

ASYMPTOTIC HÖLDER REGULARITY FOR THE ELLIPSOID PROCESS

ÁNGEL ARROYO^{1,*} AND MIKKO PARVIAINEN²

Abstract. We obtain an asymptotic Hölder estimate for functions satisfying a dynamic programming principle arising from a so-called ellipsoid process. By the ellipsoid process we mean a generalization of the random walk where the next step in the process is taken inside a given space dependent ellipsoid. This stochastic process is related to elliptic equations in non-divergence form with bounded and measurable coefficients, and the regularity estimate is stable as the step size of the process converges to zero. The proof, which requires certain control on the distortion and the measure of the ellipsoids but not continuity assumption, is based on the coupling method.

Mathematics Subject Classification. 35B65, 35J15, 60H30, 60J10, 91A50.

Received October 15, 2019. Accepted May 20, 2020.

1. INTRODUCTION

1.1. Overview

The Krylov-Safonov [7] Hölder regularity result is one of the central results in the theory of non-divergence form elliptic partial differential equations with bounded and measurable coefficients. The result is not only important on its own right but also gives a flexible tool in the higher regularity and existence theory due to its rather weak assumptions on the coefficients.

In this paper, we study a quite general class of what we call ellipsoid processes. Ellipsoid processes are generalizations of the random walk where the next step in the process is taken inside a given space dependent ellipsoid E_x , and the value function u_ε satisfies the dynamic programming principle

$$u_\varepsilon(x) = \int_{E_x} u_\varepsilon(y) \, dy,$$

as explained in detail in the next section. Their role among the discrete processes is somewhat similar to the role of the linear uniformly elliptic partial differential equations with bounded and measurable coefficients among partial differential equations. Our main result, Theorem 1.1, establishes an asymptotic Hölder regularity for

Keywords and phrases: Dynamic programming principle, local Hölder estimates, stochastic games, coupling of stochastic processes, ellipsoid process, equations in non-divergence form.

¹ Department of Mathematics, University of Genoa, Via Dodecaneso 35, 16146 Genova, Italy.

² Department of Mathematics and Statistics, University of Jyväskylä, PO Box 35, 40014 Jyväskylä, Finland.

* Corresponding author: arroyo@dim.unige.it

a value function of an ellipsoid process under certain assumptions on the distortion and the measure of the ellipsoids but without any continuity assumption.

There is a classical well-known connection between the Brownian motion and the Laplace equation. It was observed in the paper of Peres, Schramm, Sheffield and Wilson [18] in discrete time that there is also a connection between the infinity Laplace equation and a two-player random turn zero-sum game called tug-of-war. Similarly, a connection exists between the p -Laplacian, $1 < p < \infty$ and different variants of tug-of-war game with noise [14, 17], as well as between the 1-Laplacian and a deterministic two-player game [6]. Our study of ellipsoid processes is partly motivated by the aim of developing a more general approach that would be widely applicable to problems including the tug-of-war games with noise. The mean value principle associated to the ellipsoids has also been discussed in the context of PDE theory for example by Pucci and Talenti in [20].

A more detailed overview of the proof is given in the next sections, but roughly speaking the proof is based on suitable couplings of probability measures related to the dynamic programming principle at different points and then look at the higher dimensional dynamic programming principle. The underlying idea is coming from the related stochastic processes: if we can show that with high probability the components of the coupled process coincide at some point, this will give a regularity estimate. In the usual random walk it is rather immediate that a good choice is a mirror point coupling of probability measures. However, since our ellipsoids can vary from point to point in a discontinuous fashion, finding good enough couplings becomes a nontrivial task. At the end of the paper, we illustrate the main points of the proof by explicit examples, and also counterexamples demonstrating the role of the assumptions.

The coupling approach to the regularity of different variants of tug-of-war with noise was first developed in [13] and applied in [1, 2, 4, 16]. As it turned out, for the continuous time diffusion processes and the Laplacian, the coupling method was utilized in connection to the regularity already at the beginning of 90's by Cranston [3], utilizing the tools developed in [12]. For more recent works, see for example [8, 9], which deal with linear PDEs under spatial continuity assumptions on the coefficients. Actually, in continuous time, the lack of regularity can have some fundamental consequences: Nadirashvili showed in [15] that there is not necessarily a unique diffusion, *i.e.* a unique solution to the martingale problem, nor is there necessarily a unique solution to the corresponding linear PDE with bounded and measurable coefficients. Some of the aspects in the method of coupling are similar to those of Ishii-Lions method [5] as pointed out in [19], but the methods seem to have developed independently.

Let us also point out that the global approach to regularity used for example in [10, 11, 14, 17] seems to be hard to adapt to our situation. This approach is based on coupling the same steps in different points so that the distance between the points is preserved, and continuing close to the boundary of the domain. Alternatively one can use the comparison with translated solutions. In both cases the translation invariance is used and this is not available for the ellipsoid process.

This paper is organized as follows. In Section 1.2, we make detailed statements. In Section 2 we review some basic notions. Section 3 presents the key lemmas in the coupling method. Section 4 is devoted to the proof of our main result in the case $|x - z| \gtrsim \varepsilon$, while Section 5 deals with the remaining case $|x - z| \lesssim \varepsilon$. Finally in Section 6 we give some examples showing that the assumptions are really needed in the method used here.

1.2. Statement of the main result

Given $0 < \lambda \leq \Lambda < \infty$, let us denote by $\mathcal{A}(\lambda, \Lambda)$ the set of all symmetric $n \times n$ real matrices A such that

$$\lambda|\xi|^2 \leq \xi^\top A \xi \leq \Lambda|\xi|^2$$

for every $\xi \in \mathbb{R}^n$. We say that the matrices in $\mathcal{A}(\lambda, \Lambda)$ are *uniformly elliptic*. We also refer to λ and Λ as the ellipticity constants of A .

It turns out that each symmetric and positive definite matrix A determines the shape and the orientation of an ellipsoid $E_A \subset \mathbb{R}^n$ centered at the origin and given by the formula

$$E_A := \{y \in \mathbb{R}^n : y^\top A^{-1} y < 1\} = A^{1/2} \mathbb{B},$$

where \mathbb{B} denotes the open unit ball in \mathbb{R}^n and $A^{1/2}$ stands for the *principal square root* of A . Moreover, the length of the principal semi-axes of E_A is determined by the square root of the eigenvalues of A (see Sect. 2 for more details), and thus the distortion of E_A coincides with the square root of the quotient between the largest and the smallest eigenvalue of A . In particular, for any $A \in \mathcal{A}(\lambda, \Lambda)$, the distortion of E_A is bounded from above by $\sqrt{\Lambda/\lambda}$. Motivated by this, we say that the quotient Λ/λ is the (maximum) *distortion* of the matrices in $\mathcal{A}(\lambda, \Lambda)$.

Let $\Omega \subset \mathbb{R}^n$ be a domain. We denote by $A : \Omega \rightarrow \mathcal{A}(\lambda, \Lambda)$ a matrix-valued function in Ω with measurable coefficients and values in $\mathcal{A}(\lambda, \Lambda)$. Given $\varepsilon > 0$, for each $x \in \Omega$ we can define an ellipsoid E_x centered at x by

$$E_x := E_{\varepsilon, A, x} := x + \varepsilon A(x)^{1/2} \mathbb{B}. \quad (1.1)$$

We also assume that the ellipsoids have the same measure, in other words, $\det\{A(x)^{1/2}\}$ is constant for every $x \in \Omega$. This does not seriously affect the applicability of our result as discussed at the end of Section 2.

For fixed $\varepsilon > 0$, in this paper we deal with solutions u_ε of the *dynamic programming principle (DPP)*

$$u_\varepsilon(x) = \int_{E_x} u_\varepsilon(y) \, dy \quad (1.2)$$

for $x \in \Omega$. That is, u_ε is a function whose value at a point $x \in \Omega$ coincides with its mean value over the ellipsoid E_x (with respect a uniform probability distribution).

Now we are in position of stating our main result, which asserts that, assuming a bound for the distortion of the ellipsoids, the solutions u_ε to (1.2) are asymptotically Hölder continuous.

Theorem 1.1. *Let $\Omega \subset \mathbb{R}^n$ be a domain, $n \geq 2$ and $0 < \lambda \leq \Lambda < \infty$ such that*

$$1 \leq \frac{\Lambda}{\lambda} < \frac{n+1}{n-1}. \quad (1.3)$$

Suppose that $A : \Omega \rightarrow \mathcal{A}(\lambda, \Lambda)$ is a measurable mapping with constant determinant and $B_{2r}(x_0) \subset \Omega$ for some $r > 0$. If u_ε is a solution to (1.2), then there exists some $\alpha = \alpha(n, \lambda, \Lambda) \in (0, 1)$ such that

$$|u_\varepsilon(x) - u_\varepsilon(z)| \leq C(|x - z|^\alpha + \varepsilon^\alpha),$$

holds for every $x, z \in B_r(x_0)$ and some constant $C > 0$ depending on $n, \lambda, \Lambda, r, \alpha$ and $\sup_{B_{2r}} |u|$, but independent of ε .

1.3. The ellipsoid process and heuristic idea of the proof

The *ellipsoid process* can be seen as a generalization of the random walk. Let us consider a sequence of points $\{x_0, x_1, x_2, \dots\} \subset \Omega$ describing the location of a particle at each time $j = 0, 1, 2, \dots$. Steps of the particle are decided according to the following rule: suppose that the particle is placed at some point $x_j \in \Omega$, then the next particle position x_{j+1} is chosen randomly in the ellipsoid E_{x_j} (according to a uniform probability distribution on E_{x_j}). This step is repeated until the particle exits Ω for the first time at some $x_\tau \notin \Omega$. Then the process stops and the amount $F(x_\tau)$ is collected, where F is a pay-off function defined outside Ω . The one step rule above defines a probability measure at every point. These probability measures induce a probability measure on

the space of sequences according to the Kolmogorov construction. Using this probability measure, the expected pay-off of the ellipsoid process starting from $x_0 \in \Omega$ is then

$$u_\varepsilon(x_0) := \mathbb{E}[F(x_\tau) | x_0].$$

Moreover, the process is a Markov chain and by the Markov property the expected pay-off at a point x coincides with the average of the expected pay-offs over the ellipsoid E_x , and thus it satisfies the DPP (1.2).

Given any solution $u_\varepsilon : \Omega \rightarrow \mathbb{R}$ to the original DPP (1.2), we perform a change of variables to rewrite the DPP as

$$u_\varepsilon(x) = \int_{\mathbb{B}} u_\varepsilon(x + \varepsilon A(x)^{1/2} y) dy. \quad (1.4)$$

Then, we construct a $2n$ -dimensional dynamic programming principle by defining $G : \Omega \times \Omega \rightarrow \mathbb{R}$ as

$$G(x, z) := u_\varepsilon(x) - u_\varepsilon(z)$$

and we get that

$$G(x, z) = \int_{\mathbb{B}} G(x + \varepsilon A(x)^{1/2} y, z + \varepsilon A(z)^{1/2} y) dy. \quad (1.5)$$

Now, by means of the function G , the problem of the regularity of u_ε becomes a question about the size of a solution G to (1.5) in $\Omega \times \Omega$.

Observe that, due to the invariance of the unit ball \mathbb{B} under orthogonal transformations, performing an orthogonal change of variables in (1.4) for $u_\varepsilon(z)$, we see that the $2n$ -dimensional DPP (1.5) is equivalent to

$$G(x, z) = \int_{\mathbb{B}} G(x + \varepsilon A(x)^{1/2} y, z + \varepsilon A(z)^{1/2} Q y) dy, \quad (1.6)$$

for any orthogonal matrix Q . Note that the value of the integral with G does not depend on Q . However, the choice of this matrix will become relevant in Section 4.2, where we estimate the right-hand side of (1.6) for a certain explicit comparison function f and show that it is a supersolution for (1.6) with a strict inequality. We will also show that the solution G satisfies comparison principle with the supersolution f . By the explicit structure of f , this implies the desired regularity result.

In stochastic terms, the above proof can be described by looking at a $2n$ -dimensional stochastic process induced by the ellipsoid process started at x_0 and z_0 , and by their probability measures coupled using the above coupling. The process continues until either one of the particles exits Ω for the first time (and then we impose the pay-off $2 \sup_{\Omega} |u_\varepsilon|$), or both particles occupy the same position (and we impose the pay-off 0). In this context, the key would be to show that the latter case occurs with a high enough probability.

Let us further illustrate the proof by taking as an example the particular case in which both E_x and E_z are unit balls centered at x and z , respectively. In this case the ellipsoid process reduces to a random walk. Then, a suitable choice for the orthogonal matrix Q , *i.e.* the coupling, would be the matrix describing the reflection with respect to the $(n - 1)$ -dimensional hyperplane orthogonal to $x - z$ and passing through $(x + z)/2$.

However, in the general case where we have two different ellipsoids E_x and E_z , it is not clear which couplings to choose. Actually, for ellipsoids with large distortion it may happen that there is no *good enough* coupling that lead us to obtain the sufficient estimates (see the examples in Sect. 6).

2. PRELIMINARIES AND NOTATION

In this work, $O(n)$ stands for the n -dimensional orthogonal group defined as

$$O(n) := \{P \in \mathbb{R}^{n \times n} : P^\top P = PP^\top = I\},$$

where P^\top stands for the transpose of P .

Note that, since $\lambda > 0$, then every matrix A in $\mathcal{A}(\lambda, \Lambda)$ is, in particular, positive definite. This together with the symmetry of A implies that there exists an orthogonal matrix $R \in O(n)$ and a diagonal matrix D (with real and positive elements) such that

$$A = RDR^\top.$$

We denote by $D^{1/2}$ the square root of D , obtained by taking the square root of each element in the diagonal of D . In consequence, we define the *principal square root* of a symmetric and positive semidefinite matrix A as

$$A^{1/2} := RD^{1/2}R^\top. \quad (2.1)$$

In view of (2.1), we see that the length of the principal semi-axes of the ellipsoid E_A is given by the square root of the eigenvalues of A (that is the diagonal entries of $D^{1/2}$), while the orientation of those semi-axes is given by the orthogonal matrix R . Hence, by uniform ellipticity, the inclusions

$$B_{\sqrt{\lambda}} \subset E_A \subset B_{\sqrt{\Lambda}} \quad (2.2)$$

hold uniformly for every $A \in \mathcal{A}(\lambda, \Lambda)$. Moreover, the measure of the ellipsoid E_A can be easily computed as $|E_A| = |\mathbb{B}| \sqrt{\det\{A\}}$.

As we have defined in (1.1), let us fix $\varepsilon > 0$ and $A : \Omega \rightarrow \mathcal{A}(\lambda, \Lambda)$, and let us explain the connection between the ellipsoids E_x , the DPP (1.2) and the elliptic equation in non-divergence form

$$\text{Tr}\{A(x)D^2u(x)\} = \sum_{i,j=1}^n a_{ij}(x) u_{x_i x_j}(x) = 0, \quad (2.3)$$

where the coefficients a_{ij} are the entries of $A(x)$. We start by recalling the second order Taylor's expansion of a twice differentiable function u at $x \in \Omega$,

$$u(x+y) = u(x) + \nabla u(x)^\top y + \frac{1}{2} y^\top D^2 u(x) y + o(|y|^2),$$

for every small enough $y \in \mathbb{R}^n$. Here we have used the notation $v^\top w$ for the scalar product of two (column) vectors $v, w \in \mathbb{R}^n$. Next, we compute the average with respect to y over the ellipsoid $\varepsilon A(x)^{1/2} \mathbb{B}$ of the expansion above. Then the first order term vanishes by the symmetry of $\varepsilon A(x)^{1/2} \mathbb{B}$, that is

$$\int_{\varepsilon A(x)^{1/2} \mathbb{B}} \nabla u(x)^\top y \, dy = \nabla u(x)^\top \left(\int_{\varepsilon A(x)^{1/2} \mathbb{B}} y \, dy \right) = 0,$$

while for the second order term we first rewrite it in trace form by recalling the equality $w^\top M v = \text{Tr}\{M v w^\top\}$. Then we need to compute the averaged integral over the ellipsoid of each entry in the matrix yy^\top ,

$$\begin{aligned} \int_{\varepsilon A(x)^{1/2} \mathbb{B}} yy^\top \, dy &= \int_{\mathbb{B}} (\varepsilon A(x)^{1/2} y) (\varepsilon A(x)^{1/2} y)^\top \, dy \\ &= \varepsilon^2 A(x)^{1/2} \left(\int_{\mathbb{B}} yy^\top \, dy \right) A(x)^{1/2} = \frac{\varepsilon^2}{n+2} A(x), \end{aligned}$$

where in the first equality a change of variables has been performed and in the last equality we have used that the averaged integral in parenthesis is equal to the identity matrix divided by $n+2$. Thus, the expansion for the average of u over E_x becomes

$$\int_{E_x} u(y) \, dy = u(x) + \frac{\varepsilon^2}{2(n+2)} \text{Tr}\{A(x) D^2 u(x)\} + o(\varepsilon^2).$$

Therefore, we get the following asymptotic mean value property related to (2.3): let $u \in C^2(\Omega)$, then

$$\text{Tr}\{A(x) D^2 u(x)\} = 0 \quad \iff \quad u(x) = \int_{E_x} u(y) \, dy + o(\varepsilon^2).$$

However, as already pointed out, the solutions to (2.3) with bounded and measurable coefficients, with no further regularity assumptions, are not necessarily unique. By a scaling argument we may assume that the determinant of A is a constant function on Ω , and thus the ellipsoids E_x defined in (1.1) have all the same measure.

3. THE COUPLING METHOD FOR ASYMPTOTIC REGULARITY

3.1. The comparison function

One of the keys in the proof of Theorem 1.1 is the construction of a suitable comparison function f that is a supvalue for the function G in the $2n$ -dimensional DPP (1.5), and possesses the desired regularity properties. We construct such a comparison function for the $2n$ -dimensional DPP (1.5) by defining explicitly the function $f_1 : \mathbb{R}^{2n} \rightarrow \mathbb{R}$ by

$$f_1(x, z) = C|x - z|^\alpha + \tilde{C}|x + z|^2, \tag{3.1}$$

where $\alpha \in (0, 1)$ is a small enough exponent depending on n, λ and Λ (see (4.16)),

$$\tilde{C} = \frac{2}{3r^2} \sup_{B_{2r}} |u|,$$

and

$$C > \frac{2}{r^\alpha} \sup_{B_{2r}} |u| \tag{3.2}$$

is a constant depending on $n, \lambda, \Lambda, r, \alpha$ and $\sup_{B_{2r}} |u|$ (see (4.17), (5.2) and (5.11)). Here and in what follows, we use a shorthand $u := u_\varepsilon$. Note that the first term in (3.1) plays the role of the modulus of α -Hölder continuity needed for obtaining the regularity result, while the second is just a small correction term introduced in order to ensure that

$$|u(x) - u(z)| \leq f_1(x, z) \quad \text{when } x, z \in B_{2r} \setminus B_r. \tag{3.3}$$

Indeed, if $x, z \in B_{2r} \setminus B_r$ such that $|x - z| \leq r$, then

$$|x + z|^2 = 2|x|^2 + 2|z|^2 - |x - z|^2 \geq 3r^2$$

and

$$|u(x) - u(z)| \leq 2 \sup_{B_{2r}} |u| = 3\tilde{C}r^2 \leq \tilde{C}|x + z|^2 \leq f_1(x, z).$$

Otherwise, if $|x - z| > r$, using (3.2) we get

$$|u(x) - u(z)| \leq 2 \sup_{B_{2r}} |u| < Cr^\alpha < C|x - z|^\alpha < f_1(x, z).$$

However, due to the discrete nature of the DPP, the solutions $u = u_\varepsilon$ to (1.2) can present jumps in the small ε -scale. For that reason, we need to introduce an additional step term in the comparison function in order to control the behavior of u in this situation. This is, an *annular step* function f_2 defined by

$$f_2(x, z) = \begin{cases} C^{2(2N-i)}\varepsilon^\alpha & \text{if } \frac{i-1}{2} < \frac{|x-z|}{\sqrt{\lambda}\varepsilon} \leq \frac{i}{2}, \quad (i = 0, 1, \dots, 2N), \\ 0 & \text{if } \frac{|x-z|}{\sqrt{\lambda}\varepsilon} > N \end{cases} \quad (3.4)$$

where $N \in \mathbb{N}$ is a large enough constant depending on λ, Λ and C that will be chosen later (see (4.7), (4.9) and Rem. 4.2). Note that $f_2 = 0$ whenever $|x - z| > N\sqrt{\lambda}\varepsilon$ and that $\sup f_2 = C^{4N}\varepsilon^\alpha$ is reached on the set

$$S_0 = \{(x, z) \in \mathbb{R}^{2n} : x = z\}.$$

3.2. The counter assumption

We choose

$$f(x, z) := f_1(x, z) - f_2(x, z),$$

as our comparison function. Since the terms in f_1 have been chosen so that (3.3) holds and $\sup f_2 = C^{4N}\varepsilon^\alpha$, then

$$|u(x) - u(z)| \leq f(x, z) + C^{4N}\varepsilon^\alpha \quad \text{when } x, z \in B_{2r} \setminus B_r. \quad (3.5)$$

In order to check that this estimate is also satisfied inside B_r , let us define a constant K measuring the maximum difference between $|u(x) - u(z)|$ and $f(x, z)$ for x and z in B_r , *i.e.*

$$K := \sup_{x, z \in B_r} (u(x) - u(z) - f(x, z)).$$

Our aim is to show that there exist suitable constants C and N so that the inequality (3.5) also holds in B_r , *i.e.*

$$|u(x) - u(z)| \leq f(x, z) + C^{4N}\varepsilon^\alpha \quad \text{when } x, z \in B_r. \quad (3.6)$$

We proceed by contradiction. Assume that (3.6) does not hold. Then, this implies that

$$K > C^{4N} \varepsilon^\alpha. \quad (3.7)$$

As a direct consequence of the counter assumption, observe that from the definition of K together with (3.5) it holds that

$$u(x) - u(z) \leq f(x, z) + K \quad \text{for every } x, z \in B_{2r}. \quad (3.8)$$

3.3. Statement of the key inequalities

Let us assume that the counter assumption (3.7) holds. It follows directly from the definition of K that, for any $\eta > 0$ to be fixed later, there exist $x_\eta, z_\eta \in B_r$ such that

$$u(x_\eta) - u(z_\eta) - f(x_\eta, z_\eta) \geq K - \eta. \quad (3.9)$$

In order to get a contradiction, we need to distinguish two different cases depending on the distance between the points x_η and z_η from (3.9). The case $|x_\eta - z_\eta| < \frac{1}{2} \sqrt{\lambda} \varepsilon$ is simpler due to a cancellation effect that happens when the distance is small. This special case is presented in Section 5.2. Now we focus on the case in which $|x_\eta - z_\eta| \geq \frac{1}{2} \sqrt{\lambda} \varepsilon$.

Our strategy is to utilize first the counter assumption together with (3.9) in order to obtain an inequality in terms of the comparison function f (see (3.11)). Then, by using the explicit form of the comparison function f , we choose adequately the constants C and N in such a way that f satisfies the opposite strict inequality (Lem. 3.1). This then is the desired contradiction giving the main result.

Starting from (3.9) and recalling the DPP (1.2), we get

$$\begin{aligned} K - \eta &\leq u(x_\eta) - u(z_\eta) - f(x_\eta, z_\eta) \\ &= \int_{\mathbb{B}} u(x_\eta + \varepsilon A(x_\eta)^{1/2} y) \, dy - \int_{\mathbb{B}} u(z_\eta + \varepsilon A(z_\eta)^{1/2} y) \, dy - f(x_\eta, z_\eta) \\ &= \int_{\mathbb{B}} \left[u(x_\eta + \varepsilon A(x_\eta)^{1/2} y) - u(z_\eta + \varepsilon A(z_\eta)^{1/2} Q y) \right] \, dy - f(x_\eta, z_\eta) \\ &\leq K + \int_{\mathbb{B}} f(x_\eta + \varepsilon A(x_\eta)^{1/2} y, z_\eta + \varepsilon A(z_\eta)^{1/2} Q y) \, dy - f(x_\eta, z_\eta), \end{aligned} \quad (3.10)$$

where in the second equality we have performed the orthogonal change of variables $y \mapsto Qy$ in the second integral, and in the last inequality we have recalled (3.8). We remark that the counter assumption has been applied here when using (3.8).

Thus for any $\eta > 0$, let $x_\eta, z_\eta \in B_r$ that satisfy (3.9), it follows that

$$f(x_\eta, z_\eta) \leq \int_{\mathbb{B}} f(x_\eta + \varepsilon A(x_\eta)^{1/2} y, z_\eta + \varepsilon A(z_\eta)^{1/2} Q y) \, dy + \eta, \quad (3.11)$$

where $Q \in O(n)$ is any fixed orthogonal matrix.

Now, the idea is to thrive for a contradiction by showing that the opposite strict inequality for (3.11) holds for a certain choice of the coupling matrix Q . As we will see in Section 4.2, this can only be done assuming a bound for the distortion of the matrices in $\mathcal{A}(\lambda, \Lambda)$.

Lemma 3.1. *Let $f = f_1 - f_2$ be the comparison function, where f_1 and f_2 are the functions defined in (3.1) and (3.4), respectively. Let $\eta = \eta(\varepsilon) = \varepsilon^2 > 0$ and assume that Λ/λ is bounded as in (1.3). Let $x, z \in B_r$ such*

that $|x - z| > \frac{1}{2}\sqrt{\lambda}\varepsilon$, then there exists an orthogonal matrix $Q \in O(n)$ such that

$$f(x, z) > \int_{\mathbb{B}} f(x + \varepsilon A(x)^{1/2}y, z + \varepsilon A(z)^{1/2}Qy) \, dy + \eta \quad (3.12)$$

holds.

Observe that, choosing x and z so that $x = x_\eta$ and $z = z_\eta$ in Lemma 3.1, we get a contradiction between (3.11) and (3.12), and this implies the falseness of (3.7), so (3.6) holds. This, together with the short distance case presented in Section 5.2, yields the desired asymptotic Hölder regularity estimate and the proof of Theorem 1.1 is complete.

We can assume without loss of generality that both x_η and z_η from (3.9) lie in the first coordinate axis of \mathbb{R}^n , i.e. $x_\eta, z_\eta \in \text{span}\{e_1\}$. Furthermore, from now on we will assume that

$$\frac{x_\eta - z_\eta}{|x_\eta - z_\eta|} = e_1,$$

where $e_1 = (1, 0, \dots, 0)^\top$.

We split the proof of Lemma 3.1 into two cases depending on the size of $|x - z|$ in comparison with the constant $N\sqrt{\lambda}\varepsilon$, using different arguments and distinguishing between the case in which $f_2 = 0$ (Sect. 4) and the case in which $f_2 \neq 0$ (Sect. 5).

4. CASE $|x_\eta - z_\eta| > N\sqrt{\lambda}\varepsilon$

In this section, we prove Lemma 3.1 in the case $|x - z| > N\sqrt{\lambda}\varepsilon$, and thus, as pointed out in Section 3.3, complete the proof of the main result Theorem 1.1 in this particular case.

Since $|x - z| > N\sqrt{\lambda}\varepsilon$, by (3.4) we have that $f_2(x, z) = 0$. Thus

$$f(x, z) = f_1(x, z) = C|x - z|^\alpha + \tilde{C}|x + z|^2. \quad (4.1)$$

Our aim is to show that the inequality (3.12) in Lemma 3.1 holds. The idea is to use the explicit form of the comparison function in order to estimate the right-hand side of (3.11), and then check that, for a convenient choice of the constants, the estimate is strictly bounded from above by $f(x, z)$.

4.1. Taylor's expansion for f

In this section we compute the Taylor's expansion of f_1 in order to obtain the desired estimates for the comparison function. This is similar to [1, 13] and [2]. However, for the convenience of the reader and expository reasons we write down the details.

Lemma 4.1. *Let f be the comparison function (4.1) and suppose that $x, z \in B_r$ satisfy $(x - z)/|x - z| = e_1$. Then the inequality*

$$\begin{aligned} & f(x + h_x, z + h_z) - f(x, z) \\ & \leq C \alpha |x - z|^{\alpha-1} e_1^\top (h_x - h_z) + 2\tilde{C}(x + z)^\top (h_x + h_z) \\ & \quad + \frac{C}{2} \alpha |x - z|^{\alpha-2} \text{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} (h_x - h_z)(h_x - h_z)^\top \right\} \\ & \quad + (16\tilde{C}\Lambda r^{2-\alpha} + 1) |x - z|^{\alpha-2} \varepsilon^2 \end{aligned} \quad (4.2)$$

holds for every $|h_x|, |h_z| \leq \sqrt{\Lambda}\varepsilon$.

Proof. First observe that,

$$\begin{aligned}
& |(x + h_x) + (z + h_z)|^2 - |x + z|^2 \\
&= 2(x + z)^\top (h_x + h_z) + |h_x + h_z|^2 \\
&\leq 2(x + z)^\top (h_x + h_z) + 4\Lambda \varepsilon^2 \\
&\leq 2(x + z)^\top (h_x + h_z) + 16\Lambda r^{2-\alpha} |x - z|^{\alpha-2} \varepsilon^2
\end{aligned} \tag{4.3}$$

for every $|h_x|, |h_z| < \sqrt{\Lambda} \varepsilon$, where in the last inequality we have used that $|x - z| \leq 2r$ and $\alpha - 2 < 0$. Next, let us recall the second order Taylor's expansion of a $2n$ -dimensional function $\phi(x, z)$,

$$\begin{aligned}
& \phi(x + h_x, z + h_z) - \phi(x, z) \\
&= D\phi(x, z)^\top \begin{pmatrix} h_x \\ h_z \end{pmatrix} + \frac{1}{2} \begin{pmatrix} h_x^\top & h_z^\top \end{pmatrix} D^2\phi(x, z) \begin{pmatrix} h_x \\ h_z \end{pmatrix} + \mathcal{E}_{x,z}(h_x, h_z),
\end{aligned}$$

where $\mathcal{E}_{x,z}(h_x, h_z)$ is the error term in the Taylor's expansion. Developing the terms in the expansion, we equivalently have that

$$\begin{aligned}
& \phi(x + h_x, z + h_z) - \phi(x, z) \\
&= D_x\phi(x, z)^\top h_x + D_z\phi(x, z)^\top h_z \\
&\quad + \frac{1}{2} \text{Tr} \{ D_{xx}\phi(x, z) h_x h_x^\top + D_{zz}\phi(x, z) h_z h_z^\top + 2D_{xz}\phi(x, z) h_z h_x^\top \} \\
&\quad + \mathcal{E}_{x,z}(h_x, h_z).
\end{aligned} \tag{4.4}$$

We use the formulas

$$D_x|x| = \frac{x}{|x|} \quad \text{and} \quad D_{xx}|x| = \frac{1}{|x|} \left(I - \frac{xx^\top}{|x|^2} \right),$$

which, since $(x - z)/|x - z| = e_1$ by assumption, give us that

$$D_x|x - z| = e_1 \quad \text{and} \quad D_{xx}|x - z| = -D_{xz}|x - z| = \frac{1}{|x - z|} (I - e_1 e_1^\top).$$

Thus, differentiating $|x - z|^\alpha$ with respect to x and z we get

$$D_x|x - z|^\alpha = \alpha|x - z|^{\alpha-1} e_1, \quad D_z|x - z|^\alpha = -\alpha|x - z|^{\alpha-1} e_1 \tag{4.5}$$

and

$$\begin{aligned}
D_{xx}|x - z|^\alpha &= D_{zz}|x - z|^\alpha = -D_{xz}|x - z|^\alpha \\
&= \alpha|x - z|^{\alpha-2} [-(1 - \alpha)e_1 e_1^\top + (I - e_1 e_1^\top)] \\
&= \alpha|x - z|^{\alpha-2} \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I_{n-1} \end{pmatrix}.
\end{aligned} \tag{4.6}$$

Next we estimate the error term. Since

$$\frac{\partial^3}{\partial t^3} t^\alpha = \alpha(1-\alpha)(2-\alpha)t^{\alpha-3},$$

by Taylor's theorem, whenever $|x-z| > 2\sqrt{\Lambda}\varepsilon$, then

$$\mathcal{E}_{x,z}(h_x, h_z) \leq \frac{\alpha(1-\alpha)(2-\alpha)}{6} \left| \begin{pmatrix} h_x \\ h_z \end{pmatrix} \right|^3 (|x-z| - 2\sqrt{\Lambda}\varepsilon)^{\alpha-3},$$

holds for every $|h_x|, |h_z| \leq \sqrt{\Lambda}\varepsilon$. Fix

$$N \geq 4\sqrt{\frac{\Lambda}{\lambda}}, \quad (4.7)$$

then by hypothesis $|x-z| > N\sqrt{\lambda}\varepsilon \geq 4\sqrt{\Lambda}\varepsilon$, and since $0 < \alpha < 1$, we can estimate

$$(|x-z| - 2\sqrt{\Lambda}\varepsilon)^{\alpha-3} \leq \left(\frac{|x-z|}{2} \right)^{\alpha-3}.$$

Inserting this in the estimate for the error term together with $|h_x|, |h_z| \leq \sqrt{\Lambda}\varepsilon$ we obtain

$$\begin{aligned} \mathcal{E}_{x,z}(h_x, h_z) &\leq \frac{1}{3} (2\Lambda\varepsilon^2)^{3/2} \left(\frac{|x-z|}{2} \right)^{\alpha-3} \\ &= \frac{2^{9/2-\alpha}}{3} \Lambda^{3/2} \frac{\varepsilon}{|x-z|} |x-z|^{\alpha-2} \varepsilon^2 \\ &< 2^{7/2} \Lambda^{3/2} \frac{\varepsilon}{|x-z|} |x-z|^{\alpha-2} \varepsilon^2 \\ &< \frac{2^{7/2}\Lambda}{N} \sqrt{\frac{\Lambda}{\lambda}} |x-z|^{\alpha-2} \varepsilon^2, \end{aligned} \quad (4.8)$$

where in the last inequality we have used $|x-z| > N\sqrt{\lambda}\varepsilon$ again. Plugging (4.5), (4.6) and (4.8) into the terms in (4.4) with $\phi(x, z) = |x-z|^\alpha$ we obtain

$$\begin{aligned} &|(x+h_x) - (z+h_z)|^\alpha - |x-z|^\alpha \\ &= \alpha|x-z|^{\alpha-1} e_1^\top (h_x - h_z) \\ &\quad + \frac{1}{2} \alpha|x-z|^{\alpha-2} \text{Tr} \left\{ \begin{pmatrix} \alpha-1 & 0 \\ 0 & I \end{pmatrix} (h_x - h_z)(h_x - h_z)^\top \right\} \\ &\quad + \frac{2^{7/2}\Lambda}{N} \sqrt{\frac{\Lambda}{\lambda}} |x-z|^{\alpha-2}, \end{aligned}$$

Thus combining this and (4.3) as in (4.1) and choosing a large enough natural number $N \in \mathbb{N}$ such that

$$N > 2^{7/2}\Lambda \sqrt{\frac{\Lambda}{\lambda}} C, \quad (4.9)$$

we get (4.2). \square

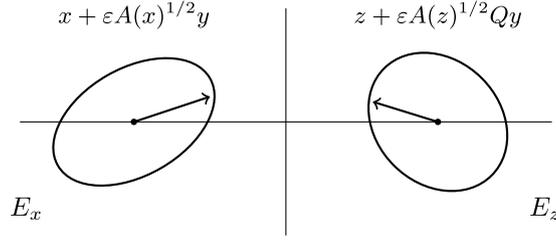


FIGURE 1. Illustration of the coupling aiming at negativity of the left-hand side in (4.11).

Remark 4.2. We could choose $N \in \mathbb{N}$ taking into account the fact that, as we will see later in Proposition 4.7, the distortion needs to be bounded by certain constant depending on n and α which is less than 3. Then, it is enough to choose $N \geq 4\sqrt{3}$ and $N \geq \sqrt{2^7} \cdot 3\Lambda C$ instead of (4.7) and (4.9), respectively.

The desired estimate then follows after averaging the inequality (4.2) from the previous Lemma.

Lemma 4.3. *Let f be the comparison function (4.1) and suppose that $x, z \in B_r$ satisfy $(x - z)/|x - z| = e_1$. Then the inequality*

$$\begin{aligned} & \int_{\mathbb{B}} f(x + \varepsilon A(x)^{1/2}y, z + \varepsilon A(z)^{1/2}Qy) \, dy - f(x, z) \\ & \leq |x - z|^{\alpha-2} \left[\frac{C\alpha}{2} T_\alpha(A(x), A(z), Q) + 16\tilde{C}\Lambda r^{2-\alpha} + 1 \right] \varepsilon^2 \end{aligned} \quad (4.10)$$

holds for any fixed orthogonal matrix $Q \in O(n)$, where

$$\begin{aligned} & T_\alpha(A_1, A_2, Q) \\ & = \int_{\mathbb{B}} \text{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} \left((A_1^{1/2} - A_2^{1/2}Q)y \right) \left((A_1^{1/2} - A_2^{1/2}Q)y \right)^\top \right\} \, dy. \end{aligned}$$

Proof. Replace $h_x = \varepsilon A(x)^{1/2}y$ and $h_z = \varepsilon A(z)^{1/2}Qy$ in (4.2) and note that, by symmetry, averaging over \mathbb{B} , the first order terms vanish and we get (4.10). \square

4.2. The optimal orthogonal coupling

In view of Lemma 4.3, we need to ensure first the negativity of the right-hand side of (4.10) in order to show that (3.12) holds. For that reason, in this subsection, our aim is to check that for every pair of matrices $A_1, A_2 \in \mathcal{A}(\lambda, \Lambda)$ there exists an orthogonal matrix $Q \in O(n)$ (depending on A_1, A_2 and α) so that

$$\int_{\mathbb{B}} \text{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} \left((A_1^{1/2} - A_2^{1/2}Q)y \right) \left((A_1^{1/2} - A_2^{1/2}Q)y \right)^\top \right\} \, dy < 0 \quad (4.11)$$

holds.

This is equivalent to the question of which is the best coupling matrix in (1.6). Heuristically, since $\alpha - 1 < 0$, the idea is to choose the orthogonal matrix Q such that the difference (in average) between $A_1^{1/2}y$ and $A_2^{1/2}Qy$ projected over the first component is much larger than projected over the orthogonal subspace $\{e_1\}^\perp$ (see Fig. 1).

We start by rewriting (4.11) in a more suitable form. By the linearity of the trace, we can integrate term by term inside the trace so we get

$$\begin{aligned}
& \int_{\mathbb{B}} \text{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} \left((A_1^{1/2} - A_2^{1/2}Q)y \right) \left((A_1^{1/2} - A_2^{1/2}Q)y \right)^\top \right\} dy \\
&= \text{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} (A_1^{1/2} - A_2^{1/2}Q) \left(\int_{\mathbb{B}} yy^\top dy \right) (A_1^{1/2} - A_2^{1/2}Q)^\top \right\} \\
&= \frac{1}{n+2} \text{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} (A_1^{1/2} - A_2^{1/2}Q)(A_1^{1/2} - A_2^{1/2}Q)^\top \right\} \\
&= \frac{1}{n+2} \text{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} (A_1 + A_2 - 2A_2^{1/2}QA_1^{1/2}) \right\},
\end{aligned} \tag{4.12}$$

where in the second equality we have used that

$$\int_{\mathbb{B}} yy^\top dy = \frac{1}{n+2} I.$$

Then, (4.11) holds if and only if

$$\text{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} (A_1 + A_2 - 2A_2^{1/2}QA_1^{1/2}) \right\} < 0.$$

In order to accomplish this, our strategy is to find a matrix in $O(n)$ minimizing the map

$$Q \mapsto \text{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} (A_1 + A_2 - 2A_2^{1/2}QA_1^{1/2}) \right\}, \tag{4.13}$$

where Q ranges the orthogonal group $O(n)$. For that reason, first we remark that the above map is continuous and defined on the compact space $O(n)$. Thus, there exists an orthogonal matrix, say Q_0 , depending on A_1 and A_2 such that

$$\text{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} A_2^{1/2}Q_0A_1^{1/2} \right\} \geq \text{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} A_2^{1/2}QA_1^{1/2} \right\}$$

for every $Q \in O(n)$. The matrix Q_0 represents the best possible coupling in (4.11). Then, the following step will be to impose the negativity of the resulting term as a sufficient condition for (3.12). As a consequence of this, it will be made clear the need of imposing a bound for the distortion of the matrices A_1 and A_2 .

We start by showing the following lemma, which is stated for any square matrix M and it is a result of optimization over the orthogonal group by using the singular value decomposition of matrices.

Lemma 4.4. *Let $n \geq 2$ and $M \in \mathbb{R}^{n \times n}$. Then $M^\top M$ is symmetric and positive semidefinite and satisfies*

$$\max_{Q \in O(n)} \text{Tr}\{MQ\} = \text{Tr}\{(M^\top M)^{1/2}\} \geq n |\det\{M\}|^{1/n}. \tag{4.14}$$

Proof. By the singular value decomposition, any square matrix M can be factorized as $M = R_1 S^{1/2} R_2^\top$, where $R_1, R_2 \in O(n)$ and S is a diagonal matrix containing the eigenvalues of $M^\top M$, which are real and non-negative since $M^\top M$ is symmetric and a positive semidefinite. Then, for any $Q \in O(n)$,

$$\text{Tr}\{MQ\} = \text{Tr}\{R_1 S^{1/2} R_2^\top Q\} = \text{Tr}\{S^{1/2} R_2^\top Q R_1\},$$

where $\text{Tr}\{AB\} = \text{Tr}\{BA\}$ has been used in the second equality. Since $O(n)$ together with the usual matrix multiplication has group structure, then we can select $\tilde{Q} = R_2^\top Q R_1 \in O(n)$ and thus

$$\max_{Q \in O(n)} \text{Tr}\{MQ\} = \max_{\tilde{Q} \in O(n)} \text{Tr}\{S^{1/2}\tilde{Q}\}.$$

Observe that, since $S^{1/2}$ is a diagonal matrix with non-negative entries, the maximum in the right-hand side of the previous equations is attained for $\tilde{Q} = I$, *i.e.*

$$\max_{Q \in O(n)} \text{Tr}\{S^{1/2}Q\} = \text{Tr}\{S^{1/2}\} = \text{Tr}\{(M^\top M)^{1/2}\}.$$

Therefore, we derive the equality in (4.14). In order to show the inequality in (4.14), let us write $\tilde{M} = (M^\top M)^{1/2}$ and recall the following inequality

$$\text{Tr}\{\tilde{M}\} \geq n(\det\{\tilde{M}\})^{1/n},$$

which follows from the inequality of arithmetic and geometric means and the fact that all the eigenvalues of \tilde{M} are real and non-negative. Then the proof is completed by observing that $\det\{\tilde{M}\} = |\det\{M\}|$. \square

Remark 4.5. Out of curiosity, observe that if $Q \in O(n)$ and $M \in \mathbb{R}^{n \times n}$, the distance between Q and M^\top is then given by

$$\begin{aligned} \|M^\top - Q\|^2 &= \text{Tr}\{(M^\top - Q)^\top(M^\top - Q)\} \\ &= \text{Tr}\{MM^\top + I - 2MQ\} \\ &= \|M\|^2 + n - 2\text{Tr}\{MQ\}. \end{aligned}$$

Therefore, the problem of finding the nearest orthogonal matrix to a given matrix M^\top is equivalent to the problem of maximizing $\text{Tr}\{MQ\}$ among all matrices Q in $O(n)$. Moreover, assuming that $\det\{M\} \neq 0$, the solution to this problem is attained at the orthogonal matrix $Q_0 = (M^\top M)^{-1/2}M^\top$, *i.e.* Q_0 is the nearest orthogonal matrix to M^\top and it satisfies

$$\text{Tr}\{MQ\} \leq \text{Tr}\{MQ_0\} = \text{Tr}\{(M^\top M)^{1/2}\}$$

for every $Q \in O(n)$.

Next we apply the previous lemma to obtain the following estimate.

Lemma 4.6. *Let $0 < \lambda \leq \Lambda < \infty$ and $\alpha \in (0, 1)$. Then*

$$\begin{aligned} \min_{Q \in O(n)} \text{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} (A_1 + A_2 - 2A_2^{1/2}QA_1^{1/2}) \right\} \\ \leq 2 \left[(n-1)\Lambda - ((1-\alpha) + n(1-\alpha)^{1/n})\lambda \right]. \end{aligned} \tag{4.15}$$

holds for every pair of matrices $A_1, A_2 \in \mathcal{A}(\lambda, \Lambda)$.

Proof. First notice that the first two terms in the trace of (4.15) do not depend on Q , so we can bound them directly by recalling the uniform ellipticity of the matrices in $\mathcal{A}(\lambda, \Lambda)$, i.e.,

$$\begin{aligned} \operatorname{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} (A_1 + A_2) \right\} &\leq 2 \max_{A \in \mathcal{A}(\lambda, \Lambda)} \operatorname{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} A \right\} \\ &= 2 \left[(n-1)\Lambda - (1-\alpha)\lambda \right]. \end{aligned}$$

for every pair of matrices $A_1, A_2 \in \mathcal{A}(\lambda, \Lambda)$. For the remaining term, we recall Lemma 4.4 with

$$M = A_1^{1/2} \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} A_2^{1/2}.$$

Then,

$$\begin{aligned} \max_{Q \in O(n)} \operatorname{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} A_2^{1/2} Q A_1^{1/2} \right\} &\geq n \left| \det \left\{ A_1^{1/2} \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} A_2^{1/2} \right\} \right|^{1/n} \\ &= n(1-\alpha)^{1/n} |\det\{A_1\} \det\{A_2\}|^{\frac{1}{2n}} \\ &\geq n(1-\alpha)^{1/n} \lambda, \end{aligned}$$

where in the second inequality we have used the uniform ellipticity to get that the determinant of each matrix is bounded from below by λ^n . Then (4.15) follows. \square

The following is the main result of this section and provides a sufficient condition for (3.12) whenever $|x - z| > N\sqrt{\lambda}\varepsilon$, concluding the proof of Lemma 3.1 in this case.

Proposition 4.7. *Let $n \geq 2$, $0 < \lambda \leq \Lambda < \infty$ and $\alpha \in (0, 1)$ such that*

$$1 \leq \frac{\Lambda}{\lambda} < \frac{n(1-\alpha)^{1/n} + (1-\alpha)}{n-1}. \quad (4.16)$$

For $x, z \in B_r$ such that $|x - z| > N\sqrt{\lambda}\varepsilon$, there exists a coupling matrix $Q_0 \in O(n)$ and a large enough constant $C > 0$ such that

$$\int_{\mathbb{B}} f(x + \varepsilon A(x)^{1/2} y, z + \varepsilon A(z)^{1/2} Q_0 y) \, dy < f(x, z) - \varepsilon^2,$$

where $f(x, z) = f_1(x, z) = C|x - z|^\alpha + \tilde{C}|x + z|^2$.

Proof. Select $Q_0 \in O(n)$ minimizing (4.13) with $A_1 = A(x)$, $A_2 = A(z)$. Then, recalling Lemma 4.3 together with (4.12) and Lemma 4.6, it turns out that

$$\begin{aligned} &\int_{\mathbb{B}} f(x + \varepsilon A(x)^{1/2} y, z + \varepsilon A(z)^{1/2} Q_0 y) \, dy - f(x, z) \\ &\leq |x - z|^{\alpha-2} \left(-\tau C + 16\tilde{C}\Lambda r^{2-\alpha} + 1 \right) \varepsilon^2, \end{aligned}$$

where

$$\tau = \tau(n, \lambda, \Lambda) := -\frac{\alpha}{n+2} \left[(n-1)\Lambda - ((1-\alpha) + n(1-\alpha)^{1/n})\lambda \right] > 0$$

by hypothesis. Finally, since $|x - z| < 2r$, choosing large enough

$$C > \frac{1}{\tau} (16\tilde{C}\Lambda r^{2-\alpha} + 4r^{2-\alpha} + 1) \quad (4.17)$$

we ensure that the right-hand side in the previous inequality is less than $-\varepsilon^2$ and the proof is finished. \square

Remark 4.8. Note that, as a consequence of Proposition 4.7, the condition (1.3) in the statement of Theorem 1.1 can be replaced by (4.16). Indeed, since the right-hand side of (4.16) is decreasing on α and

$$\frac{n(1-\alpha)^{1/n} + (1-\alpha)}{n-1} \xrightarrow{\alpha \rightarrow 0} \frac{n+1}{n-1},$$

it turns out that the condition (4.16) is weaker than (1.3). Therefore, this improves the bound for the distortion in Theorem 1.1 and it provides an explicit dependence between the dimension, the ellipticity constants and the Hölder exponent of the solutions.

Remark 4.9. It is worth noting that in the case of continuous coefficients (that is, when $A : \Omega \rightarrow \mathcal{A}(\lambda, \Lambda)$ is continuous) the bound for the distortion (1.3) can be dropped from the assumptions in Theorem 1.1. This is due to the fact that under the continuity assumption the difference between $A_1 = A(x)$ and $A_2 = A(z)$ is small for every x and z close enough. In other words, we can assume that A_1 and A_2 are almost the same matrix. Then the negativity in (4.11) can be deduced from the constant coefficient case $A_1 = A_2$, for which (4.11) holds independently of the ellipticity constants.

To see this, let us first recall that one of the key steps in the proof of the asymptotic Hölder estimate is the minimization of (4.13) among all orthogonal matrices $Q \in O(n)$, and then to deduce the conditions under which such minimum is negative. This condition is expressed through the formula (4.16) and is only relevant in the large distance case. In fact, in the medium and short distance cases we assume that $\Lambda/\lambda < 3$ just for convenience, but we could adapt the proof without this assumption. In order to replace the assumption (1.3) by the continuity of $A(\cdot)$, we first focus on the constant coefficients case, that is, when both A_1 and A_2 in (4.13) are equal to a given fixed matrix $A \in \mathcal{A}(\lambda, \Lambda)$. Then it is possible to show the following estimate

$$\min_{Q \in O(n)} \text{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} (2A - 2A^{1/2}QA^{1/2}) \right\} \leq -4(1-\alpha)\lambda, \quad (4.18)$$

which holds for every $A \in \mathcal{A}(\lambda, \Lambda)$ and $\alpha \in (0, 1)$. As a direct consequence of (4.18), it turns out that (4.11) holds in the case $A_1 = A_2$ independently of $0 < \lambda \leq \Lambda < \infty$ and $\alpha \in (0, 1)$. Moreover, by virtue of Remark 4.5, this minimum is attained for the nearest orthogonal matrix to

$$A^{1/2} \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} A^{1/2}.$$

Now, returning to the general case, if $A_1, A_2 \in \mathcal{A}(\lambda, \Lambda)$, we can rewrite (4.13) as follows,

$$\begin{aligned} & \text{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} (A_1 + A_2 - 2A_2^{1/2}QA_1^{1/2}) \right\} \\ &= \text{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} (A_1^{1/2} + A_2^{1/2} - 2A_2^{1/2}Q)(A_1^{1/2} - A_2^{1/2}) \right\} \\ & \quad + \text{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} (2A_2 - 2A_2^{1/2}QA_2^{1/2}) \right\}. \end{aligned}$$

Since $\alpha \in (0, 1)$, by the uniform ellipticity of A_1 and A_2 we can estimate the first term in the right-hand side by

$$4\sqrt{n\Lambda} \|A_1^{1/2} - A_2^{1/2}\|.$$

Thus, taking the minimum at both sides of the inequality and recalling (4.18) with $A = A_2$ we get

$$\begin{aligned} & \min_{Q \in O(n)} \operatorname{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} (A_1 + A_2 - 2A_2^{1/2}QA_1^{1/2}) \right\} \\ & \leq 4 \left[\sqrt{n\Lambda} \|A_1^{1/2} - A_2^{1/2}\| - (1 - \alpha)\lambda \right]. \end{aligned}$$

In consequence, the desired negativity is obtained for small enough $\|A_1^{1/2} - A_2^{1/2}\|$. In particular, imposing the condition

$$\|A_1^{1/2} - A_2^{1/2}\| = \operatorname{Tr} \{(A_1^{1/2} - A_2^{1/2})^2\} \leq \frac{(1 - \alpha)\lambda}{2\sqrt{n\Lambda}},$$

we ensure that

$$\min_{Q \in O(n)} \operatorname{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} (A_1 + A_2 - 2A_2^{1/2}QA_1^{1/2}) \right\} \leq -2(1 - \alpha)\lambda,$$

so (4.11) follows for every $\alpha \in (0, 1)$.

Finally, in the continuous coefficients case, we can always find a small enough $r > 0$ such that

$$\|A(x)^{1/2} - A(z)^{1/2}\| \leq \frac{(1 - \alpha)\lambda}{2\sqrt{n\Lambda}}$$

for every $x, z \in B_r$.

5. CASE $|x_\eta - z_\eta| \leq N\sqrt{\lambda}\varepsilon$

Remember that for fixed $\eta > 0$, we selected x_η and z_η so that (3.9) holds. In this section we deal with the case in which the distance between these two points is bounded by $N\sqrt{\lambda}\varepsilon$. The idea is to obtain a contradiction by using two separate arguments depending on how large the distance between x_η and z_η is. We call these the *medium* and the *short distance case*, and they happen whenever $|x_\eta - z_\eta| > \frac{1}{2}\sqrt{\lambda}\varepsilon$ and $|x_\eta - z_\eta| \leq \frac{1}{2}\sqrt{\lambda}\varepsilon$, respectively.

As we noted in Section 3.3, by using the DPP (1.2) and the counter assumption (3.7), the inequality (3.11) holds for x_η and z_η satisfying (3.9). Hence, for the medium distance case, the contradiction will follow from Lemma 3.1 that we prove below in Section 5.1.

On the other hand, since Lemma 3.1 is stated for $|x - z| > \frac{1}{2}\sqrt{\lambda}\varepsilon$, and in order to obtain a contradiction, for the short distance case we use a slightly different coupling. We address this case in Section 5.2.

In contrast to the previous section where $f \equiv f_1$, in this case the key term in our comparison function is the step annular function f_2 . Hence, in what follows, it will be enough to use the following rough estimate for the term f_1 in the comparison function.

Lemma 5.1. *Let f_1 be the function defined in (3.1) and $x, z \in B_r$. Then*

$$f_1(x + h_x, z + h_z) \leq f_1(x, z) + 3C\Lambda^{\alpha/2}\varepsilon^\alpha \quad (5.1)$$

for every $|h_x|, |h_z| < \sqrt{\Lambda}\varepsilon$.

Proof. First we use the concavity of $t \mapsto t^\alpha$ to estimate

$$|(x + h_x) - (z + h_z)|^\alpha - |x - z|^\alpha \leq |h_x - h_z|^\alpha \leq 2(\sqrt{\Lambda} \varepsilon)^\alpha$$

for every $|h_x|, |h_z| < \sqrt{\Lambda} \varepsilon$, where $\alpha \in (0, 1)$ has been used here. On the other hand, since $x, z \in B_r$, then

$$\begin{aligned} |(x + h_x) + (z + h_z)|^2 - |x + z|^2 &= 2(x + z)^\top (h_x + h_z) + |h_x + h_z|^2 \\ &\leq 8r\sqrt{\Lambda} \varepsilon + 4\Lambda \varepsilon^2 \\ &\leq 12r\sqrt{\Lambda} \varepsilon \\ &\leq 12r\Lambda^{\alpha/2} \varepsilon^\alpha, \end{aligned}$$

where we have recalled that $\sqrt{\Lambda} \varepsilon < \min\{1, r\}$ and $0 < \alpha < 1$. Thus, recalling (3.1) and choosing large enough

$$C > 12\tilde{C}r, \tag{5.2}$$

we get (5.1). \square

Before moving into details, observe that, since $n \geq 2$, in this section we can weaken the assumption (1.3) by imposing the following bounds for the distortion,

$$1 \leq \frac{\Lambda}{\lambda} \leq 3, \tag{5.3}$$

with no dependence on the dimension. Hence, recalling (2.2), the inclusions

$$\mathbb{B} \subset \frac{1}{\sqrt{\lambda}} A(x)^{1/2} \mathbb{B} \subset \sqrt{3} \mathbb{B} \tag{5.4}$$

hold uniformly for every $x \in \Omega$.

5.1. The medium distance case: $\frac{1}{2} \sqrt{\lambda} \varepsilon < |x_\eta - z_\eta| \leq N\sqrt{\lambda} \varepsilon$

In this section, we prove Lemma 3.1 in the case $\frac{1}{2} \sqrt{\lambda} \varepsilon < |x - z| \leq N\sqrt{\lambda} \varepsilon$.

Recalling definition of f_2 in (3.4), this function can be expressed as a finite disjoint sum

$$f_2(x, z) = \sum_{i=0}^{2N} C^{2(2N-i)} \varepsilon^\alpha \chi_{S_i}(x, z), \tag{5.5}$$

where

$$S_i := \left\{ (x, z) \in \mathbb{R}^{2n} : \frac{i-1}{2} < \frac{|x-z|}{\sqrt{\lambda} \varepsilon} \leq \frac{i}{2} \right\} \tag{5.6}$$

for each $i = 0, 1, \dots, 2N$ and $\chi_{S_i}(x, z)$ stands for the function which is equal to 1 whenever $(x, z) \in S_i$ and 0 otherwise. By the assumptions of this section, we can fix $j = 2, 3, \dots, 2N$ so that (x, z) belongs to S_j . The next is the main result of this section.

Lemma 5.2. *Let f_2 be the function defined in (3.4) and $x, z \in B_r$ such that $(x - z)/|x - z| = e_1$ and*

$$\frac{1}{2} \sqrt{\lambda} \varepsilon < |x - z| \leq N \sqrt{\lambda} \varepsilon.$$

Define the vectors $\nu_x = A(x)^{-1/2} e_1$ and $\nu_z = A(z)^{-1/2} e_1$ and fix an orthogonal matrix $Q \in O(n)$ such that

$$Q \frac{\nu_x}{|\nu_x|} = - \frac{\nu_z}{|\nu_z|}. \quad (5.7)$$

Then

$$\int_{\mathbb{B}} f_2(x + \varepsilon A(x)^{1/2} y, z + \varepsilon A(z)^{1/2} Q y) dy \geq \gamma C^2 f_2(x, z), \quad (5.8)$$

where $\gamma \in (0, 1)$ is a fixed constant depending only on n .

Proof. For the sake of simplicity, let us write $A_1 = A(x)$ and $A_2 = A(z)$. Similarly as in (5.5), given $A_1, A_2 \in \mathcal{A}(\lambda, \Lambda)$ and $Q \in O(n)$ we write

$$f_2(x + \varepsilon A_1^{1/2} y, z + \varepsilon A_2^{1/2} Q y) = \sum_{i=0}^{2N} C^{2(2N-i)} \varepsilon^\alpha \chi_{S_i^*}(y),$$

where

$$S_i^* := \{y \in \mathbb{B} : (x + \varepsilon A_1^{1/2} y, z + \varepsilon A_2^{1/2} Q y) \in S_i\},$$

for $i = 0, 1, \dots, 2N$.

Therefore, computing the averaged integral we get that

$$\begin{aligned} & \int_{\mathbb{B}} f_2(x + \varepsilon A_1^{1/2} y, z + \varepsilon A_2^{1/2} Q y) dy \\ &= \frac{1}{|\mathbb{B}|} \sum_{i=0}^{2N} C^{2(2N-i)} \varepsilon^\alpha |S_i^*| \geq \frac{1}{|\mathbb{B}|} \sum_{i=0}^{j-1} C^{2(2N-i)} \varepsilon^\alpha |S_i^*| \\ &\geq C^{2(2N-j+1)} \varepsilon^\alpha \frac{1}{|\mathbb{B}|} \sum_{i=0}^{j-1} |S_i^*| = C^2 f_2(x, z) \frac{1}{|\mathbb{B}|} \sum_{i=0}^{j-1} |S_i^*|, \end{aligned}$$

where in the first inequality we have used that $f_2 \geq 0$ to discard all terms in the sum with $i \geq j$ and in the second inequality we have used that $C^{-2i} \geq C^{-2(j-1)}$ for $i = 0, 1, \dots, j-1$. Therefore, it only remains to show that the sum in the right-hand side of the previous estimate has a lower bound depending only on n, λ and Λ . Recalling the definition of the sets S_i^* together with the fact that they are pairwise disjoint we obtain

$$\begin{aligned} \frac{1}{|\mathbb{B}|} \sum_{i=0}^{j-1} |S_i^*| &= \frac{1}{|\mathbb{B}|} \left| \left\{ y \in \mathbb{B} : |x - z + \varepsilon (A_1^{1/2} - A_2^{1/2} Q) y| \leq \frac{j-1}{2} \sqrt{\lambda} \varepsilon \right\} \right| \\ &= \frac{1}{|\mathbb{B}|} \left| \left\{ y \in \mathbb{B} : \left| \frac{|x-z|}{\sqrt{\lambda} \varepsilon} e_1 + (I - A_2^{1/2} Q A_1^{-1/2}) \frac{1}{\sqrt{\lambda}} A_1^{1/2} y \right| \leq \frac{j-1}{2} \right\} \right|, \end{aligned}$$

where the hypothesis $(x - z)/|x - z| = e_1$ has been used here. Note that, letting $w = \lambda^{-1/2}A_1^{1/2}y$ and recalling (5.4) we obtain that

$$\begin{aligned} & \frac{1}{|\mathbb{B}|} \sum_{i=0}^{j-1} |S_i^*| \\ &= \frac{1}{|\lambda^{-1/2}A_1^{1/2}\mathbb{B}|} \left| \left\{ w \in \lambda^{-1/2}A_1^{1/2}\mathbb{B} : \left| \frac{|x-z|}{\sqrt{\lambda}\varepsilon} e_1 + (I - A_2^{1/2}QA_1^{-1/2})w \right| \leq \frac{j-1}{2} \right\} \right|, \\ &\geq 3^{-n/2} \frac{1}{|\mathbb{B}|} \left| \left\{ w \in \mathbb{B} : \left| \frac{|x-z|}{\sqrt{\lambda}\varepsilon} e_1 + (I - A_2^{1/2}QA_1^{-1/2})w \right| \leq \frac{j-1}{2} \right\} \right|. \end{aligned}$$

Next, we claim that there exists small enough fixed constant $\varrho > 0$ so that the following inequality holds

$$\left| \frac{|x-z|}{\sqrt{\lambda}\varepsilon} e_1 + (I - A_2^{1/2}QA_1^{-1/2})w \right| \leq \frac{j-1}{2} \quad \text{for every } w \in B_\varrho(-\frac{1}{3}e_1). \quad (5.9)$$

Indeed, assuming (5.9) we have that

$$\frac{1}{|\mathbb{B}|} \left| \left\{ w \in \mathbb{B} : \left| \frac{|x-z|}{\sqrt{\lambda}\varepsilon} e_1 + (I - A_2^{1/2}QA_1^{-1/2})w \right| \leq \frac{j-1}{2} \right\} \right| \geq \varrho^n$$

and (5.8) follows from this and the previous estimates with $\gamma = 3^{-n/2}\varrho^n$.

To prove that there exists $\varrho > 0$ such that (5.9) holds, let Q be an orthogonal matrix satisfying (5.7), then

$$A_2^{1/2}QA_1^{-1/2}e_1 = -\frac{|\nu_x|}{|\nu_z|}e_1.$$

Observe that by (5.3) and the definition of ν_x and ν_z we have in particular that

$$\frac{1}{\sqrt{3}} \leq \frac{|\nu_x|}{|\nu_z|} \leq \sqrt{3}. \quad (5.10)$$

Now fix any $w = -\frac{1}{3}e_1 + \varrho\zeta$ with $\zeta \in \mathbb{B}$ and insert it in the left-hand side of (5.9) to get

$$\left| \frac{|x-z|}{\sqrt{\lambda}\varepsilon} e_1 + (I - A_2^{1/2}QA_1^{-1/2})w \right| \leq \left| \frac{|x-z|}{\sqrt{\lambda}\varepsilon} - \frac{1}{3} \left(1 + \frac{|\nu_x|}{|\nu_z|} \right) \right| + 3\varrho,$$

where the second term in the right-hand side of this inequality follows from the rough estimate

$$\left| \zeta - A_2^{1/2}QA_1^{-1/2}\zeta \right| \leq 1 + \sqrt{3} < 3,$$

which holds uniformly for every $A_1, A_2 \in \mathcal{A}(\lambda, 3\lambda)$ and $\zeta \in \mathbb{B}$. Next, recalling that $(x, z) \in S_j$ with $j \geq 2$, we use the condition in the definition of the set S_j in (5.6) together with (5.10) to see that

(i) if $\frac{|x-z|}{\sqrt{\lambda}\varepsilon} > \frac{1}{3} \left(1 + \frac{|\nu_x|}{|\nu_z|}\right)$, then

$$\begin{aligned} \left| \frac{|x-z|}{\sqrt{\lambda}\varepsilon} - \frac{1}{3} \left(1 + \frac{|\nu_x|}{|\nu_z|}\right) \right| - \frac{j-1}{2} &\leq \frac{j}{2} - \frac{1}{3} \left(1 + \frac{|\nu_x|}{|\nu_z|}\right) - \frac{j-1}{2} \\ &= \frac{1}{6} - \frac{1}{3} \cdot \frac{|\nu_x|}{|\nu_z|} \\ &\leq \frac{1}{6} - \frac{1}{3\sqrt{3}} < -\frac{1}{100}, \end{aligned}$$

(ii) otherwise, if $\frac{|x-z|}{\sqrt{\lambda}\varepsilon} \leq \frac{1}{3} \left(1 + \frac{|\nu_x|}{|\nu_z|}\right)$, then

$$\left| \frac{|x-z|}{\sqrt{\lambda}\varepsilon} - \frac{1}{3} \left(1 + \frac{|\nu_x|}{|\nu_z|}\right) \right| - \frac{j-1}{2} < \frac{1}{3} \left(1 + \frac{|\nu_x|}{|\nu_z|}\right) - \frac{j-1}{2} - \frac{j-1}{2} \leq \frac{4}{3} + \frac{1}{\sqrt{3}} - j.$$

Observe that the last term is strictly less than $2-j$, so the negativity of this term is ensured since $j \geq 2$.

Therefore, we have shown that if $(x, z) \in S_j$ for some $j \geq 2$, then

$$\left| \frac{|x-z|}{\sqrt{\lambda}\varepsilon} - \frac{1}{3} \left(1 + \frac{|\nu_x|}{|\nu_z|}\right) \right| < \frac{j-1}{2} - \frac{1}{100}.$$

Finally, choosing $\varrho = \frac{1}{300}$, the inequality (5.9) holds for every $w \in B_{1/300}(-\frac{1}{3}e_1)$. This completes the proof of the lemma. \square

Hence, since $f = f_1 - f_2$, combining Lemmas 5.1 and 5.2 we get

$$\begin{aligned} \int_{\mathbb{B}} f(x + \varepsilon A(x)^{1/2} y, z + \varepsilon A(z)^{1/2} Q y) dy - f(x, z) \\ \leq 3C\Lambda^{\alpha/2}\varepsilon^\alpha - (\gamma C^2 - 1)f_2(x, z). \end{aligned}$$

Now, by definition of f_2 , it turns out that $f_2(x, z) \geq \varepsilon^\alpha$ for every $|x-z| \leq N\sqrt{\lambda}\varepsilon$, and thus choosing large enough C such that

$$\gamma C^2 - 3C\Lambda^{\alpha/2} - 2 > 0, \tag{5.11}$$

we ensure that the right-hand side of the previous inequality is less than $-\varepsilon^\alpha < -\varepsilon^2$. Then (3.12) follows, so the proof of Lemma 3.1 is complete in the case $|x-z| \leq N\sqrt{\lambda}\varepsilon$.

5.2. The short distance case: $|x_\eta - z_\eta| \leq \frac{1}{2} \sqrt{\lambda}\varepsilon$

In contrast with the large distance case and the medium distance case addressed in Sections 4 and 5.1, respectively, where the contradiction was obtained by proving Lemma 3.1, in this section we thrive for a contraction in a slightly different way. The main difference is that, in this case, we take an advantage from the full cancellation effect that happens when the two points x_η and z_η are close enough.

Recall that $x_\eta, z_\eta \in B_r$ satisfy (3.9). Observe that, since

$$B_{\sqrt{\lambda}\varepsilon}(x) \subset E_x$$

by uniform ellipticity, if the distance between x_η and z_η is small enough, then the corresponding ellipsoids E_{x_η} and E_{z_η} have non-empty intersection. Indeed, since $0 < |x_\eta - z_\eta| \leq \frac{1}{2} \sqrt{\lambda} \varepsilon$ by hypothesis, it is easy to check that

$$B_{\sqrt{\lambda} \varepsilon / 4} \left(\frac{x_\eta + z_\eta}{2} \right) \subset E_{x_\eta} \cap E_{z_\eta},$$

and thus

$$\frac{|E_{x_\eta} \cap E_{z_\eta}|}{|E_{x_\eta}|} \geq \gamma(n, \lambda, \Lambda) := \left(\frac{1}{4} \sqrt{\frac{\lambda}{\Lambda}} \right)^n. \quad (5.12)$$

This means that, in the $2n$ -dimensional process described in Section 1.3, starting from (x_η, z_η) we can choose a point $y \in \mathbb{R}^n$ belonging to both E_{x_η} and E_{z_η} . This choice gives a full cancellation but we also need that the measure of the ellipsoids is the same, that is, $\det\{A(x_\eta)\} = \det\{A(z_\eta)\}$. In fact, this is the only place in the paper where such an assumption is needed.

Then we repeat the argument in (3.10) but with a different coupling taking advantage of the full cancellation where the ellipsoids overlap. Starting from (3.9) and using the DPP (1.2) for estimating the difference $u(x_\eta) - u(z_\eta)$ we get

$$\begin{aligned} f(x_\eta, z_\eta) - \eta &\leq u(x_\eta) - u(z_\eta) - K = \int_{E_{x_\eta}} u(\zeta) \, d\zeta - \int_{E_{z_\eta}} u(\xi) \, d\xi - K \\ &= \frac{1}{|E_{x_\eta}|} \left[\int_{E_{x_\eta} \setminus E_{z_\eta}} u(\zeta) \, d\zeta - \int_{E_{z_\eta} \setminus E_{x_\eta}} u(\xi) \, d\xi \right] - K \\ &= \frac{|E_{x_\eta} \setminus E_{z_\eta}|}{|E_{x_\eta}|} \int_{E_{x_\eta} \setminus E_{z_\eta}} \int_{E_{z_\eta} \setminus E_{x_\eta}} [u(\zeta) - u(\xi)] \, d\xi \, d\zeta - K \\ &\leq \frac{|E_{x_\eta} \setminus E_{z_\eta}|}{|E_{x_\eta}|} \left[K + \int_{E_{x_\eta} \setminus E_{z_\eta}} \int_{E_{z_\eta} \setminus E_{x_\eta}} f(\zeta, \xi) \, d\xi \, d\zeta \right] - K \\ &= \frac{|E_{x_\eta} \setminus E_{z_\eta}|}{|E_{x_\eta}|} \int_{E_{x_\eta} \setminus E_{z_\eta}} \int_{E_{z_\eta} \setminus E_{x_\eta}} f(\zeta, \xi) \, d\xi \, d\zeta - \frac{|E_{x_\eta} \cap E_{z_\eta}|}{|E_{x_\eta}|} K \\ &\leq \int_{E_{x_\eta} \setminus E_{z_\eta}} \int_{E_{z_\eta} \setminus E_{x_\eta}} f(\zeta, \xi) \, d\xi \, d\zeta - \gamma K, \end{aligned} \quad (5.13)$$

where (3.8) has been recalled in the second inequality and in the last inequality we have used that $|E_{x_\eta} \setminus E_{z_\eta}| < |E_{x_\eta}|$ and (5.12). Since $f = f_1 - f_2 \leq f_1$ and $E_x \subset B_{\sqrt{\lambda} \varepsilon}(x)$ by uniform ellipticity, we can estimate the integral part in the right-hand side above using the estimate for f_1 from Lemma 5.1:

$$\begin{aligned} \int_{E_{x_\eta} \setminus E_{z_\eta}} \int_{E_{z_\eta} \setminus E_{x_\eta}} f(\zeta, \xi) \, d\xi \, d\zeta &\leq \int_{E_{x_\eta} \setminus E_{z_\eta}} \int_{E_{z_\eta} \setminus E_{x_\eta}} f_1(\zeta, \xi) \, d\xi \, d\zeta \\ &\leq f_1(x_\eta, z_\eta) + 3C\Lambda^{\alpha/2} \varepsilon^\alpha. \end{aligned}$$

For the remaining term, recalling the definition of f_2 in (3.4) together with the fact that $0 < |x_\eta - z_\eta| \leq \frac{1}{2} \sqrt{\lambda} \varepsilon$, we have that $f_2(x_\eta, z_\eta) = C^{2(2N-1)} \varepsilon^\alpha$. On the other hand, combining this with the counter assumption (3.7) we get

$$K > C^{4N} \varepsilon^\alpha = C^2 f_2(x_\eta, z_\eta).$$

Hence, putting all these estimates together in (5.13) we have that

$$f(x_\eta, z_\eta) - \eta \leq f_1(x_\eta, z_\eta) + 3C\Lambda^{\alpha/2}\varepsilon^\alpha - \gamma C^2 f_2(x_\eta, z_\eta).$$

Let $C > \gamma^{-1/2}$. Since $f = f_1 - f_2$ and $f_2(x_\eta, z_\eta) \geq \varepsilon^\alpha$, then

$$(\gamma C^2 - 1)\varepsilon^\alpha \leq 3C\Lambda^{\alpha/2}\varepsilon^\alpha + \eta.$$

Finally, we obtain a contradiction with this inequality by letting $\eta = \varepsilon^\alpha$ and choosing large enough C so that

$$\gamma C^2 - 3C\Lambda^{\alpha/2} - 2 > 0.$$

Therefore, this implies the falseness of the counter assumption (3.7). Thus (3.6) holds and the proof of Theorem 1.1 is finished.

6. EXAMPLES

6.1. Diagonal case

As it has been proved in the previous sections, the inequality (4.16) provides a sufficient condition for (3.12), from which the asymptotic Hölder regularity of solutions u_ε to the DPP (1.2) follows. One of the key results for deriving this condition is Lemma 4.4, which states that, given a matrix $M \in \mathbb{R}^{n \times n}$, the following holds,

$$\max_{Q \in O(n)} \operatorname{Tr}\{MQ\} = \operatorname{Tr}\{(M^\top M)^{1/2}\} \geq n |\det\{M\}|^{1/n}.$$

Then the idea of the proof consisted basically in obtaining a lower estimate for the right-hand side in the previous formula (see the proof of Lem. 4.6). However, we could try to obtain the desired estimate for the left-hand side directly without using the inequality above. Therefore, using the equality $\max_{Q \in O(n)} \operatorname{Tr}\{MQ\} = \operatorname{Tr}\{(M^\top M)^{1/2}\}$ we obtain

$$\begin{aligned} & \min_{Q \in O(n)} \operatorname{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} (A_1 + A_2 - 2A_2^{1/2}QA_1^{1/2}) \right\} \\ &= \operatorname{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} (A_1 + A_2) - 2 \left[\begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} A_1 \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} A_2 \right]^{1/2} \right\}, \end{aligned}$$

where $A_1, A_2 \in \mathcal{A}(\lambda, \Lambda)$ and $\alpha \in (0, 1)$. Then a stronger sufficient condition for (3.12) to be satisfied is that the matrix inequality

$$\begin{aligned} & \frac{1}{2} \operatorname{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} (A_1 + A_2) \right\} \\ & < \operatorname{Tr} \left\{ \left[\begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} A_1 \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} A_2 \right]^{1/2} \right\} \end{aligned} \quad (6.1)$$

holds for every pair of matrices $A_1, A_2 \in \mathcal{A}(\lambda, \Lambda)$. Inequality (6.1) can be seen as some sort of reverse inequality of arithmetic and geometric means for traces. Therefore, it seems reasonable to expect that the assumption on the bound of the distortion of the ellipsoids (4.16) can be weakened. For example, let us consider the case in which the axes of the ellipsoids are aligned with the coordinate axes. This is the case in which both A_1 and A_2 are diagonal matrices in $\mathcal{A}(\lambda, \Lambda)$. We denote by $\lambda_1(A_1), \lambda_2(A_1), \dots, \lambda_n(A_1)$ and $\lambda_1(A_2), \lambda_2(A_2), \dots, \lambda_n(A_2)$ the

elements in the diagonals of A_1 and A_2 , respectively. Recall that all these elements coincide with the eigenvalues of A_1 and A_2 and thus they are bounded between λ and Λ .

Lemma 6.1. *The inequality (6.1) holds for every diagonal matrices A_1 and A_2 in $\mathcal{A}(\lambda, \Lambda)$ if and only if*

$$1 \leq \frac{\Lambda}{\lambda} < \left(1 + 2\sqrt{\frac{1-\alpha}{n-1}}\right)^2. \quad (6.2)$$

Proof. Inserting $A_1 = \text{diag}(\lambda_1(A_1), \dots, \lambda_n(A_1))$ and $A_2 = \text{diag}(\lambda_1(A_2), \dots, \lambda_n(A_2))$ in (6.1) we get

$$\begin{aligned} & -(1-\alpha)(\lambda_1(A_1) + \lambda_1(A_2)) + \sum_{i=2}^n (\lambda_i(A_1) + \lambda_i(A_2)) \\ & < 2(1-\alpha)\sqrt{\lambda_1(A_1)\lambda_1(A_2)} + 2\sum_{i=2}^n \sqrt{\lambda_i(A_1)\lambda_i(A_2)}. \end{aligned}$$

Rearranging the terms we obtain

$$\sum_{i=2}^n (\sqrt{\lambda_i(A_1)} - \sqrt{\lambda_i(A_2)})^2 < (1-\alpha)(\sqrt{\lambda_1(A_1)} + \sqrt{\lambda_1(A_2)})^2. \quad (6.3)$$

Now observe that, by uniform ellipticity,

$$\sum_{i=2}^n (\sqrt{\lambda_i(A_1)} - \sqrt{\lambda_i(A_2)})^2 \leq (n-1)(\sqrt{\Lambda} - \sqrt{\lambda})^2$$

and

$$(1-\alpha)(\sqrt{\lambda_1(A_1)} + \sqrt{\lambda_1(A_2)})^2 \geq 4(1-\alpha)\lambda.$$

Moreover, the equality is attained in both inequalities for $A_1 = \lambda I$ and A_2 the diagonal matrix with entries $\lambda, \Lambda, \Lambda, \dots, \Lambda$. Thus, (6.3) holds for every diagonal matrices A_1 and A_2 in $\mathcal{A}(\lambda, \Lambda)$ if and only if

$$(n-1)(\sqrt{\Lambda} - \sqrt{\lambda})^2 < 4(1-\alpha)\lambda.$$

Hence, the bound in (6.2) follows. \square

Observe that the bound in (6.2) is larger than in (4.16). However, in this result we do not consider the whole family of ellipsoids, only the ones aligned with the coordinate axes, and hence we can not say that (6.2) is a sufficient condition for asymptotic Hölder regularity.

6.2. The mirror point coupling

This work is related to the results in [12] except that our setting is discrete, we do not require continuity for the ellipsoids, and our proof is written in purely analytic form *i.e.* we avoid the probabilistic tools. The Lindvall–Rogers coupling (see Eq. (2) in their paper), can be written without the drift as

$$\begin{aligned} dX_t &= \sigma(X_t) dB_t, \\ dX'_t &= \sigma(X'_t)H(X_t, X'_t) dB_t, \end{aligned}$$

where B is the Brownian motion. If we formally discretize this, we get

$$\begin{aligned} X_{k+1} &= X_k + \sigma(X_k)U_k, \\ X'_{k+1} &= X'_{k+1} + \sigma(X'_k)H(X_k, X'_k)U_k, \end{aligned}$$

where U_1, U_2, U_3, \dots is an i.i.d. sequence of uniform random variables on the ball B_ε . Above, $H(X_k, X'_k)$ is an orthogonal reflection matrix that can depend on the two coupled walks. Observe that our discrete results imply similar regularity results for the limit. Furthermore, the discrete results can also be interesting from the point of view of discretization of PDEs.

On the other hand, it is known that the so-called mirror point reflection coupling can be used to obtain Lipschitz estimates for functions satisfying the mean value property with balls. If x and z denote the centers of two balls, this coupling consists on reflecting the points of one of the balls into the other with respect to the hyperplane orthogonal to $z - x$ passing through $(x + z)/2$. Therefore, it is reasonable to expect that the same type of coupling should work if the balls are replaced by ellipsoids with sufficiently small distortion. In this subsection we adapt the mirror point coupling to our setting and we study under which conditions this coupling works. Moreover, we compare it with the optimal coupling arising from the minimization of (4.13).

Recalling Remark 4.5, the optimal coupling matrix Q for which the minimum of (4.13) is attained depends on A_1, A_2 and α , and can be explicitly written as follows,

$$Q = \left[\left(A_1^{1/2} \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} A_2^{1/2} \right)^2 \right]^{-1/2} A_1^{1/2} \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} A_2^{1/2}.$$

Observe that Q is indeed an orthogonal matrix with determinant -1 , so Q can be also seen as a reflection in the Euclidean n -dimensional space. For this choice of Q , (4.11) holds if the condition (4.16) is satisfied.

However, it is possible to fix Q without any dependence on the matrices or the exponent in such a way that (4.11) still holds under certain restriction on the distortion. More precisely, this choice corresponds to the reflection matrix on the first coordinate axis, that is

$$Q = J_0 := \begin{pmatrix} -1 & 0 \\ 0 & I \end{pmatrix}.$$

To see this, replacing $\begin{pmatrix} \alpha - 1 & 0 \\ 0 & 1 \end{pmatrix} = J_0 + \alpha e_1 e_1^\top$ we obtain

$$\begin{aligned} & \text{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} (A_1 + A_2 - 2A_2^{1/2} Q A_1^{1/2}) \right\} \\ &= \text{Tr} \{ J_0 (A_1 + A_2 - 2A_2^{1/2} J_0 A_1^{1/2}) \} + \alpha |(A_1^{1/2} - A_2^{1/2} J_0)^\top e_1|^2. \end{aligned}$$

By uniform ellipticity we can easily obtain a bound for the second term,

$$\alpha |(A_1^{1/2} - A_2^{1/2} J_0)^\top e_1|^2 \leq \alpha (|A_1^{1/2} e_1| + |A_2^{1/2} e_1|)^2 \leq 4\alpha\Lambda,$$

while for the other term, observing that $J_0 A_2^{1/2} J_0 \in \mathcal{A}(\sqrt{\lambda}, \sqrt{\Lambda})$, we get that

$$\begin{aligned} \text{Tr} \{ J_0 (A_1 + A_2 - 2A_2^{1/2} J_0 A_1^{1/2}) \} &= \text{Tr} \{ J_0 (A_1 + A_2) \} - 2 \text{Tr} \{ J_0 A_2^{1/2} J_0 A_1^{1/2} \} \\ &\leq 2[-\lambda + (n-1)\Lambda] - 2n\lambda \\ &= 2[(n-1)\Lambda - (n+1)\lambda]. \end{aligned}$$

Putting together these estimates we have

$$\mathrm{Tr} \left\{ \begin{pmatrix} \alpha - 1 & 0 \\ 0 & I \end{pmatrix} (A_1 + A_2 - 2A_2^{1/2}QA_1^{1/2}) \right\} \leq 2[(n-1+2\alpha)\Lambda - (n+1)\lambda],$$

so (4.11) holds with $Q = J_0$ if

$$\frac{\Lambda}{\lambda} < \frac{n+1}{n-1+2\alpha}. \quad (6.4)$$

Comparing this with (4.16), we observe that (6.4) improves the exponent α when the distortion Λ/λ is close enough to 1. However, for sufficiently large distortion (but still below $\frac{n+1}{n-1}$), (4.16) gives a better exponent. In any case, both conditions give the same upper bound for the distortion when $\alpha \rightarrow 0$, that is $\Lambda/\lambda < \frac{n+1}{n-1}$ for every $\alpha \in (0, 1)$.

It is worth noting that, since Q was originally chosen to be the orthogonal matrix minimizing the left-hand side in (4.11), it is reasonable to expect that a refinement of the argument in Lemma 4.6 would produce better estimates, as it was already pointed out in the discussion at the beginning of Section 6. In this direction, in the following example we show (in the constant coefficients case, $A_1 = A_2$) that, even when the distortion Λ/λ is small, the optimal choice of Q minimizing (4.13) differs from the reflection J_0 .

Example 6.2. In this example we discuss the effect of choosing the mirror point coupling in (4.11) when $A_1 = A_2$ is certain 2×2 symmetric matrix with given distortion $\Lambda/\lambda = \delta \geq 1$. In particular, we compare the conditions under which (4.11) holds when Q is either J_0 or the optimal coupling minimizing (4.13).

Let A be a 2×2 symmetric and positive definite matrix. Then the inequality (4.11) for $A_1 = A_2 = A$ is equivalent to

$$\mathrm{Tr} \{ A^{1/2} J_\alpha A^{1/2} (I - Q) \} < 0, \quad (6.5)$$

where we have denoted

$$J_\alpha = \begin{pmatrix} \alpha - 1 & 0 \\ 0 & 1 \end{pmatrix}$$

for simplicity. Fix $n = 2$ and

$$A = \begin{pmatrix} \delta + 1 & \delta - 1 \\ \delta - 1 & \delta + 1 \end{pmatrix},$$

with $\delta \geq 1$. Then A has eigenvalues $\lambda = 2$ and $\Lambda = 2\delta$, so its distortion is $\Lambda/\lambda = \delta$. In this example we compare the conditions under which (6.5) holds when Q is either J_0 or the nearest orthogonal matrix to $A^{1/2}J_\alpha A^{1/2}$, for which the minimum is attained.

First, if $Q = J_0$, then $I - J_0 = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$ and thus

$$\mathrm{Tr} \{ A^{1/2} J_\alpha A^{1/2} (I - J_0) \} = 2e_1^\top A^{1/2} J_\alpha A^{1/2} e_1.$$

Computing the square root of A we get

$$A^{1/2} = \frac{1}{\sqrt{2}} \begin{pmatrix} \sqrt{\delta} + 1 & \sqrt{\delta} - 1 \\ \sqrt{\delta} - 1 & \sqrt{\delta} + 1 \end{pmatrix},$$

so

$$\mathrm{Tr} \{A^{1/2} J_\alpha A^{1/2} (I - J_0)\} = -(1 - \alpha)(\sqrt{\delta} + 1)^2 + (\sqrt{\delta} - 1)^2.$$

Hence, (6.5) holds with $Q = J_0$ if and only if

$$\alpha < 1 - \left(\frac{\sqrt{\delta} - 1}{\sqrt{\delta} + 1} \right)^2,$$

or equivalently

$$\frac{\Lambda}{\lambda} = \delta < \left(\frac{1 + \sqrt{1 - \alpha}}{1 - \sqrt{1 - \alpha}} \right)^2.$$

On the other hand, by the equality in (4.14) from Lemma 4.4 with $M = A^{1/2} J_\alpha A^{1/2}$ we have that

$$\min_{Q \in O(n)} \mathrm{Tr} \{A^{1/2} J_\alpha A^{1/2} (I - Q)\} = \mathrm{Tr} \{A^{1/2} J_\alpha A^{1/2}\} - \mathrm{Tr} \{[(A^{1/2} J_\alpha A^{1/2})^2]^{1/2}\}.$$

It is easy to check that

$$\mathrm{Tr} \{A^{1/2} J_\alpha A^{1/2}\} = \mathrm{Tr} \{J_\alpha A\} = \alpha(\delta + 1).$$

For the other term we recall that

$$\mathrm{Tr} \{M^{1/2}\}^2 = \mathrm{Tr} \{M\} + 2\sqrt{\det\{M\}}$$

for every positive definite symmetric 2×2 matrix M , so

$$\mathrm{Tr} \{[(A^{1/2} J_\alpha A^{1/2})^2]^{1/2}\} = \sqrt{\mathrm{Tr} \{(J_\alpha A)^2\} + 2\sqrt{\det\{(J_\alpha A)^2\}}},$$

where

$$\mathrm{Tr}\{(J_\alpha A)^2\} = \alpha^2(\delta + 1)^2 + 8(1 - \alpha)\delta$$

and

$$\sqrt{\det\{(J_\alpha A)^2\}} = (1 - \alpha) \det\{A\} = 4(1 - \alpha)\delta.$$

Hence, replacing above we obtain

$$\min_{Q \in O(n)} \mathrm{Tr} \{A^{1/2} J_\alpha A^{1/2} (I - Q)\} = \alpha(\delta + 1) - \sqrt{\alpha^2(\delta + 1)^2 + 16(1 - \alpha)\delta},$$

which is clearly always negative, independently of the relation between the distortion $\Lambda/\lambda = \delta$ and the exponent α .

In conclusion:

$$\mathrm{Tr} \{A^{1/2} J_\alpha A^{1/2} (I - J_0)\} < 0 \quad \iff \quad 1 \leq \frac{\Lambda}{\lambda} < \left(\frac{1 + \sqrt{1 - \alpha}}{1 - \sqrt{1 - \alpha}} \right)^2,$$

and

$$\min_{Q \in O(n)} \operatorname{Tr} \{A^{1/2} J_\alpha A^{1/2} (I - Q)\} < 0 \quad \iff \quad \frac{\Lambda}{\lambda} \geq 1 \quad \text{and} \quad \alpha \in (0, 1).$$

Example 6.3. The mirror point coupling $Q = J_0$ fails to give (4.11) when the distortion is large. Let

$$A = \begin{pmatrix} 5 & -12 \\ -12 & 29 \end{pmatrix},$$

which is a positive definite matrix and has the following square root:

$$A^{1/2} = \begin{pmatrix} 1 & -2 \\ -2 & 5 \end{pmatrix}.$$

Then

$$A^{1/2} J_\alpha A^{1/2} = \begin{pmatrix} 3 & -8 \\ -8 & 21 \end{pmatrix} + \alpha \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix}$$

for every $\alpha \in (0, 1)$. If we choose $Q = J_0$, then $I - Q = 2e_1 e_1^\top$, and thus

$$\operatorname{Tr} \{A^{1/2} J_\alpha A^{1/2} (I - Q)\} = 2e_1^\top A^{1/2} J_\alpha A^{1/2} e_1 = 2(3 + \alpha) > 0$$

for every $\alpha \in (0, 1)$, so (6.5) (and thus (4.11)) does not hold in general for $Q = J_0$. On the other hand, for Q the optimal coupling, (6.5) follows immediately from recalling (4.18).

6.3. Counterexamples for large distortion

The following examples show that the bound on the distortion Λ/λ in the assumptions of our main theorem is really needed in this method. We first remark that this bound is only needed in the case $|x - z| > N\sqrt{\lambda}\varepsilon$. Indeed, when $|x - z| \leq N\sqrt{\lambda}\varepsilon$, the only place where the distortion crucially appears is in the inequality (5.12), and for the proof in this case any bound on Λ/λ is sufficient: for convenience we selected $\Lambda/\lambda \leq 3$ there. Thus, the only part of the paper where the bound on the distortion comes crucially into a play is in Section 4, more precisely, in the condition (4.11).

As we explained in Section 4.2, given $A_1, A_2 \in \mathcal{A}(\lambda, \Lambda)$ our strategy was based on showing (4.11) for some $\alpha \in (0, 1)$. Let us reformulate the problem by defining the ellipsoids $E_1 := E_{A_1} = A_1^{1/2}\mathbb{B}$ and $E_2 := E_{A_2} = A_2^{1/2}\mathbb{B}$ and letting $\phi(y) = A_2^{1/2} Q A_1^{-1/2} y$. Then ϕ maps E_1 into E_2 . Thus, performing a change of variables we see that (4.11) is equivalent to

$$\int_{E_1} |P_i(y - \phi(y))|^2 dy > \int_{E_1} |(I - P_1)(y - \phi(y))|^2 dy, \quad (6.6)$$

where $P_i h$ stands for the projection of a vector $h \in \mathbb{R}^n$ over the i th coordinate axis for $i = 1, \dots, n$, i.e. $P_i h = (e_i^\top h) e_i = (e_i e_i^\top) h$. That is, the method in Section 4 works if the length of the projection on $\{e_1\}^\perp$ of the difference between $y \in E_1$ and $\phi(y) \in E_2$ is larger (in average) than the length of such projection over $\operatorname{span}\{e_1\}$.

Here, we can consider any measure preserving coupling map ϕ such that $\phi(E_1) = E_2$. This is the case in the following examples, which are constructed in such a way that the condition (6.6) does not hold for any measure preserving map ϕ . This is achieved by considering a pair of ellipsoids with large distortion oriented in such a way that the largest principal axes are aligned with coordinate axes orthogonal to e_1 , then the difference $\phi(y) - y$ project over the first coordinate axis will necessarily be smaller (in average) than projected over one

of the orthogonal axes to e_1 . In other words, the average distance in terms of f increases, no matter which coupling we use.

6.3.1. A counterexample in two dimensions

Let us consider an ellipse centered at 0 with its largest axis oriented in the direction of e_2 and a ball,

$$E_1 = \begin{pmatrix} 1/10 & 0 \\ 0 & 10 \end{pmatrix} \mathbb{B} \quad \text{and} \quad E_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mathbb{B}.$$

In this case, both the ellipses have the same area, $|E_1| = |E_2| = \pi$, but different distortion: E_1 has distortion $\sqrt{100}$ while the distortion of E_2 is equal to 1. Let $\phi : E_1 \rightarrow E_2$ be any measure preserving map. Let us denote $y = (y_1, y_2)^\top \in \mathbb{R}^2$. Since $|y_1| < \frac{1}{10}$ in E_1 and $|y_1| < 1$ in E_2 , it turns out that $|P_1(\phi(y) - y)| < \frac{11}{10}$ in E_1 , and thus

$$\int_{E_1} |P_1(\phi(y) - y)|^2 dy \leq 1.21.$$

On the other hand, since $|y_2| < 1$ in E_2 , then $|P_2\phi(y)| < 1$ in E_1 . Moreover, since $E_1 \cap \{|y_2| \geq 5\}$ is a non-empty subset of E_1 with positive measure, we can estimate $|P_2(\phi(y) - y)|$ in $E_1 \cap \{|y_2| \geq 5\}$ from below by 4. Indeed, by the triangle inequality,

$$|P_2(\phi(y) - y)| \geq ||P_2\phi(y)| - |y_2|| > |y_2| - 1 \geq 4$$

for every $y \in E_1$ such that $|y_2| \geq 5$. Then, averaging we get

$$\begin{aligned} \int_{E_1} |P_2(\phi(y) - y)|^2 dy &\geq \frac{1}{|E_1|} \int_{E_1 \cap \{|y_2| \geq 5\}} |P_2(\phi(y) - y)|^2 dy \\ &\geq 16 \frac{|E_1 \cap \{|y_2| \geq 5\}|}{|E_1|}. \end{aligned}$$

Since $|E_1| = \pi$ and

$$E_1 \cap \{|y_2| \geq 5\} = \begin{pmatrix} 1/10 & 0 \\ 0 & 10 \end{pmatrix} (\mathbb{B} \cap \{|y_2| \geq 1/2\}),$$

then

$$\begin{aligned} 16 \frac{|E_1 \cap \{|y_2| \geq 5\}|}{|E_1|} &= \frac{16}{\pi} |\mathbb{B} \cap \{|y_2| \geq 1/2\}| = \frac{64}{\pi} \int_{1/2}^1 \sqrt{1-t^2} dt \\ &= 16 \left(\frac{2}{3} - \frac{\sqrt{3}}{2\pi} \right) \simeq 6.26 \dots \end{aligned}$$

Hence,

$$\int_{E_1} |P_1(\phi(y) - y)|^2 dy < \int_{E_1} |P_2(\phi(y) - y)|^2 dy,$$

which contradicts (6.6).

6.3.2. A counterexample in three dimensions

In this case we consider two ellipsoids with the same shape but with different orientations,

$$E_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 100 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbb{B} \quad \text{and} \quad E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 100 \end{pmatrix} \mathbb{B}.$$

Observe that the largest axes of the ellipsoids are aligned with e_2 and e_3 , respectively. Let us denote $y = (y_1, y_2, y_3)^\top \in \mathbb{R}^3$. Then $|y_1| < 1$ in both E_1 and E_2 , that is, for any measure preserving map $\phi : E_1 \rightarrow E_2$, we have $|y_1| < 1$ and $|P_1\phi(y)| < 1$, and thus

$$\int_{E_1} |P_1(\phi(y) - y)|^2 dy \leq 4.$$

On the other hand, let us consider the set $E_1 \cap \{|y_2| \geq 5\}$. Since $|y_2| < 1$ in E_2 , then the distance between a point in E_2 and $E_1 \cap \{|y_2| \geq 5\}$ is bounded from below by 4. In particular, $|P_2(\phi(y) - y)| \geq |y_2| - |P_2\phi(y)| \geq 4$ in $E_1 \cap \{|y_2| \geq 5\}$, and thus

$$\begin{aligned} \int_{E_1} |(I - P_1)(\phi(y) - y)|^2 dy &= \int_{E_1} |P_2(\phi(y) - y)|^2 dy + \int_{E_1} |P_3(\phi(y) - y)|^2 dy \\ &\geq \int_{E_1} |P_2(\phi(y) - y)|^2 dy \\ &\geq \frac{1}{|E_1|} \int_{E_1 \cap \{|y_2| \geq 5\}} |P_2(\phi(y) - y)|^2 dy \\ &\geq 16 \frac{|E_1 \cap \{|y_2| \geq 5\}|}{|E_1|} \geq 8. \end{aligned}$$

The last inequality follows by the fact that the set $\{|y_2| \geq 5\}$ takes at least one half of the volume of the ellipsoid. Then

$$\int_{E_1} |P_1(\phi(y) - y)|^2 dy < \int_{E_1} |(I - P_1)(\phi(y) - y)|^2 dy,$$

which also contradicts (6.6).

Acknowledgements. The authors have been supported by the Academy of Finland project #298641. Á. A. was partially supported by the grant MTM2017-85666-P. The authors would like to thank Eero Saksman for discussions related to this paper.

REFERENCES

- [1] Á. Arroyo, J. Heino and M. Parviainen, Tug-of-war games with varying probabilities and the normalized $p(x)$ -Laplacian. *Commun. Pure Appl. Anal.* **16** (2017) 915–944.
- [2] Á. Arroyo, H. Luiro, M. Parviainen and E. Ruosteenoja, Asymptotic lipschitz regularity for tug-of-war games with varying probabilities. Preprint [arXiv:1806.10838](https://arxiv.org/abs/1806.10838) (2018).
- [3] M. Cranston, Gradient estimates on manifolds using coupling. *J. Funct. Anal.* **99** (1991) 110–124.
- [4] J. Han, Local lipschitz regularity for functions satisfying a time-dependent dynamic programming principle. Preprint [arXiv:1812.00646](https://arxiv.org/abs/1812.00646) (2019).
- [5] H. Ishii and P.-L. Lions, Viscosity solutions of fully nonlinear second-order elliptic partial differential equations. *J. Differ. Equ.* **83** (1990) 26–78.
- [6] R.V. Kohn and S. Serfaty, A deterministic-control-based approach to motion by curvature. *Comm. Pure Appl. Math.* **59** (2006) 344–407.

- [7] N.V. Krylov and M.V. Safonov, An estimate for the probability of a diffusion process hitting a set of positive measure. *Dokl. Akad. Nauk SSSR* **245** (1979) 18–20.
- [8] S. Kusuoka, Hölder continuity and bounds for fundamental solutions to nondivergence form parabolic equations. *Anal. Partial Differ. Equ.* **8** (2015) 1–32.
- [9] S. Kusuoka, Hölder and Lipschitz continuity of the solutions to parabolic equations of the non-divergence type. *J. Evol. Equ.* **17** (2017) 1063–1088.
- [10] M. Lewicka, Random tug of war games for the p -Laplacian: $1 < p < \infty$. Preprint [arXiv:1810.03413](https://arxiv.org/abs/1810.03413) (2019).
- [11] M. Lewicka and J.J. Manfredi, The obstacle problem for the p -laplacian via optimal stopping of tug-of-war games. *Probab. Theory Related Fields* **167** (2017) 349–378.
- [12] T. Lindvall and L.C.G. Rogers, Coupling of multidimensional diffusions by reflection. *Ann. Probab.* **14** (1986) 860–872.
- [13] H. Luiro and M. Parviainen, Regularity for nonlinear stochastic games. *Ann. Inst. Henri Poincaré Anal. Non Linéaire* **35** (2018) 1435–1456.
- [14] J.J. Manfredi, M. Parviainen and J.D. Rossi, On the definition and properties of p -harmonic functions. *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* **11** (2012) 215–241.
- [15] N. Nadirashvili, Nonuniqueness in the martingale problem and the Dirichlet problem for uniformly elliptic operators. *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* **24** (1997) 537–549.
- [16] M. Parviainen and E. Ruosteenoja, Local regularity for time-dependent tug-of-war games with varying probabilities. *J. Differ. Equ.* **261** (2016) 1357–1398.
- [17] Y. Peres and S. Sheffield, Tug-of-war with noise: a game-theoretic view of the p -Laplacian. *Duke Math. J.* **145** (2008) 91–120.
- [18] Y. Peres, O. Schramm, S. Sheffield and D.B. Wilson, Tug-of-war and the infinity Laplacian. *J. Amer. Math. Soc.* **22** (2009) 167–210.
- [19] A. Porretta and E. Priola, Global Lipschitz regularizing effects for linear and nonlinear parabolic equations. *J. Math. Pures Appl.* **100** (2013) 633–686.
- [20] C. Pucci and G. Talenti, Elliptic (second-order) partial differential equations with measurable coefficients and approximating integral equations. *Adv. Math.* **19** (1976) 48–105.