (4.18), we have $\delta \geq \widetilde{\varepsilon}$, we arrive at

$$(4.34) \quad |\mathcal{Z}|^{-1} \sum_{y \in \mathcal{Z}} \mathbb{E}\left[\fint_{\square_t(y)} |\nabla \eta_y(x)|^2 \, dx \right]$$

$$\leq C\left(\left(\frac{l}{r-t}\right)^2 \left(\frac{s}{l}\right)^{2k+2} + \left(\frac{r^{1-\delta}}{r-t}\right)^2 \right) \mathbb{E}\left[\|\nabla v\|_{\underline{L}^2(\square_R)}^2 \right] + CR^{-d-1}.$$

By the choice of k and t in (4.18) and (4.32), respectively, we have

$$\left(\frac{l}{r-t}\right)^2 \left(\frac{s}{l}\right)^{2k+2} = R^{2(m\varepsilon + \widetilde{\varepsilon} - k\varepsilon)} = R^{-2\varepsilon\left(k - \frac{6dm(1+\theta)}{\theta} - m\right)} \leq R^{-4dm\varepsilon},$$

and by the condition for ε in (4.18),

$$\left(\frac{r^{1-\delta}}{r-t}\right)^2 = R^{-2(1-\varepsilon(m+2))\left(\delta - \frac{6dm(1+\theta)\varepsilon}{\theta}\right)} \leq R^{-4dm\varepsilon}.$$

Therefore the induction hypothesis yields by way of (4.9) that

$$(4.35) \qquad |\mathcal{Z}|^{-1} \sum_{y \in \mathcal{Z}} \mathbb{E}\left[\fint_{\square_t(y)} |\nabla \eta_y(x)|^2 \, dx \right] \leq CKR^{-d\alpha - 4dm\varepsilon}.$$

To treat the second term in (4.33), Meyers' estimates (4.19) and (4.20) provide us, via Hölder's inequality,

$$\frac{1}{|\square_r|} \int_{y + \square_r \smallsetminus \square_t} |\nabla \eta(x)|^2 \, dx \leq \left(\frac{|\square_r \smallsetminus \square_t|}{|\square_r|} \right)^{\frac{\theta}{1+\theta}} \left(\fint_{\square_r(y)} |\nabla \eta(x)|^{2+2\theta} \, dx \right)^{\frac{1}{1+\theta}}$$

$$\leq C\left(\frac{r-t}{r} \right)^{\frac{\theta}{1+\theta}} \fint_{\square_{3r}(y)} \left(|\nabla v(x)| + |\nabla w_{[y]}(x)| \right)^2 \, dx.$$

Following Step 1 in the proof of Lemma 4.4 and applying (4.25) once more we obtain

$$\mathbb{E}\left[\frac{1}{|\square_r|} \int_{y + \square_r \smallsetminus \square_t} |\nabla \eta(x)|^2 \, dx \right] \leq C\left(\frac{r-t}{r} \right)^{\frac{\theta}{1+\theta}} \left(\mathbb{E}\left[\|\nabla v\|_{\underline{L}^2(\square_l(y))}^2 \right] + CR^{-2d-1} \right).$$

By the choice of t in (4.32) and ε in (4.18) we have

$$\left(\frac{r-t}{r} \right)^{\frac{\theta}{1+\theta}} = R^{-(1-\varepsilon(m+2))\frac{\widetilde{\varepsilon}\theta}{1+\theta}} \leq R^{-4dm\varepsilon}.$$

Therefore the summation yields, as before,

$$(4.36) \qquad |\mathcal{Z}|^{-1} \sum_{y \in \mathcal{Z}} \mathbb{E}\left[\frac{1}{|\square_r|} \int_{y + \square_r \smallsetminus \square_t} |\nabla \eta(x)|^2 \, dx \right] \leq CKR^{-d\alpha - 4dm\varepsilon}.$$

Combining (4.33) with (4.35) and (4.36) yields the statement of the lemma. $\quad\square$

Next we show, using the previous lemma and the uniform convexity of $\mathcal{J}$ in p and q separately, that the expected difference between $\mathcal{J}(v, \square_r(y), p, q)$ and $\mathcal{J}(v_y, \square_r(y), p, q)$ is small, after averaging over all $y \in \mathcal{Z}$.

Lemma 4.6. *Assume* (4.16) *and* (4.18). *There exists* $C(d, \Lambda, k) < \infty$ *such that*

$$(4.37) \qquad |\mathcal{Z}|^{-1} \sum_{y \in \mathcal{Z}} \mathbb{E}\big[\,|\mathcal{J}(v, \square_r(y), p, q) - \mathcal{J}(v_y, \square_r(y), p, q)|\,\big] \leq CKR^{-d(\alpha+\varepsilon)}.$$

Proof. For convenience, let $\widetilde{u}_y := u(\cdot, \square_r(y), p, q)$ denote the minimizer in the definition of $J(\square_r(y), p, q)$. Using (3.3) twice and summing, we find that

$$(4.38) \quad \mathcal{J}(v, \square_r(y), p, q) - \mathcal{J}(v_y, \square_r(y), p, q)$$
$$= \frac{1}{2}\left\langle \nabla v - \nabla v_y, \mathbf{a}\left(2\nabla\widetilde{u}_y - \nabla v - \nabla v_y\right)\right\rangle_{\square_r(y)}.$$

Thus

$$(4.39) \quad |\mathcal{J}(v, \square_r(y), p, q) - \mathcal{J}(v_y, \square_r(y), p, q)|$$
$$\leq C \fint_{\square_r(y)} |\nabla v(x) - \nabla v_y(x)| \left(|\nabla\widetilde{u}_y(x)| + |\nabla v(x)| + |\nabla v_y(x)|\right) dx.$$

We will estimate the term on the right. By (4.22) (with l replaced by r),

$$\mathbb{E}\left[\|\nabla\widetilde{u}_y\|^2_{\underline{L}^2(\square_r(y))}\right] \leq C\mathbb{E}\left[J(\square_r(y), p, q)\right]$$
$$\leq C\mathbb{E}\left[J(\square_r(y), p, Q(\square_r)p)\right] + CKr^{-d\alpha}$$
$$\leq CKr^{-d\alpha}.$$

On the other hand, by the induction hypothesis and (4.25), we have

$$|\mathcal{Z}|^{-1} \sum_{y \in \mathcal{Z}} \mathbb{E}\left[\|\nabla v_y\|^2_{\underline{L}^2(\square_r(y))}\right] \leq C|\mathcal{Z}|^{-1} \sum_{y \in \mathcal{Z}} \mathbb{E}\left[\|\nabla w_{[y]}\|^2_{\underline{L}^2(\square_r(y))}\right]$$
$$\leq CKR^{-d\alpha}.$$

Lemma 4.5 further gives

$$|\mathcal{Z}|^{-1} \sum_{y \in \mathcal{Z}} \mathbb{E}\left[\|\nabla v - \nabla v_y\|^2_{\underline{L}^2(\square_r(y))}\right] \leq CKR^{-d\alpha}.$$

Combining above three displays and using the triangle inequality yield

$$|\mathcal{Z}|^{-1} \sum_{y \in \mathcal{Z}} \mathbb{E}\left[\|\nabla\widetilde{u}_y\|^2_{\underline{L}^2(\square_r(y))} + \|\nabla v_y\|^2_{\underline{L}^2(\square_r(y))} + \|\nabla v\|^2_{\underline{L}^2(\square_r(y))}\right] \leq CKr^{-d\alpha}.$$

We now use Hölder's inequality, Lemma 4.5 and the display above to get

$$|\mathcal{Z}|^{-1} \sum_{y \in \mathcal{Z}} \mathbb{E}\left[\fint_{\square_r(y)} |\nabla v(x) - \nabla v_y(x)| \left(|\nabla\widetilde{u}_y(x)| + |\nabla v(x)| + |\nabla v_y(x)|\right) dx\right]$$
$$\leq C\left(\mathbb{E}\left[|\mathcal{Z}|^{-1} \sum_{y \in \mathcal{Z}} \|\nabla v - \nabla v_y\|^2_{\underline{L}^2(\square_r(y))}\right]\right)^{\frac{1}{2}} CK^{\frac{1}{2}} r^{-d\alpha/2}$$
$$\leq CKR^{-d\alpha/2 - 2dm\varepsilon} r^{-d\alpha/2}.$$

Now we return to (4.39) and use the previous estimate to obtain

$$|\mathcal{Z}|^{-1} \sum_{y \in \mathcal{Z}} \mathbb{E}\big[\,|\mathcal{J}(v, \square_r(y), p, q) - \mathcal{J}(v_y, \square_r(y), p, q)|\,\big] \leq CKr^{-d\alpha/2} R^{-d\alpha/2} R^{-2dm\varepsilon}.$$

By the choice of r in (4.14), we have $r^{-d\alpha/2} = CR^{-d\alpha/2} R^{d\alpha(m+2)\varepsilon/2}$, and therefore we get the desired result using $\alpha \leq 1$ and $m \geq \max\{d, 2\}$. $\qquad\square$

4.5. **Improving the exponent using independence.** As in the proof of Lemma 4.6, we set $\widetilde{u}_y := u(\cdot, \square_r(y), p, q)$. Applying (3.2) and (4.30), we find that, for every $y \in \mathcal{Z}$,

$$(4.40) \qquad \mathcal{J}(v_y, \square_r(y), p, q) \leq \langle -\mathbf{a}p + q, \nabla v_y \rangle_{\square_r(y)}$$
$$= \langle \mathbf{a}\nabla\widetilde{u}_y, \nabla v_y \rangle_{\square_r(y)} = \langle \mathbf{a}\nabla\widetilde{u}_y, \nabla w_{[y]} \rangle_{\square_r(y)}.$$

This together with (4.21) and Lemma 4.6 imply that

$$(4.41) \quad \mathbb{E}[J(\square_R, p, q)] \leq C\left(\frac{R}{r}\right)^{-d} \mathbb{E}\left[\sum_{y \in \mathcal{Z}} \langle \mathbf{a}\nabla\widetilde{u}_y, \nabla w_{[y]} \rangle_{\square_r(y)}\right]$$
$$+ CK\left(\frac{R}{l}\right)^{-1} l^{-d\alpha} + CKR^{-d(\alpha+\varepsilon)}.$$

The sum on the right side of (4.41) allows us to see a CLT-type scaling because the terms are essentially independent for each y inside a single larger mesoscopic cube $\square_s(z)$. This is the mechanism which improves the exponent α. The precise statement we need is formalized in the following lemma. We remark that this is the only point in the proof that we use the choice $q = Q(\square_R)p$.

Lemma 4.7. *Assume* (4.16) *and* (4.18) *hold. There exists* $C(d, \Lambda, k) < \infty$ *such that, for every* $z \in s\mathbb{Z}^d \cap \square_R^\circ$,

$$(4.42) \quad \mathbb{E}\left[\sup_{w \in \mathcal{P}_k} \left\{ \left(\|\nabla w\|_{L^\infty(\square_s(z))} + l\|\nabla^2 w\|_{L^\infty(\square_s(z))}\right)^{-2} \right.\right.$$
$$\left.\left. \times \left(\sum_{y \in r\mathbb{Z}^d \cap \square_s(z)} \langle \mathbf{a}\nabla\widetilde{u}_y, \nabla w \rangle_{\square_r(y)}\right)^2 \right\}\right]$$
$$\leq CK\left(\frac{s}{r}\right)^{2d} r^{-d\alpha}\left(\left(\frac{s}{r}\right)^{-d} + \left(\frac{l}{r}\right)^{-2} + Kr^{-d\alpha}\right).$$

Proof. We fix $z \in s\mathbb{Z}^d \cap \square_R^\circ$ throughout the proof.

We first prove a bound for each particular $w \in \mathcal{P}_k$, and then put the supremum over w inside the expectation, using linearity and that $\mathcal{P}_k$ is a finite dimensional vector space. We begin with an estimate of the expectation of each term in the sum.

Step 1. We show that, for every $w \in \mathcal{P}_k$ and $y \in r\mathbb{Z}^d \cap \square_s(z)$,

$$(4.43) \quad \left|\mathbb{E}\left[\langle \mathbf{a}\nabla\widetilde{u}_y, \nabla w \rangle_{\square_r(y)}\right]\right|$$
$$\leq CKr^{-d\alpha}\|\nabla w\|_{L^\infty(\square_s(z))} + CK^{\frac{1}{2}}r^{1-\frac{d\alpha}{2}}\|\nabla^2 w\|_{L^\infty(\square_s(z))}.$$

We first observe that

$$\left|\langle \mathbf{a}\nabla\widetilde{u}_y, \nabla w \rangle_{\square_r(y)} - \nabla w(y) \cdot \fint_{\square_r(y)} \mathbf{a}(x)\nabla\widetilde{u}_y(x)\,dx\right|$$
$$\leq C \sup_{x \in \square_r(y)} |\nabla w(x) - \nabla w(y)| \fint_{\square_r(x)} |\nabla\widetilde{u}_y(x)|\,dx.$$

Thus, by (4.22) and the induction hypothesis,

$$(4.44) \qquad \mathbb{E}\left[\left|\langle \mathbf{a}\nabla\widetilde{u}_y, \nabla w\rangle_{\square_r(y)} - \nabla w(y) \cdot \fint_{\square_r(y)} \mathbf{a}(x)\nabla\widetilde{u}_y(x)\,dx\right|\right]$$

$$\leq Cr\left\|\nabla^2 w\right\|_{L^\infty(\square_r(y))} \mathbb{E}\left[\fint_{\square_r(x)} |\nabla\widetilde{u}_y(x)|^2\,dx\right]^{\frac{1}{2}}$$

$$\leq CK^{\frac{1}{2}} r^{1-\frac{d\alpha}{2}} \left\|\nabla^2 w\right\|_{L^\infty(\square_r(y))}.$$

On the other hand, using Lemma 3.6 and the induction hypothesis again, we get, for every $\square \in \mathcal{C}$,

$$\left|\bar{\mathbf{a}}^{-1} - P(\square)\right| \leq CK\,|\square|^{-\alpha},$$

from which we deduce, using also (4.22), that

$$|p - P(\square_r(y))q| \leq CKr^{-d\alpha}.$$

Therefore

$$\left|\mathbb{E}\left[\fint_{\square_r(y)} \mathbf{a}(x)\nabla\widetilde{u}_y(x)\,dx\right]\right| = \left|\mathbb{E}\left[\nabla_p J\left(\square_r(y), p, q\right)\right]\right|$$

$$\leq \left|\mathbb{E}\left[\nabla_p J\left(\square_r(y), P(\square_r(y))q, q\right)\right]\right| + CKr^{-d\alpha}$$

$$= CKr^{-d\alpha},$$

as $\mathbb{E}\left[\nabla_p J\left(\square_r(y), P(\square_r(y))q, q\right)\right] = 0$ by (3.20), and hence

$$(4.45) \qquad \left|\mathbb{E}\left[\nabla w(y) \cdot \fint_{\square_r(y)} \mathbf{a}(x)\nabla\widetilde{u}_y(x)\,dx\right]\right| \leq CK\,|\nabla w(y)|\,r^{-d\alpha}.$$

We now obtain (4.43) by combining (4.44) and (4.45) and the triangle inequality.

Step 2. We show using independence that, for every $w \in \mathcal{P}_k$,

$$(4.46) \qquad \mathrm{var}\left[\sum_{y\in r\mathbb{Z}^d\cap\square_s(z)} \langle \mathbf{a}\nabla\widetilde{u}_y, \nabla w\rangle_{\square_r(y)}\right] \leq CKr^{-d\alpha}\left(\frac{s}{r}\right)^d \|\nabla w\|_{L^\infty(\square_s(z))}^2.$$

We expand the variance by writing

$$\mathrm{var}\left[\sum_{y\in r\mathbb{Z}^d\cap\square_s(z)} \langle \mathbf{a}\nabla\widetilde{u}_y, \nabla w\rangle_{\square_r(y)}\right]$$

$$= \sum_{y,y'\in r\mathbb{Z}^d\cap\square_s(z)} \mathrm{cov}\left[\langle \mathbf{a}\nabla\widetilde{u}_y, \nabla w\rangle_{\square_r(y)}, \langle \mathbf{a}\nabla\widetilde{u}_{y'}, \nabla w\rangle_{\square_r(y')}\right].$$

Using independence, the fact that $\langle \mathbf{a}\nabla\widetilde{u}_y, \nabla w\rangle_{\square_r(y)}$ is $\mathcal{F}(\square_r(y))$–measurable and each cube $\square_{r+1}(y)$ has nonempty intersection with at most C cubes of the form $\square_{r+1}(y')$ with $y' \in r\mathbb{Z}^d \cap \square_s(z)$, we obtain

$$\sum_{y,y'\in r\mathbb{Z}^d\cap\square_s(z)} \mathrm{cov}\left[\langle \mathbf{a}\nabla\widetilde{u}_y, \nabla w\rangle_{\square_r(y)}, \langle \mathbf{a}\nabla\widetilde{u}_{y'}, \nabla w\rangle_{\square_r(y')}\right]$$

$$\leq C\sum_{y\in r\mathbb{Z}^d\cap\square_s(z)} \mathrm{var}\left[\langle \mathbf{a}\nabla\widetilde{u}_y, \nabla w\rangle_{\square_r(y)}\right].$$

Finally, we observe that the induction hypothesis gives, for each $y \in r\mathbb{Z}^d \cap \square_s(z)$,

$$\mathrm{var}\left[\langle \mathbf{a}\nabla\widetilde{u}_y, \nabla w\rangle_{\square_r(y)}\right] \le C\mathbb{E}\left[\|\nabla\widetilde{u}_y\|_{\underline{L}^2(\square_r(y))}^2\right]\|\nabla w\|_{L^\infty(\square_s(z))}^2$$
$$\le CKr^{-d\alpha}\|\nabla w\|_{L^\infty(\square_s(z))}^2.$$

This completes the proof of (4.46).

Step 3. We complete the proof. Observe that combining the results of the first two steps gives, for every $w \in \mathcal{P}_k$,

$$(4.47) \quad \mathbb{E}\left[\left(\sum_{y \in r\mathbb{Z}^d \cap \square_s(z)} \langle \mathbf{a}\nabla\widetilde{u}_y, \nabla w\rangle_{\square_r(y)}\right)^2\right]$$
$$\le CK\left(\frac{s}{r}\right)^{2d} r^{-d\alpha}\left(\left(\frac{s}{r}\right)^{-d} + \left(\frac{l}{r}\right)^{-2} + Kr^{-d\alpha}\right)$$
$$\times \left(\|\nabla w\|_{L^\infty(\square_s(z))} + l\left\|\nabla^2 w\right\|_{L^\infty(\square_s(z))}\right)^2.$$

To conclude, it remains to smuggle the supremum over $w \in \mathcal{P}_k$ inside the expectation. As $\mathcal{P}_k$ is a finite dimensional vector space with dimension depending only on (k,d), there exists an integer $N(k,d) \in \mathbb{N}$ and $w_1, \dots, w_N \in \mathcal{P}_k$ such that $\{w_1, \dots, w_N\}$ is a basis for $\mathcal{P}_k$, each w_j satisfies

$$\|\nabla w_j\|_{L^\infty(\square_s(z))} + l\left\|\nabla^2 w_j\right\|_{L^\infty(\square_s(z))} = 1,$$

and for any $w \in \mathcal{P}_k$ expressed in the form

$$w = c_1 w_1 + \dots + c_N w_N,$$

we have, for some $C(k,d) \ge 1$,

$$|c_1| + \dots + |c_N| \le C\left(\|\nabla w\|_{L^\infty(\square_s(z))} + l\left\|\nabla^2 w\right\|_{L^\infty(\square_s(z))}\right).$$

It follows that

$$\sup_{w \in \mathcal{P}_k}\left(\|\nabla w\|_{L^\infty(\square_s(z))} + l\left\|\nabla^2 w\right\|_{L^\infty(\square_s(z))}\right)^{-1}\left|\sum_{y \in r\mathbb{Z}^d \cap \square_s(z)} \langle \mathbf{a}\nabla\widetilde{u}_y, \nabla w\rangle_{\square_r(y)}\right|$$
$$\le C\sum_{j=1}^{N}\left|\sum_{y \in r\mathbb{Z}^d \cap \square_s(z)} \langle \mathbf{a}\nabla\widetilde{u}_y, \nabla w_j\rangle_{\square_r(y)}\right|.$$

Squaring and taking the expectation of the previous inequality, using that $N \le C$ and then applying (4.47), we get

$$\mathbb{E}\left[\sup_{w\in\mathcal{P}_k}\left(\|\nabla w\|_{L^\infty(\square_s(z))} + l\,\|\nabla^2 w\|_{L^\infty(\square_s(z))}\right)^{-2}\left(\sum_{y\in r\mathbb{Z}^d\cap\square_s(z)}\langle\mathbf{a}\nabla\widetilde{u}_y,\nabla w\rangle_{\square_r(y)}\right)^2\right]$$

$$\le C\mathbb{E}\left[\left(\sum_{j=1}^{N}\left|\sum_{y\in r\mathbb{Z}^d\cap\square_s(z)}\langle\mathbf{a}\nabla\widetilde{u}_y,\nabla w_j\rangle_{\square_r(y)}\right|\right)^2\right]$$

$$\le C\sum_{j=1}^{N}\mathbb{E}\left[\left(\sum_{y\in r\mathbb{Z}^d\cap\square_s(z)}\langle\mathbf{a}\nabla\widetilde{u}_y,\nabla w_j\rangle_{\square_r(y)}\right)^2\right]$$

$$\le CK\left(\frac{s}{r}\right)^{2d}r^{-d\alpha}\left(\left(\frac{s}{r}\right)^{-d}+\left(\frac{l}{r}\right)^{-2}+Kr^{-d\alpha}\right).$$

This completes the proof. $\qquad\square$

Combining (4.40) and Lemma 4.7, we obtain

$$\sum_{y\in\mathcal{Z}}\mathbb{E}\left[\mathcal{J}(v_y,\square_r(y),p,q)\right]\le CK^{\frac{1}{2}}\left(\frac{s}{r}\right)^d r^{-\frac{d\alpha}{2}}\left(\left(\frac{s}{r}\right)^{-\frac{d}{2}}+\left(\frac{l}{r}\right)^{-1}+K^{\frac{1}{2}}r^{-\frac{d\alpha}{2}}\right)$$

$$\times\sum_{z\in s\mathbb{Z}^d\cap\square_R^\circ}\mathbb{E}\left[\left(\|\nabla w_z\|_{L^\infty(\square_s(z))}^2 + l^2\,\|\nabla^2 w_z\|_{L^\infty(\square_s(z))}^2\right)\right]^{\frac{1}{2}}.$$

By Lemma 4.4 and (4.9), the sum in the line above is bounded by

$$C\left(\frac{R}{s}\right)^d K^{\frac{1}{2}}R^{-\frac{d\alpha}{2}},$$

so that

$$\sum_{y\in\mathcal{Z}}\mathbb{E}\left[\mathcal{J}(u_y,\square_r(y),p,q)\right]\le CK\left(\frac{R}{r}\right)^d R^{-\frac{d\alpha}{2}}r^{-\frac{d\alpha}{2}}\left(\left(\frac{s}{r}\right)^{-\frac{d}{2}}+\left(\frac{l}{r}\right)^{-1}+K^{\frac{1}{2}}r^{-\frac{d\alpha}{2}}\right).$$

Combining this with (4.41) and using $K\ge 1$, we obtain

$$\mathbb{E}\left[J(\square_R,p,q)\right]$$

$$\le CK^{\frac{3}{2}}\left(R^{-\frac{d\alpha}{2}}r^{-\frac{d\alpha}{2}}\left(\left(\frac{s}{r}\right)^{-\frac{d}{2}}+\left(\frac{l}{r}\right)^{-1}\right)+R^{-\frac{d\alpha}{2}}r^{-d\alpha}+\left(\frac{R}{l}\right)^{-1}l^{-d\alpha}+R^{-d(\alpha+\varepsilon)}\right).$$

Using the definitions of the mesoscales,

$$\begin{cases} R^{-\frac{d\alpha}{2}}r^{-\frac{d\alpha}{2}}\left(\frac{s}{r}\right)^{-\frac{d}{2}} = CR^{-d\alpha}R^{-d\varepsilon\left(\frac{m}{2}(1-\alpha)-\alpha\right)} \\[2mm] R^{-\frac{d\alpha}{2}}r^{-\frac{d\alpha}{2}}\left(\frac{l}{r}\right)^{-1} = CR^{-d\alpha}R^{-\varepsilon(m+1-d\alpha(m+2)/2)} \\[2mm] R^{-\frac{d\alpha}{2}}r^{-d\alpha} = CR^{-d\alpha}R^{d\varepsilon(m+2)-\frac{d\alpha}{2}} \\[2mm] \left(\frac{R}{l}\right)^{-1}l^{-d\alpha} = CR^{-d\alpha}R^{\varepsilon(d\alpha-1)}. \end{cases}$$

It is easy to check that that the choices made in (4.16), (4.17) and (4.18) guarantee that the first three terms above are bounded by $R^{-d(\alpha+\varepsilon)}$.

In view of these choices of the parameters, it is clear that the third term (the error in removing the boundary layer) is the limiting one, and we deduce that

$$\mathbb{E}\left[J\left(\square_R, p, q\right)\right] \le CK\left(R^{-d\alpha}R^{\varepsilon(d\alpha-1)} + R^{-d(\alpha+\varepsilon)}\right) \le CKR^{-d\alpha}R^{\varepsilon(d\alpha-1)}.$$

Thus $\mathcal{S}(\alpha+\varepsilon(1-d\alpha)/d, CK)$ holds, for $\varepsilon \ge c(d, \lambda)$ and $C(d, \Lambda) < \infty$ which, after a redefinition of ε, completes the proof of the claim (4.6) and therefore the proof of Proposition 4.1.

5. Improvement of stochastic integrability by subadditivity

The aim of this section is to complete the proof of Theorem 1.1, by strengthening the stochastic integrability of Proposition 4.1 (or Corollary 4.2) from L^1 to exponential moments. The rough argument is as follows. We decompose $\square_R$ into subcubes $(\square_r(y))$. By subadditivity, we can bound the upper fluctuations of $\nu(\square_R, p)$ by the upper fluctuations of the average over y of $\nu(\square_r(y), p)$, up to an error controlled by the difference $\mathbb{E}[\nu(\square_r, p)] - \mathbb{E}[\nu(\square_R, p)]$, which is small by Corollary 4.2. By independence, the average over y of $\nu(\square_r(y), p)$ is unlikely to be large. The same argument applied to $-\mu$ gives a control of the lower fluctuations of μ. By duality, we can then control upper and lower fluctuations of both quantities.

We make this idea precise in the following general statement. (Recall our slightly non-standard definition of subadditivity in (1.8), and that $\mathcal{C}$ is the set of cubes $\square$ such that $|\square| \ge 1$.)

Theorem 5.1. *Let $\delta < 1$, $R_0 < \infty$, and let $\widetilde{\mu}(\cdot) \le \widetilde{\nu}(\cdot)$ be respectively super- and subadditive quantities, such that for every $R \ge R_0$ and $x \in \mathbb{R}^d$,*

$$\widetilde{\mu}(\square_R(x)) \text{ and } \widetilde{\nu}(\square_R(x)) \text{ are } \mathcal{F}(\square_{R+R^\delta}(x))\text{-measurable.}$$

Let $\alpha > 0$ and $\beta \in (0, \alpha) \cap (0, 1/2]$. Assume that there exists $c < \infty$ such that for every cube $\square \in \mathcal{C}$,

$$(5.1) \qquad\qquad \mathbb{E}[\widetilde{\nu}(\square)] - \mathbb{E}[\widetilde{\mu}(\square)] \le \frac{c}{|\square|^\alpha}.$$

Assume furthermore that there exists $c' < \infty$ such that for every cube $\square$ satisfying $1 \le |\square| \le 2^d$ and every $\lambda \in \mathbb{R}$,

$$(5.2) \qquad\qquad \mathbb{E}[\exp(\lambda\widetilde{\mu}(\square))] \vee \mathbb{E}[\exp(\lambda\widetilde{\nu}(\square))] \le c'(1 + \lambda^2).$$

Then there exists $C = C(d, \delta, R_0, \alpha, \beta, c, c') < \infty$ such that for every cube $\square \in \mathcal{C}$ and every $\lambda \in \mathbb{R}$,

$$(5.3) \qquad\qquad \log\mathbb{E}[\exp(\lambda|\square|^\beta(\widetilde{\nu}(\square) - \mathbb{E}[\widetilde{\nu}(\square)]))] \le C\lambda^2,$$

and the same estimate holds with $\widetilde{\nu}$ replaced by $\widetilde{\mu}$.

Remark 5.2. The conclusion of Theorem 5.1 can be strengthened in several directions. First, one can replace λ^2 by $\lambda^2 \wedge \lambda^{\frac{1}{1-\beta}}$ in the right side of (5.3). Moreover, we believe that (5.3) holds with $\alpha = \beta$ for $\alpha < \frac{1}{2}$ (and with a logarithmic correction when $\alpha = \frac{1}{2}$), and in fact, the proof given below can be adapted to show this stronger result when α is sufficiently small. On the other

hand, even if $\widetilde{\mu}$ and $\widetilde{\nu}$ were additive, one could not improve the conclusion to $\beta > \frac{1}{2}$, by the central limit theorem.

The proof of Theorem 5.1 makes use of the following lemma.

Lemma 5.3 (parabolicity of log-Laplace). *There exists $\lambda_0 \in (0,1]$ such that the following holds. Let $a \geq 0$ and X be a random variable such that $\mathbb{E}[\exp(|X|)] < \infty$ and $\mathbb{E}[X] = 0$. If the inequality*

$$(5.4) \qquad\qquad \log \mathbb{E}[\exp(\lambda X)] \leq a\lambda^2$$

holds for every λ such that $|\lambda| \in [\lambda_0, 1]$, then it holds for every $\lambda \in [-1, 1]$.

Proof. Since $\mathbb{E}[\exp(|X|)] < \infty$, the function $\Psi := \lambda \mapsto \log \mathbb{E}[\exp(\lambda X)]$ is infinitely differentiable on $(-1, 1)$. Its value and first derivative at 0 vanish, while its second derivative at $|\lambda| < 1$ is bounded by $\mathbb{E}[X^2 \exp(\lambda X)]$. We choose $\lambda_0 > 0$ sufficiently small that for every $\lambda \in [-\lambda_0, \lambda_0]$,

$$\forall x \in \mathbb{R}, \quad x^2 \exp(\lambda x) \leq \exp(x) + \exp(-x).$$

(By symmetry and monotonicity, it suffices to check the inequality for $\lambda = \lambda_0$.) In particular, for every $\lambda \in [-\lambda_0, \lambda_0]$, we have

$$\Psi''(\lambda) \leq \mathbb{E}[\exp(X)] + \mathbb{E}[\exp(-X)] \leq 2a.$$

The result then follows by integration. $\qquad\qquad\qquad\qquad\qquad\square$

Proof of Theorem 5.1. For any positive integer m and $\mathbf{C} \geq 0$, we denote by $\mathcal{A}_m(\mathbf{C})$ the assertion that for every cube $\square$ satisfying $1 \leq |\square| \leq 3^{dm}$ and every $\lambda \in \mathbb{R}$,

$$(5.5) \qquad\qquad \log \mathbb{E}[\exp(\lambda|\square|^\beta(\widetilde{\nu}(\square) - \mathbb{E}[\widetilde{\nu}(\square)]))] \leq \mathbf{C}\lambda^2,$$

and that the same estimate holds with $\widetilde{\nu}$ replaced by $\widetilde{\mu}$.

Step 1. We show that for any given m_0, there exists $\mathbf{C} < \infty$ such that $\mathcal{A}_{m_0}(\mathbf{C})$ holds. Every cube $\square$ satisfying $1 \leq |\square| \leq 3^{dm_0}$ can be decomposed into a finite number of subcubes of side length between 1 and 2. By subadditivity, the hypothesis and Hölder's inequality,

$$(5.6) \qquad\qquad \log \mathbb{E}[\exp(\lambda(\widetilde{\nu}(\square) - \mathbb{E}[\widetilde{\nu}(\square)]))] \lesssim 1 + \lambda^2,$$

uniformly over cubes $\square$ satisfying $1 \leq |\square| \leq 3^{dm_0}$ and $\lambda \in \mathbb{R}$. The conclusion for $\widetilde{\nu}$ then follows by Lemma 5.3, and the reasoning for $\widetilde{\mu}$ is identical.

Step 2. We show that there exists $\varepsilon > 0$ such that for every m sufficiently large and $\mathbf{C} \geq 1$,

$$\mathcal{A}_{m-1}(\mathbf{C}) \implies \mathcal{A}_m((1 + 3^{-\varepsilon m})\mathbf{C}).$$

Since $\prod_m (1 + 3^{-\varepsilon m}) < \infty$, this will complete the proof of the proposition.

Assuming $\mathcal{A}_{m-1}(\mathbf{C})$, we give ourselves a cube $\square$ of side length R such that $3^{m-1} < R \leq 3^m$. For notational convenience, we assume that $\square = (0, R)^d$. We define a partition of $\square$ into 3^d subcubes of side length L, each subcube being surrounded by a layer of smaller cubes of side length $\ell = 3^{\gamma m}$ for some $\gamma \in (\delta, 1)$. (We choose a triadic decomposition of $\square$ for coherence with the rest of the paper, but a dyadic one would be fine too.) In order for the partition to be well-defined, we ask that $L = (R - 2\ell)/3$ be an integer multiple of ℓ. This requirement can

easily be taken care of since $\ell \ll R$, so we will neglect it for clarity. We let $(z_i)_{1 \le i \le 3^d}$ be the centers of the 3^d subcubes of side length L such that

$$\left((0, L) \cup (\ell + L, \ell + 2L) \cup (2\ell + 2L, R)\right)^d = \bigcup_{i=1}^{3^d} \square_L(z_i),$$

and $(z'_j)_{1 \le j \le N}$ be the centers of the disjoint subcubes of side length ℓ such that

$$\square \smallsetminus \left(\bigcup_{i=1}^{3^d} \square_L(z_i)\right) = \bigcup_{j=1}^{N} \square_\ell(z'_j) \quad \text{up to a set of null measure,}$$

where $N = (R^d - (3L)^d)/\ell^d$. By subadditivity,

$$\widetilde{\nu}(\square) \le \sum_{i=1}^{3^d} \frac{|\square_L|}{|\square|} \widetilde{\nu}(\square_L(z_i)) + \sum_{j=1}^{N} \frac{|\square_\ell|}{|\square|} \widetilde{\nu}(\square_\ell(z'_j)).$$

Let $r, s \in (1, \infty)$ be such that $1/r + 1/s = 1$. By Hölder's inequality, for every $\lambda \ge 0$,

$$\log \mathbb{E}\left[\exp\left(\lambda |\square|^\beta \left(\widetilde{\nu}(\square) - \mathbb{E}[\widetilde{\nu}(\square)]\right)\right)\right]$$

$$(5.7) \qquad \le \frac{1}{r} \log \mathbb{E}\left[\exp\left(r\lambda |\square|^\beta \frac{|\square_L|}{|\square|} \sum_{i=1}^{3^d} (\widetilde{\nu}(\square_L(z_i)) - \mathbb{E}[\widetilde{\nu}(\square)])\right)\right]$$

$$(5.8) \qquad + \frac{1}{s} \log \mathbb{E}\left[\exp\left(s\lambda |\square|^\beta \frac{|\square_\ell|}{|\square|} \sum_{j=1}^{N} (\widetilde{\nu}(\square_\ell(z'_j)) - \mathbb{E}[\widetilde{\nu}(\square)])\right)\right].$$

We decompose the term in (5.7) into

$$(5.9) \quad \frac{1}{r} \log \mathbb{E}\left[\exp\left(r\lambda \frac{|\square_L|}{|\square|^{1-\beta}} \sum_{i=1}^{3^d} (\widetilde{\nu}(\square_L(z_i)) - \mathbb{E}[\widetilde{\nu}(\square_L(z_i))])\right)\right]$$

$$+ \lambda \frac{|\square_L|}{|\square|^{1-\beta}} \sum_{i=1}^{3^d} (\mathbb{E}[\widetilde{\nu}(\square_L(z_i))] - \mathbb{E}[\widetilde{\nu}(\square)]),$$

and likewise, the term in (5.8) into

$$(5.10) \quad \frac{1}{s} \log \mathbb{E}\left[\exp\left(s\lambda \frac{|\square_\ell|}{|\square|^{1-\beta}} \sum_{j=1}^{N} (\widetilde{\nu}(\square_\ell(z'_j)) - \mathbb{E}[\widetilde{\nu}(\square_\ell(z'_j))])\right)\right]$$

$$+ \lambda \frac{|\square_\ell|}{|\square|^{1-\beta}} \sum_{j=1}^{N} (\mathbb{E}[\widetilde{\nu}(\square_\ell(z'_j))] - \mathbb{E}[\widetilde{\nu}(\square)]).$$

It follows from sub-/superadditivity and (5.1) that

$$\mathbb{E}[\widetilde{\nu}(\square)] \ge \lim_{|\widetilde{\square}| \to \infty} \mathbb{E}[\widetilde{\nu}(\widetilde{\square})] = \lim_{|\widetilde{\square}| \to \infty} \mathbb{E}[\widetilde{\mu}(\widetilde{\square})] \ge \mathbb{E}[\widetilde{\mu}(\square_L(z_i))].$$

Hence, the second term in (5.9) can be estimated using (5.1):

$$\mathbb{E}[\widetilde{\nu}(\square_L(z_i))] - \mathbb{E}[\widetilde{\nu}(\square)] \le \frac{c}{|\square_L|^\alpha},$$

and similarly,

$$\mathbb{E}[\widetilde{\nu}(\square_\ell(z'_j))] - \mathbb{E}[\widetilde{\nu}(\square)] \le \frac{c}{|\square_\ell|^\alpha}.$$

By independence, the first term in (5.9) is equal to

$$\frac{1}{r} \sum_{i=1}^{3^d} \log \mathbb{E}\left[\exp\left(r\lambda \frac{|\square_L|}{|\square|^{1-\beta}}(\widetilde{\nu}(\square_L(z_i)) - \mathbb{E}[\widetilde{\nu}(\square_L(z_i))])\right)\right],$$

which by the induction hypothesis is bounded by

$$(5.11) \qquad \frac{3^d \mathbf{C}}{r}\left(\lambda r \frac{|\square_L|^{1-\beta}}{|\square|^{1-\beta}}\right)^2 \le \mathbf{C} r \lambda^2,$$

since $\beta \le 1/2$ and $3^d|\square_L| \le |\square|$. In order to estimate the first term in (5.10), we split the set $\{z'_j, j' \le N\}$ into 3^d subsets $\{z'_j, j \in Z_1\}$, ..., $\{z'_j, j \in Z_{3^d}\}$ in such a way that if $j_1 \in Z_{k_1}$ and $j_2 \in Z_{k_2}$ with $k_1 \ne k_2$, then the cubes $\square_\ell(z_{j_1})$ and $\square_\ell(z_{j_2})$ are at distance at least ℓ from one another. By Hölder's inequality and independence, we get that the first term in (5.10) is bounded by

$$\frac{1}{3^d s} \sum_{j=1}^N \log \mathbb{E}\left[\exp\left(3^d s \lambda \frac{|\square_\ell|}{|\square|^{1-\beta}}(\widetilde{\nu}(\square_\ell(z'_j)) - \mathbb{E}[\widetilde{\nu}(\square_\ell(z'_j))])\right)\right],$$

which by the induction hypothesis is bounded by

$$(5.12) \qquad \frac{\mathbf{C}N}{3^d s}\left(3^d s \lambda \frac{|\square_\ell|^{1-\beta}}{|\square|^{1-\beta}}\right)^2 \le \mathbf{C}3^d s \frac{N|\square_\ell|}{|\square|}\lambda^2.$$

To sum up, we have shown that

$$\log \mathbb{E}\left[\exp\left(\lambda|\square|^\beta(\widetilde{\nu}(\square) - \mathbb{E}[\widetilde{\nu}(\square)])\right)\right] \le \mathbf{C}\left(r + 3^d s \frac{N|\square_\ell|}{|\square|}\right)\lambda^2 + \lambda c \frac{|\square|^\beta}{|\square_\ell|^\alpha}$$

(recall that $3^d|\square_L| + N|\square_\ell| = |\square|$), where $r, s \in (1, \infty)$ such that $1/r + 1/s = 1$ are arbitrary. Recall that $\ell = 3^{\gamma m}$ with $\gamma \in (\delta, 1)$. Since $\beta < \alpha$, we can choose γ sufficiently close to 1 that

$$\frac{|\square|^\beta}{|\square_\ell|^\alpha} = 3^{-m\varepsilon}$$

for some $\varepsilon > 0$. Moreover, there exists a constant c_d such that $N \le c_d 3^{m(1-\gamma)(d-1)}$, hence

$$\log \mathbb{E}\left[\exp\left(\lambda|\square|^\beta(\widetilde{\nu}(\square) - \mathbb{E}[\widetilde{\nu}(\square)])\right)\right] \le \mathbf{C}\left(r + sc_d 3^d 3^{-m(1-\gamma)}\right)\lambda^2 + \lambda c\, 3^{-m\varepsilon}.$$

We can now choose $s = 3^{m(1-\gamma)/2}$ and thus obtain that for every $\lambda \ge 0$,

$$(5.13) \quad \log \mathbb{E}\left[\exp\left(\lambda|\square|^\beta(\widetilde{\nu}(\square) - \mathbb{E}[\widetilde{\nu}(\square)])\right)\right]$$
$$\le \mathbf{C}\left((1 - 3^{-m(1-\gamma)/2})^{-1} + c_d 3^{d-m(1-\gamma)/2}\right)\lambda^2 + \lambda c\, 3^{-m\varepsilon}.$$

The same reasoning applied to $-\widetilde{\mu}$ shows that for every $\lambda \le 0$,

$$(5.14) \quad \log \mathbb{E}\left[\exp\left(\lambda|\square|^\beta(\widetilde{\mu}(\square) - \mathbb{E}[\widetilde{\mu}(\square)])\right)\right]$$
$$\le \mathbf{C}\left((1 - 3^{-m(1-\gamma)/2})^{-1} + c_d 3^{d-m(1-\gamma)/2}\right)\lambda^2 - \lambda c\, 3^{-m\varepsilon}.$$

Moreover, by (5.1), we can replace $\mathbb{E}[\widetilde{\mu}(\square)]$ by $\mathbb{E}[\widetilde{\nu}(\square)]$ in (5.14) provided we replace c by $2c$ in the right side of (5.14). Since $\widetilde{\nu} \ge \widetilde{\mu}$, we deduce that for every

$\lambda \in \mathbb{R}$,

$$\log \mathbb{E}\left[\exp\left(\lambda|\Box|^{\beta}\left(\widetilde{\nu}(\Box) - \mathbb{E}[\widetilde{\nu}(\Box)]\right)\right)\right]$$
$$\leq \mathbf{C}\left(\left(1 - 3^{-m(1-\gamma)/2}\right)^{-1} + c_d 3^{d-m(1-\gamma)/2}\right)\lambda^2 + 2|\lambda|c\,3^{-m\varepsilon}.$$

The result for $\widetilde{\nu}$ follows by Lemma 5.3 and similar reasoning applies to $\widetilde{\mu}$. $\quad\square$

We now complete the proof of Theorem 1.1 by combining Proposition 4.1 and Theorem 5.1.

Proof of Theorem 1.1. Let $\alpha < \alpha' < 1/d$, $p \in \mathbb{R}^d$ be a unit vector, and $q = \bar{\mathbf{a}}p$. By Proposition 4.1, there exists $c < \infty$ such that for every $\Box \in \mathcal{C}$,

$$(5.15) \qquad \mathbb{E}[\nu(\Box, p)] - \mathbb{E}[\mu(\Box, q)] - p \cdot q \leq \frac{c}{|\Box|^{\alpha'}}.$$

We apply Theorem 5.1 with $\widetilde{\nu} = \nu(\,\cdot\,, p)$ and $\widetilde{\mu} = \mu(\,\cdot\,, q) + p \cdot q$. The measurability assumption on $\widetilde{\nu}$ and $\widetilde{\mu}$ and the property that $\widetilde{\nu} \leq \widetilde{\mu}$ clearly hold. Assumption (5.2) is also satisfied, since $\widetilde{\nu}$ and $\widetilde{\mu}$ are bounded. We thus obtain the existence of a constant $C < \infty$ such that for every cube $\Box \in \mathcal{C}$ and $\lambda \in \mathbb{R}$,

$$\log \mathbb{E}[\exp(\lambda|\Box|^{\alpha}(\nu(\Box, p) - \mathbb{E}[\nu(\Box, p)]))] \leq C\lambda^2.$$

Moreover, the constant C does not depend on the unit vector p (since the same is true of the constant c in (5.15)). By Corollary 4.2, we obtain, for every $\lambda \in \mathbb{R}$,

$$\log \mathbb{E}\left[\exp\left(\lambda|\Box|^{\alpha}\left|\nu(\Box, p) - \frac{1}{2}p \cdot \bar{\mathbf{a}}p\right|\right)\right] \leq C(1 + \lambda^2).$$

The extension to arbitrary $p \in \mathbb{R}^d$ then follows by homogeneity. The reasoning for μ is identical. $\quad\square$

6. Sublinear growth of the correctors

In this section we prove Theorem 1.2. The proof naturally divides into two main steps: first we introduce a functional inequality, which appears to be new and which we term the *multiscale Poincaré* inequality. It converts control of spatial averages of gradients into control over the function itself. Then we show, using Theorem 1.1, that estimates of spatial averages of the gradient of the correctors can be reduced to the convergence of the subadditive energy quantities.

In this section it is convenient to work with triadic cubes, so we change the notation from the rest of the paper: for every $m \in \mathbb{N}$,

$$\Box_m := \left(-\frac{1}{2}3^m, \frac{1}{2}3^m\right)^d.$$

6.1. **Multiscale Poincaré inequality.** Here we present an inequality which gives an estimate of the H^{-1} norm of ∇u in terms of spatial averages of ∇u in cubes. This can be seen as a generalization of the usual Poincaré inequality giving the bound, for every $u \in H^1(\Box_m)$,

$$(6.1) \qquad \fint_{\Box_m} \left|u(x) - (u)_{\Box_m}\right|^2 dx \leq C(d)3^{2m} \fint_{\Box_m} |\nabla u(x)|^2 \, dx.$$

The sharpness of the scaling of the constant $C3^{2m}$ in the Poincaré inequality is, of course, realized by considering an affine function. In the following proposition,

we show that this scaling can be improved for functions with gradients having small spatial averages relative to their absolute size: roughly, if the gradient is canceling itself out, then the function has smaller oscillation.

Proposition 6.1 (Multiscale Poincaré inequality)**.** *Fix $m \in \mathbb{N}$ and, for each $n \in \mathbb{N}$, $n \le m$, define $\mathcal{Z}_n := 3^n \mathbb{Z}^d \cap \square_m$. There exists a constant $C(d) < \infty$ such that, for every $u \in H^1(\square_m)$,*

$$(6.2) \quad \left\| u - (u)_{\square_m} \right\|_{\underline{L}^2(\square_m)} + \left\| \nabla u \right\|_{\underline{H}^{-1}(\square_m)}$$

$$\le C \left\| \nabla u \right\|_{\underline{L}^2(\square_m)} + C \sum_{n=0}^{m-1} 3^n \left(|\mathcal{Z}_n|^{-1} \sum_{y \in \mathcal{Z}_n} \left| (\nabla u)_{y+\square_n} \right|^2 \right)^{\frac{1}{2}}.$$

Proof. We first prove the estimate for $\left\| \nabla u \right\|_{\underline{H}^{-1}(\square_m)}$ and then deduce the estimate for $\left\| u - (u)_{\square_m} \right\|_{\underline{L}^2(\square_m)}$ as a simple consequence. Without loss of generality, we may suppose $(u)_{\square_m} = 0$.

Step 1. The estimate for $\left\| \nabla u \right\|_{\underline{H}^{-1}(\square_m)}$. Fix $\eta \in H^1(\square_m; \mathbb{R}^d)$ with

$$\fint_{\square_m} |\nabla \eta(x)|^2 \, dx = 1.$$

We denote by $w \in H^2(\square_m)$ the unique (up to additive constants) solution of the Neumann problem

$$\begin{cases} -\Delta w = -\nabla \cdot \eta + b & \text{in } \square_m, \\ \partial_\nu w = 0 & \text{on } \partial\square_m, \end{cases}$$

where $b := \fint_{\square_m} \nabla \cdot \eta(x) \, dx$ is chosen to ensure solvability. Then according to [16, 1], we have $w \in H^2(\square_m)$ and

$$(6.3) \qquad \fint_{\square_m} \left| \nabla^2 w(x) \right|^2 \, dx \le C \fint_{\square_m} |\nabla \eta(x)|^2 \, dx = C.$$

Testing the equation for w with u and using $(u)_{\square_m} = 0$ yields

$$(6.4) \qquad \fint_{\square_m} \nabla u(x) \cdot \eta(x) \, dx = \fint_{\square_m} \nabla u(x) \cdot \nabla w(x) \, dx.$$

For every $n \in \{1, \dots, m\}$ and $z \in \mathcal{Z}_n$, we have

$$\int_{z+\square_n} \nabla u(x) \cdot \left(\nabla w(x) - (\nabla w)_{z+\square_n} \right) dx$$

$$= \sum_{y \in \mathcal{Z}_{n-1} \cap (z+\square_n)} \int_{y+\square_{n-1}} \nabla u(x) \cdot \left(\nabla w(x) - (\nabla w)_{y+\square_{n-1}} \right) dx$$

$$+ |\square_{n-1}| \sum_{y \in \mathcal{Z}_{n-1} \cap (z+\square_n)} \left((\nabla w)_{z+\square_n} - (\nabla w)_{y+\square_{n-1}} \right) \cdot (\nabla u)_{y+\square_{n-1}}.$$

By the Poincaré inequality,

$$\sum_{y \in \mathcal{Z}_{n-1} \cap (z+\square_n)} \left| (\nabla w)_{z+\square_n} - (\nabla w)_{y+\square_{n-1}} \right|^2 \le C 3^{2n} \fint_{z+\square_n} \left| \nabla^2 w(x) \right|^2 \, dx.$$

Therefore, after summing over $z \in \mathcal{Z}_n$ and using Hölder's inequality, we get

$$\sum_{z \in \mathcal{Z}_n} \int_{z+\square_n} \nabla u(x) \cdot \left(\nabla w(x) - (\nabla w)_{z+\square_n} \right) dx$$

$$\leq \sum_{y \in \mathcal{Z}_{n-1}} \int_{y+\square_{n-1}} \nabla u(x) \cdot \left(\nabla w(x) - (\nabla w)_{y+\square_{n-1}} \right) dx$$

$$+ C 3^{n(1+d/2)} \left(\int_{\square_m} \left| \nabla^2 w(x) \right|^2 dx \right)^{\frac{1}{2}} \left(\sum_{y \in \mathcal{Z}_{n-1}} \left| (\nabla u)_{y+\square_{n-1}} \right|^2 \right)^{\frac{1}{2}}.$$

Iterating this and using (6.3),

$$\int_{\square_m} \nabla u(x) \cdot \nabla w(x) \, dx \leq \sum_{z \in \mathcal{Z}_0} \int_{z+\square_0} \nabla u(x) \cdot \left(\nabla w(x) - (\nabla w)_{z+\square_0} \right) dx$$

$$+ C \left| \square_m \right|^{\frac{1}{2}} \sum_{n=0}^{m-1} 3^{n(1+d/2)} \left(\sum_{y \in \mathcal{Z}_n} \left| (\nabla u)_{y+\square_n} \right|^2 \right)^{\frac{1}{2}}.$$

By the Poincaré inequality and (6.3),

$$\sum_{z \in \mathcal{Z}_0} \int_{z+\square_0} |\nabla w(x) - (\nabla w)_{z+\square_0}|^2 \, dx \leq C \int_{\square_m} |\nabla \eta(x)|^2 \, dx = C |\square_m|.$$

Thus using Hölder's and Young's inequalities, we obtain

$$\int_{\square_m} \nabla u(x) \cdot \nabla w(x) \, dx$$

$$\leq C |\square_m|^{\frac{1}{2}} \left(\int_{\square_m} |\nabla u(x)|^2 \, dx \right)^{\frac{1}{2}} + C |\square_m|^{\frac{1}{2}} \sum_{n=0}^{m-1} 3^{n(1+d/2)} \left(\sum_{y \in \mathcal{Z}_n} \left| (\nabla u)_{y+\square_n} \right|^2 \right)^{\frac{1}{2}}.$$

Using (6.4) and rearranging the inequality yields the desired conclusion after taking the supremum over all such η.

Step 2. The estimate for $\|u\|_{\underline{L}^2(\square_m)}$. By the representation theorem, there exists $\phi \in L^2(\square_m)$ such that $\fint_{\square_m} |\phi(x)|^2 \, dx = 1$ and

$$(6.5) \quad \left(\fint_{\square_m} |u(x)|^2 \, dx \right)^{\frac{1}{2}} = \fint_{\square_m} u(x) \phi(x) \, dx = \fint_{\square_m} u(x) \left(\phi(x) - (\phi)_{\square_m} \right) dx.$$

Denote by $w \in H^2(\square_m)$ the unique (up to additive constants) solution of the Neumann problem

$$\begin{cases} -\Delta w = \phi - (\phi)_{\square_m} & \text{in } \square_m, \\ \partial_\nu w = 0 & \text{on } \partial \square_m. \end{cases}$$

We have that $w \in H^2(\square_m)$ and

$$\fint_{\square_m} \left| \nabla^2 w(x) \right|^2 dx \leq C \fint_{\square_m} \left| \phi(x) - (\phi)_{\square_m} \right|^2 dx \leq C.$$

Testing the equation for w with u yields that

$$\fint_{\square_m} u(x) \left(\phi(x) - (\phi)_{\square_m} \right) dx = \fint_{\square_m} \nabla u(x) \cdot \nabla w(x) \, dx.$$

The conclusion now follows from the previous step and (6.5). $\qquad \square$

6.2. **Proof of Theorem 1.2.** Throughout this subsection, we let

$$v(x, U, p) := -u(x, U, p, 0) - p \cdot x.$$

That is, $v(\cdot, U, p) \in H_0^1(U)$ is the minimizer in the definition of $\nu(U, p)$ with the plane $x \mapsto p \cdot x$ subtracted.

The multiscale Poincaré inequality motivates us to prove the sublinearity of the correctors by studying the spatial averages of its gradient in mesoscopic cubes. This is accomplished by a very simple energy comparison argument combined with the Lipschitz estimate, which reduce the needed estimates to the convergence of the subadditive quantities. The statements we need are given in the following two lemmas.

In the first lemma, we compare $\nabla v(\cdot, \square_m, p)$ in the large cube $\square_m$ to the gradient of the function obtained by gluing together the functions $v(\cdot, z + \square_m, p)$ on the mesoscopic subgrid $\{z + \square_n : z \in \mathcal{Z}_n\}$ and bound the difference in terms of the cell problem energies in these cubes.

Lemma 6.2. *For every $p \in \mathbb{R}^d$ and $m, n \in \mathbb{N}$ with $m \geq n$,*

$$(6.6) \quad |\mathcal{Z}_n|^{-1} \sum_{z \in \mathcal{Z}_n} \left\| \nabla v(\cdot, \square_m, p) - \nabla v(\cdot, z + \square_n, p) \right\|_{\underline{L}^2(z + \square_n)}^2$$

$$\leq -\nu(\square_m, p) + |\mathcal{Z}_n|^{-1} \sum_{z \in \mathcal{Z}_n} \nu(z + \square_n, p),$$

where $\mathcal{Z}_n := 3^n \mathbb{Z}^d \cap \square_m$.

Proof. The lemma is a simple consequence of the uniform convexity of the energy functional and a proof can be found in [4, Lemma 2.1]. For the reader's convenience, we also provide the argument here. Define $V \in H_0^1(\square_m)$ to be the function obtained by gluing together the functions $v(\cdot, z + \square_n, p)$ for $z \in 3^n \mathbb{Z}^d \cap \square_m$. In other words, $V \equiv v(\cdot, z + \square_n, p)$ in $z + \square_n$, for each $z \in 3^n \mathbb{Z}^d \cap \square_m$. Since $x \mapsto p \cdot x + v(x, \square_m, p)$ is a solution of (1.1), we have

$$\langle \mathbf{a}(p + \nabla v(\cdot, \square_m, p)), \nabla V \rangle_{\square_m} = 0.$$

Using this and a direct computation, we find that

$$\frac{1}{2} \left\| \nabla v(\cdot, \square_m, p) - \nabla V \right\|_{\underline{L}^2(\square_m)}^2$$

$$\leq \frac{1}{2} \langle \nabla v(\cdot, \square_m, p) - \nabla V, \mathbf{a}(\nabla v(\cdot, \square_m, p) - \nabla V) \rangle_{\square_m}$$

$$= \frac{1}{2} \langle p + \nabla V, \mathbf{a}(p + \nabla V) \rangle_{\square_m} - \frac{1}{2} \langle p + \nabla v(x, \square_m, p), \mathbf{a}(p + \nabla v(x, \square_m, p)) \rangle_{\square_m}$$

$$= \left(|\mathcal{Z}_n|^{-1} \sum_{z \in \mathcal{Z}_n} \nu(z + \square_n, p) \right) - \nu(\square_m, p).$$

This completes the proof. $\square$

We now use the Lipschitz estimate to upgrade the previous estimate. This will give us uniform local control as we blow up the macroscopic cube $\square_m$ to the whole space and thereby yield information on the stationary correctors.

Lemma 6.3. *Let $\mathcal{X}$ be as in the statement of Theorem 2.1. There exists $C(d, \Lambda) < \infty$ such that, for every $p \in \mathbb{R}^d$, $m, n, k \in \mathbb{N}$ with $m \geq n \geq k$ and $\mathcal{X} \leq 3^k$, we have the estimate*

$$(6.7) \quad \|\nabla v(\cdot, \Box_m, p) - \nabla v(\cdot, \Box_n, p)\|_{\underline{L}^2(\Box_k)}$$

$$\leq C \sum_{l=n}^{m-1} \left(-\nu(\Box_{l+1}, p) + 3^{-d} \sum_{z \in 3^l \mathbb{Z}^d \cap \Box_{l+1}} \nu(z + \Box_l, p) \right)^{\frac{1}{2}}.$$

Proof. Fix $k \in \mathbb{N}$ with $\mathcal{X} \leq 3^k$. By Theorem 2.1 (here we just use the Lipschitz estimate) and the previous lemma, for every $l \in \mathbb{N}$, $l \geq k$, we have

$$\|\nabla v(\cdot, \Box_{l+1}, p) - \nabla v(\cdot, \Box_l, p)\|_{\underline{L}^2(\Box_k)}^2$$

$$\leq C \|\nabla v(\cdot, \Box_{l+1}, p) - \nabla v(\cdot, \Box_l, p)\|_{\underline{L}^2(\Box_l)}^2$$

$$\leq C \left(-\nu(\Box_{l+1}, p) + 3^{-d} \sum_{z \in 3^l \mathbb{Z}^d \cap \Box_{l+1}} \nu(z + \Box_l, p) \right).$$

Note that the first inequality in the display above is obvious if $l = k$, and is a consequence of Theorem 2.1 for $l \geq k+1$. Taking the square root, summing over $l = n, \ldots, m-1$ and using the triangle inequality yields (6.7). $\qquad\square$

We next combine the previous two inequalities with the multiscale Poincaré inequality to obtain estimates on the corrector itself in a large macroscopic cube terms of ν in triadic subcubes.

Lemma 6.4. *Let $\mathcal{X}$ be as in the statement of Theorem 2.1. Then, for every $m \in \mathbb{N}$ with $3^m \geq \mathcal{X}$,*

$$(6.8) \quad \left\| \Phi(\cdot, p) - (\Phi(\cdot, p))_{\Box_m} \right\|_{\underline{L}^2(\Box_m)} + \left\| \nabla \Phi(\cdot, p) \right\|_{\underline{H}^{-1}(\Box_m)}$$

$$\leq C|p| + C \sum_{n=0}^{m-1} 3^n \left(-\nu(\Box_m, p) + |\mathcal{Z}_n|^{-1} \sum_{z \in \mathcal{Z}_n} \nu(z + \Box_n, p) \right)^{\frac{1}{2}}$$

$$+ C3^m \sum_{l=m}^{\infty} \left(-\nu(\Box_{l+1}, p) + 3^{-d} \sum_{z \in 3^l \mathbb{Z}^d \cap \Box_{l+1}} \nu(z + \Box_l, p) \right)^{\frac{1}{2}}.$$

Proof. Letting $m \to \infty$ in (6.7) yields that, for every $m \in \mathbb{N}$ with $3^m \geq \mathcal{X}$,

$$(6.9) \quad \|\nabla \Phi(\cdot, p) - \nabla v(\cdot, \Box_m, p)\|_{\underline{L}^2(\Box_m)}$$

$$\leq C \sum_{l=m}^{\infty} \left(-\nu(\Box_{l+1}, p) + 3^{-d} \sum_{z \in 3^l \mathbb{Z}^d \cap \Box_{l+1}} \nu(z + \Box_l, p) \right)^{\frac{1}{2}}.$$

For $n \in \mathbb{N}$ with $n \le m$, we use the fact that $v(\cdot, z + \square_n, p) \in H_0^1(z + \square_n)$ and use integration by parts, the Hölder inequality and Lemma 6.2 to get

$$
|\mathcal{Z}_n|^{-1} \sum_{y \in \mathcal{Z}_n} \left| (\nabla v(\cdot, \square_m, p))_{y + \square_n} \right|^2
$$
$$
\le |\mathcal{Z}_n|^{-1} \sum_{z \in \mathcal{Z}_n} \| \nabla v(\cdot, \square_m, p) - \nabla v(\cdot, z + \square_n, p) \|_{\underline{L}^2(z + \square_n)}^2
$$
$$
\le C \left(-\nu(\square_m, p) + |\mathcal{Z}_n|^{-1} \sum_{z \in \mathcal{Z}_n} \nu(z + \square_n, p) \right).
$$

Applying Proposition 6.1, we obtain

$$
\| v(\cdot, \square_m, p) \|_{\underline{L}^2(\square_m)} + \| \nabla v(\cdot, \square_m, p) \|_{\underline{H}^{-1}(\square_m)}
$$
$$
\le C \| \nabla v(\cdot, \square_m, p) \|_{\underline{L}^2(\square_m)} + C \sum_{n=0}^{m-1} 3^n \left(|\mathcal{Z}_n|^{-1} \sum_{y \in \mathcal{Z}_n} \left| (\nabla v(\cdot, \square_m, p))_{y + \square_n} \right|^2 \right)^{\frac{1}{2}}
$$
$$
\le C|p| + C \sum_{n=0}^{m-1} 3^n \left(-\nu(\square_m, p) + |\mathcal{Z}_n|^{-1} \sum_{z \in \mathcal{Z}_n} \nu(z + \square_n, p) \right)^{\frac{1}{2}}.
$$

By the previous display, (6.9), the triangle inequality and the fact that the $\underline{L}^2$ norm is stronger than the $\underline{H}^{-1}$ norm, we get

$$
\| \nabla \Phi(\cdot, p) \|_{\underline{H}^{-1}(\square_m)} \le \| \nabla v(\cdot, \square_m, p) \|_{\underline{H}^{-1}(\square_m)} + \| \nabla \Phi(\cdot, p) - \nabla v(\cdot, \square_m, p) \|_{\underline{H}^{-1}(\square_m)}
$$
$$
\le \| \nabla v(\cdot, \square_m, p) \|_{\underline{H}^{-1}(\square_m)} + C \| \nabla \Phi(\cdot, p) - \nabla v(\cdot, \square_m, p) \|_{\underline{L}^2(\square_m)}
$$
$$
\le C|p| + C \sum_{n=0}^{m-1} 3^n \left(-\nu(\square_m, p) + |\mathcal{Z}_n|^{-1} \sum_{z \in \mathcal{Z}_n} \nu(z + \square_n, p) \right)^{\frac{1}{2}}
$$
$$
+ C3^m \sum_{l=m}^{\infty} \left(-\nu(\square_{l+1}, p) + 3^{-d} \sum_{z \in 3^l \mathbb{Z}^d \cap \square_{l+1}} \nu(z + \square_l, p) \right)^{\frac{1}{2}}.
$$

This is the desired estimate for $\| \nabla \Phi(\cdot, p) \|_{\underline{H}^{-1}(\square_m)}$. To complete the proof of (6.8) and obtain the estimate for $\| \Phi(x, p) - (\Phi(\cdot, p))_{\square_m} \|_{\underline{L}^2(\square_m)}$, we argue similarly, but there is an extra step in which we use the Poincaré inequality:

$$
\| \Phi(x, p) - (\Phi(\cdot, p))_{\square_m} \|_{\underline{L}^2(\square_m)}
$$
$$
\le 2 \| v(\cdot, \square_m, p) \|_{\underline{L}^2(\square_m)} + \| \Phi(x, p) - v(\cdot, \square_m, p) - (\Phi(\cdot, p) - v(\cdot, \square_m, p))_{\square_m} \|_{\underline{L}^2(\square_m)}
$$
$$
\le C \| v(\cdot, \square_m, p) \|_{\underline{L}^2(\square_m)} + C3^m \left(\fint_{\square_m} |\nabla \Phi(x, p) - \nabla v(x, \square_m, p)|^2 \, dx \right)^{\frac{1}{2}}
$$
$$
\le C|p| + C \sum_{n=0}^{m-1} 3^n \left(-\nu(\square_m, p) + |\mathcal{Z}_n|^{-1} \sum_{z \in \mathcal{Z}_n} \nu(z + \square_n, p) \right)^{\frac{1}{2}}
$$
$$
+ C3^m \sum_{l=m}^{\infty} \left(-\nu(\square_{l+1}, p) + 3^{-d} \sum_{z \in 3^l \mathbb{Z}^d \cap \square_{l+1}} \nu(z + \square_l, p) \right)^{\frac{1}{2}}. \qquad \square
$$

We are now ready to complete the proof of Theorem 1.2.

Proof of Theorem 1.2. By standard comparisons, it suffices to establish the result with B_R replaced by $\square_m$, $m \in \mathbb{N}$. We may also assume $|p| \le 1$ by homogeneity. Fix $\beta \in \left(0, \frac{1}{2}\right)$ and define, for each $n \in \mathbb{N}$,

$$
s_n := \begin{cases}
3^{-(m-n)} \left(-\nu(\square_m, p) + |\mathcal{Z}_n|^{-1} \displaystyle\sum_{z \in \mathcal{Z}_n} \nu(z + \square_n, p) \right)^{\frac{1}{2}} & \text{if } n \le m-1, \\[4ex]
\left(-\nu(\square_{n+1}, p) + 3^{-d} \displaystyle\sum_{z \in 3^n \mathbb{Z}^d \cap \square_{n+1}} \nu(z + \square_n, p) \right)^{\frac{1}{2}} & \text{if } n \ge m.
\end{cases}
$$

By Lemma 6.4, the theorem is proved if we can show that

$$
\log \mathbb{E}\left[\exp\left(\lambda 3^{2m\beta} \left(\sum_{n=0}^{\infty} s_n \right)^2 \right) \right] \le C(1 + \lambda^2).
$$

We let $\beta' \in (\beta, 1/2)$, and apply Jensen's inequality with respect to the measure $\sum_{n=0}^{\infty} 3^{-n(\beta'-\beta)} \delta_n$ and the convex function $x \mapsto \exp(x^2)$ to get

$$
\exp\left(\lambda 3^{2m\beta} \left(\sum_{n=0}^{\infty} s_n \right)^2 \right) \le C \sum_{n=0}^{\infty} 3^{-n(\beta'-\beta)} \exp\left(\lambda 3^{2m\beta + 2n(\beta'-\beta)} s_n^2 \right).
$$

We analyse first the sum over $n \ge m$. By Theorem 1.1 with $\alpha = 2\beta'/d$,

$$
\sum_{n=m}^{\infty} 3^{-n(\beta'-\beta)} \mathbb{E}\left[\exp\left(\lambda 3^{2m\beta + 2n(\beta'-\beta)} s_n^2 \right) \right] \le \sum_{n=m}^{\infty} 3^{-n(\beta'-\beta)} \exp\left[C\left(1 + 3^{-2\beta(n-m)} \lambda^2 \right) \right]
$$

$$
\le \exp\left[C(1 + \lambda^2) \right].
$$

For the sum over $n < m$, we use Jensen's inequality and Theorem 1.1 to get

$$
\log \mathbb{E}\left[\exp\left(\lambda 3^{2n\beta'} \left| \frac{1}{2} p \cdot \bar{\mathbf{a}} p - |\mathcal{Z}_n|^{-1} \sum_{z \in \mathcal{Z}_n} \nu(z + \square_n, p) \right| \right) \right] \le C(1 + \lambda^2).
$$

Therefore

$$
\sum_{n=0}^{m-1} 3^{-n(\beta'-\beta)} \mathbb{E}\left[\exp\left(\lambda 3^{2m\beta + 2n(\beta'-\beta)} s_n^2 \right) \right]
$$

$$
\le \sum_{n=0}^{m-1} 3^{-n(\beta'-\beta)} \exp\left[C\left(1 + 3^{-2(1-\beta)(m-n)} \lambda^2 \right) \right] \le \exp\left[C\left(1 + \lambda^2 \right) \right].
$$

This completes the proof. $\qquad\square$

REFERENCES

[1] V. Adolfsson and D. Jerison. L^p-integrability of the second order derivatives for the Neumann problem in convex domains. *Indiana Univ. Math. J.*, 43(4):1123–1138, 1994.

[2] S. N. Armstrong and J.-C. Mourrat. Lipschitz regularity for elliptic equations with random coefficients. *Arch. Ration. Mech. Anal.*, in press, arXiv:1411.3668.

[3] S. N. Armstrong and Z. Shen. Lipschitz estimates in almost-periodic homogenization. *Comm. Pure Appl. Math.*, in press, arXiv:1409.2094.

[4] S. N. Armstrong and C. K. Smart. Quantitative stochastic homogenization of convex integral functionals. *Ann. Sci. Éc. Norm. Supér.*, in press, arXiv:1406.0996.

[5] M. Avellaneda and F.-H. Lin. Compactness methods in the theory of homogenization. *Comm. Pure Appl. Math.*, 40(6):803–847, 1987.

[6] M. Avellaneda and F.-H. Lin. L^p bounds on singular integrals in homogenization. *Comm. Pure Appl. Math.*, 44(8-9):897–910, 1991.

[7] G. Dal Maso and L. Modica. Nonlinear stochastic homogenization. *Ann. Mat. Pura Appl. (4)*, 144:347–389, 1986.

[8] G. Dal Maso and L. Modica. Nonlinear stochastic homogenization and ergodic theory. *J. Reine Angew. Math.*, 368:28–42, 1986.

[9] J. Fischer and F. Otto. A higher-order large-scale regularity theory for random elliptic operators. Preprint, arXiv:1503.07578.

[10] A. Gloria, S. Neukamm, and F. Otto. An optimal quantitative two-scale expansion in stochastic homogenization of discrete elliptic equations. *ESAIM Math. Model. Numer. Anal.*, 48(2):325–346, 2014.

[11] 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):455–515, 2015.

[12] A. Gloria, S. Neukamm, and F. Otto. A regularity theory for random elliptic operators, Preprint, arXiv:1409.2678.

[13] A. Gloria and F. Otto. An optimal variance estimate in stochastic homogenization of discrete elliptic equations. *Ann. Probab.*, 39(3):779–856, 2011.

[14] A. Gloria and F. Otto. An optimal error estimate in stochastic homogenization of discrete elliptic equations. *Ann. Appl. Probab.*, 22(1):1–28, 2012.

[15] A. Gloria and F. Otto. Quantitative results on the corrector equation in stochastic homogenization. *J. Eur. Math. Soc.*, in press, arXiv:1409.0801.

[16] P. Grisvard. *Elliptic problems in nonsmooth domains*, volume 69 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2011.

[17] V. V. Jikov, S. M. Kozlov, and O. A. Oleĭnik. *Homogenization of differential operators and integral functionals*. Springer-Verlag, Berlin, 1994.

[18] S. M. Kozlov. Averaging of differential operators with almost periodic rapidly oscillating coefficients. *Mat. Sb. (N.S.)*, 107(149)(2):199–217, 317, 1978.

[19] D. Marahrens and F. Otto. Annealed estimates on the Green's function. *Probab. Theory Related Fields*, in press, arXiv:1304.4408.

[20] A Naddaf and T. Spencer. Estimates on the variance of some homogenization problems, 1998, Unpublished preprint.

[21] G. C. Papanicolaou and S. R. S. Varadhan. Boundary value problems with rapidly oscillating random coefficients. In *Random fields, Vol. I, II (Esztergom, 1979)*, volume 27 of *Colloq. Math. Soc. János Bolyai*, pages 835–873. North-Holland, Amsterdam, 1981.

[22] Z. Shen. Boundary estimates in elliptic homogenization. Preprint, arXiv:1505.00694.

[23] V. V. Yurinskiĭ. Averaging of symmetric diffusion in a random medium. *Sibirsk. Mat. Zh.*, 27(4):167–180, 215, 1986.

[24] V. V. Yurinskiĭ. Homogenization error estimates for random elliptic operators. In *Mathematics of random media (Blacksburg, VA, 1989)*, volume 27 of *Lectures in Appl. Math.*, pages 285–291. Amer. Math. Soc., Providence, RI, 1991.

(S. Armstrong) CEREMADE (UMR CNRS 7534), UNIVERSITÉ PARIS-DAUPHINE, PARIS, FRANCE
  E-mail address: `armstrong@ceremade.dauphine.fr`

(T. Kuusi) DEPARTMENT OF MATHEMATICS AND SYSTEMS ANALYSIS, AALTO UNIVERSITY, FINLAND
  E-mail address: `tuomo.kuusi@aalto.fi`

(J.-C. Mourrat) ECOLE NORMALE SUPÉRIEURE DE LYON, CNRS, LYON, FRANCE
  E-mail address: `jean-christophe.mourrat@ens-lyon.fr`