TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 00, Number 0, Xxxx XXXX, Pages 000–000
S 0002-9947(XX)0000-0

COUPLING LÉVY MEASURES AND COMPARISON PRINCIPLES FOR VISCOSITY SOLUTIONS

NESTOR GUILLEN, CHENCHEN MOU, AND ANDRZEJ ŚWIĘCH

Abstract. We prove new comparison principles for viscosity solutions of nonlinear integro-differential equations. The operators to which the method applies Q1 include but are not limited to those of Lévy–Itô type. The main idea is to use an optimal transport map to couple two different Lévy measures and use the resulting coupling in a doubling of variables argument.

1. Introduction

In this paper we study comparison principles for viscosity subsolutions and supersolutions of integro-differential equations of the form

$$I(u,x) = \sup_{\alpha\in\mathcal{A}} \inf_{\beta\in\mathcal{B}} \{-L^{\alpha\beta}(u,x) + c_{\alpha\beta}(x)u(x) + f_{\alpha\beta}(x)\} = 0 \qquad \text{in } \mathcal{O}, \tag{1.1}$$

where $\mathcal{O}$ is a bounded domain of $\mathbb{R}^d$, $c_{\alpha\beta}(x) \geq \lambda > 0$, and

$$L^{\alpha\beta}(u,x) = \int_{\mathbb{R}^d} [u(x+z) - u(x) - \chi_{B_1(0)}(z) Du(x)\cdot z] d\mu_x^{\alpha\beta}(z), \tag{1.2}$$

where $\mu_x^{\alpha\beta}$ are the respective Lévy measures. Equations of the form (1.1) arise in stochastic optimal control and stochastic differential games where the operators are the generators of pure jump processes. In a work by the first author and Schwab [14], it is proved (roughly speaking) that the class of operators given by a min-max as in (1.1) is the same as the class of operators satisfying the global comparison property.

Comparison principles for viscosity solutions of such equations are now well understood in two broad cases. The first case is when the operators admit a Lévy–Itô form. This means that all of the measures $\mu_x^{\alpha\beta}$ are push-forward measures of a single reference measure μ so that $\mu_x^{\alpha\beta} = (T_x^{\alpha\beta})_\#\mu$, where $T_x^{\alpha\beta} : U \to \mathbb{R}^d$ is a family of Borel measurable maps defined on some separable Hilbert space, and μ is a Lévy measure on $U \setminus \{0\}$ (see (5.17) and (5.19)). First comparison principles were obtained by Soner in [22, 23]. Further results, including results for equations Q2 with second order partial differential equation terms were obtained subsequently; see [6–8, 16]. The second case is that of equations of order less than or equal to 1. Here we mention the works of Soner [22, 23] and the papers of Awatif [19, 20], where Q3 comparison principles are proved for very general operators in the class where the operators $L^{\alpha\beta}$ are all such that the function $|z|$ is uniformly integrable with respect

---

Received by the editors July 12, 2018, and, in revised form, April 1, 2019, and April 16, 2019.
2010 *Mathematics Subject Classification.* Primary 35D40 35J60, 35R09, 45K05, 47G20.
Key words and phrases. Optimal transport, Lévy measures, nonlocal equations, viscosity solutions, comparison principles, uniqueness.

to the measures $\mu_x^{\alpha\beta}$. Also Alvarez and Tourin [3] and Alibaud [1] considered various parabolic equations with nonlocal terms of order 0, that is, with μ_x of finite mass.

Little is known when the Lévy measures arising in (1.1) and (1.2) are neither integrable with respect to $|z|$ nor of Lévy–Itô form. The second and third authors proved in [18] several comparison results for viscosity solutions which have some regularity. Chasseigne and Jakobsen proved in [10] comparison results for fully nonlinear equations involving quasi-linear nonlocal operators. We also mention continuous dependence estimates for weak entropy solutions of degenerate parabolic equations with nonlinear fractional diffusion proved by Alibaud, Cifani, and Jakobsen in [2]. Proving comparison in general is an important question, as many operators of interest are not covered by the two situations discussed above, such as the Dirichlet-to-Neumann maps for nonlinear elliptic equations or control/game problems where the processes are not classical Lévy–Itô diffusions.

In this paper we introduce optimal transport techniques in an attempt to understand this question. We obtain a comparison for nonlocal equations (1.1) and (1.2) that cover the previous two instances without requiring a Lévy–Itô structure or a restriction on the order of the operators. The idea is to use an optimal coupling for the Lévy measures arising in the nonlocal terms. Then the continuity of the Lévy measures with respect to the base point x is estimated with respect to an optimal transport based metric.

The condition we impose is Lipschitz continuity with respect to an L^p-transport metric. The exponent $p \in [1, 2]$ is related to the order of the singularity at $z = 0$ for the Lévy measures. In the case of operators of order smaller than 1, it is possible to use the metric corresponding to $p = 1$, in which case our condition is (essentially) a dual formulation of the condition used by Awatif [19]. Likewise, in the Lévy–Itô case our condition reduces to the one typically imposed in the literature [8, 16].

Unfortunately, it is rather difficult to check the Lipschitz regularity of μ_x with respect to our L^p-transport metric when $p > 1$ and μ_x is not in Lévy–Itô form (this is precisely the case where the comparison is still unknown). Such Lipschitz estimates are even nontrivial to check for Lévy measures of finite mass and fail to hold.[1] It is our hope that this paper will spur further research that will expand the class of families of measures $\{\mu_x^{\alpha\beta}\}_{x,\alpha,\beta}$, where this new approach can be applied.

1.1. **The basic idea.**

Let us illustrate the main idea of the paper in a simple situation. Consider the linear equation

$$\lambda u(x) - L(u, x) = 0 \qquad \text{in } \mathcal{O}, \tag{1.3}$$

where $\lambda > 0$ and $L(u, x)$ is an operator of the form (1.2), where we make the following simplifying assumption on the Lévy measures μ_x: μ_x is a *probability measure* with finite second moments for every x, and there is some $C > 0$ such that, for any x, y,

$$\mathrm{d}_2(\mu_x, \mu_y) \le C|x - y|. \tag{1.4}$$

Here d_2 denotes the optimal transport distance with respect to the square distance (the so-called Wasserstein distance). Suppose that u is a bounded viscosity subsolution of (1.3) and that v is a bounded viscosity supersolution of (1.3) such that $u \le v$ on $\mathbb{R}^d \setminus \mathcal{O}^c$. We start with the typical comparison proof. We assume that

[1] The authors would like to thank Alessio Figalli for helpful comments regarding this question.

$u \not\leq v$. We double the variables and penalize the doubling considering for $\varepsilon > 0$ the function

$$u(x) - v(y) - \frac{1}{\varepsilon}|x-y|^2.$$

Suppose that for all sufficiently small ε the global maximum is attained at $(x_\varepsilon, y_\varepsilon)$, where $u(x_\varepsilon) - v(y_\varepsilon) \geq \ell > 0$ for some $\ell > 0$. In such circumstances it is well known that

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon}|x_\varepsilon - y_\varepsilon|^2 = 0,$$

so for small ε we must have $(x_\varepsilon, y_\varepsilon) \in \mathcal{O} \times \mathcal{O}$. Because of the global maximum, we have

$$\begin{aligned}\left(u(x_\varepsilon + x) - u(x_\varepsilon) - 2\frac{x \cdot (x_\varepsilon - y_\varepsilon)}{\varepsilon}\right)&\\ - \left(v(y_\varepsilon + y) - v(y_\varepsilon) - 2\frac{y \cdot (x_\varepsilon - y_\varepsilon)}{\varepsilon}\right) &\leq \frac{1}{\varepsilon}|x-y|^2 \qquad \forall x, y.\end{aligned}$$

For x and y let $\pi_{x,y}$ denote a probability measure on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ_x and μ_y achieving the optimal (quadratic) transport cost between them. Then we integrate the above inequality with respect to the measure $\pi_{x_\varepsilon, y_\varepsilon}$ to obtain

$$L(u, x_\varepsilon) - L(v, y_\varepsilon) \leq \frac{1}{\varepsilon} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 \, d\pi_{x_\varepsilon, y_\varepsilon}(x,y).$$

Thus, by the definition of a viscosity solution and the definition of $\pi_{x,y}$, we get

$$\lambda(u(x_\varepsilon) - v(y_\varepsilon)) \leq \frac{1}{\varepsilon} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 d\pi_{x_\varepsilon, y_\varepsilon}(x,y) = \frac{1}{\varepsilon} \mathrm{d}_2(\mu_{x_\varepsilon}, \mu_{y_\varepsilon})^2,$$

where the last equality follows from the optimality of $\pi_{x_\varepsilon, y_\varepsilon}$. Then, using (1.4), we obtain

$$0 < \lambda \ell \leq \lambda(u(x_\varepsilon) - v(y_\varepsilon)) \leq \frac{C}{\varepsilon}|x_\varepsilon - y_\varepsilon|^2.$$

Since the right-hand side goes to 0 as $\varepsilon \to 0$, we obtain a contradiction. Thus in this model case the proof of comparison reduces to checking whether the measures μ_x satisfy the Lipschitz condition (1.4) with respect of the (quadratic) optimal transport distance.

Of course, it is atypical for a Lévy measure to also be a probability or even a finite measure of constant total mass. To deal with this issue, we will make use of an optimal transport problem featuring an "infinite mass reservoir" at 0 (after all, the mass of the Lévy measure at 0 is immaterial). This means in particular that one can consider transport between measures which may have unequal or infinite masses. This problem was studied by Figalli and Gigli in [13], motivated by questions of gradient flows with Dirichlet boundary conditions, and their work is aptly suited to our purposes.

1.2. **Outline of the paper.** The notation and definitions are explained in Section 2. The transport metric is explained in Section 3. Section 4 contains the assumptions and the statement of the main result. In Section 5 we prove the main comparison principle using the above technique. We then show how the result covers comparison principles for nonlocal equations involving nonlocal terms either of Lévy form of order $\sigma < 1$ (see Example 5.12) or of Lévy–Itô form (see Example 5.13). In Section 6 we discuss variants of our approach which we illustrate in Example 6.2 related to operators of fractional Laplacian type. We also discuss in Section 6 two other examples (Examples 6.3 and 6.4) comparing our results to these of [19]. Finally, in Section 7 we derive various comparison principles for equations which have more regular viscosity solutions. They lead to uniqueness of viscosity solutions for a class of uniformly elliptic nonlocal equations (see Example 7.6). The paper ends with an appendix which follows [13] collects the main facts about the optimal transport problem "with boundary".

2. Notation and definitions

In the whole paper we will consider equation (1.1), where the operators $L^{\alpha\beta}$ are assumed to be of the form (1.2), and $\{\mu_x^{\alpha\beta}\}_{x,\alpha,\beta}$ is a family of Lévy measures (see Definition 2.1). Denoting

$$\delta u(x,z) := u(x+z) - u(x) - \chi_{B_1(0)}(z) Du(x) \cdot z,$$

we will write

$$L^{\alpha\beta}(u,x) = \int_{\mathbb{R}^d} \delta u(x,z) d\mu_x^{\alpha\beta}(z). \tag{2.1}$$

We will denote by $B_r(x)$ the open ball in $\mathbb{R}^d$ centered at x with radius $r > 0$, and by B_r the open ball centered at 0 with radius r. Given an open set $\Omega \subset \mathbb{R}^d$ and $h > 0$, we define

$$\Omega_h = \{x \in \Omega \ \mid d(x, \partial\Omega) > h\}.$$

For a subset $A \subset \mathbb{R}^d$ we denote by A^c its complement, i.e., $A^c = \mathbb{R}^d \setminus A$, and by χ_A the characteristic function of A.

For $0 < \alpha \leq 1$ and a domain $\mathcal{O}$ in $\mathbb{R}^d$, we denote by $C^{0,\alpha}(\mathcal{O})$ the space of α-Hölder continuous functions in $\mathcal{O}$.

We write $C^k(\mathcal{O}), k = 1, 2, \dots$ for the usual spaces of k-times continuously differentiable functions in $\mathcal{O}$. The space $C_b^k(\mathcal{O})$ (resp., $C_b^p(\mathcal{O})$) consists of functions in $C^k(\mathcal{O})$ (resp., $C^p(\mathcal{O})$) which are bounded. We write $\mathrm{BUC}(\mathbb{R}^d)$ for the set of bounded and uniformly continuous functions in $\mathbb{R}^d$. For two bounded measures μ, ν we will write $\mathrm{d_{TV}}(\mu, \nu)$ to denote the total variation of $\mu - \nu$.

Let μ be a Borel measure in $\mathbb{R}^d \setminus \{0\}$, and let $1 \leq p \leq 2$. We define

$$\mathcal{N}_p(\mu) := \int_{\mathbb{R}^d \setminus \{0\}} \min\{1, |z|^p\} d\mu(z). \tag{2.2}$$

Definition 2.1. Let $1 \leq p \leq 2$. We define

$$\mathbb{L}_p(\mathbb{R}^d) := \Big\{ \mu \text{ positive Borel measure in } \mathbb{R}^d \setminus \{0\} \ \mid \ \mathcal{N}_p(\mu) < \infty \Big\}.$$

The set of all Lévy measures is $\mathbb{L}_2(\mathbb{R}^d)$. If Ω is an open subset of $\mathbb{R}^d$, we will consider the set

$$\mathbb{L}_p(\Omega) := \Big\{ \mu \in \mathbb{L}_p(\mathbb{R}^d) \ \mid \ \mathrm{spt}(\mu) \subset \Omega \Big\}.$$

Note that

$$\mathbb{L}_p(\Omega) \subset \mathbb{L}_q(\Omega) \qquad \text{whenever } p \leq q.$$

In other words, measures μ in $\mathbb{L}_p(\Omega)$ are measures in $\mathbb{L}_p(\mathbb{R}^d)$ such that $\mu(\mathbb{R}^d \backslash \Omega) = 0$. We decompose every measure $\mu \in \mathbb{L}_p(\mathbb{R}^d)$ as

$$\mu = \hat{\mu} + \check{\mu}, \tag{2.3}$$

where $\hat{\mu}(\cdot) := \mu(\cdot \cap B_1) \in \mathbb{L}_p(B_1)$ and $\check{\mu}(\cdot) := \mu(\cdot \cap (\mathbb{R}^d \setminus B_1))$. We note that $\check{\mu}$ is a bounded measure.

Consider the Lévy operator given by some measure μ,

$$L(u, x) = \int_{\mathbb{R}^d} [u(x+z) - u(x) - \chi_{B_1(0)}(z) Du(x) \cdot z] \, d\mu(z). \tag{2.4}$$

Let us decompose this operator as the sum of two operators, corresponding to the Lévy measure decomposition in (2.3),

$$L(u, x) = \hat{L}(u, x) + \check{L}(u, x),$$

where

$$\hat{L}(u, x) = \int_{\mathbb{R}^d} [u(x+z) - u(x) - \chi_{B_1(0)}(z) Du(x) \cdot z] \, d\hat{\mu}(z), \tag{2.5}$$

$$\check{L}(u, x) = \int_{\mathbb{R}^d} [u(x+z) - u(x)] \, d\check{\mu}(z). \tag{2.6}$$

Definition 2.2. Given a Lévy measure μ in $\mathbb{R}^d$, we define

$$L_\mu(u, x) := \int_{\mathbb{R}^d} [u(x+z) - u(x) - \chi_{B_1(0)}(z) Du(x) \cdot z] \, d\mu(z).$$

Definition 2.3. For $p \in (1, 2)$, a function u is said to be pointwise-C^p at a point x_0 if u is differentiable at x_0 and if there exists a constant $C > 0$ such that, for all x in a neighborhood of x_0,

$$|u(x) - u(x_0) - Du(x_0) \cdot (x - x_0)| \leq C|x - x_0|^p. \tag{2.7}$$

If u is differentiable at x_0 and (2.7) is satisfied with $p = 2$ we say that u is pointwise-$C^{1,1}$ at x_0. For $p \in (0, 1]$ a function is said to be pointwise-C^p at a point x_0 if there is a constant $C > 0$ such that, for all x in a neighborhood of x_0,

$$|u(x) - u(x_0)| \leq C|x - x_0|^p.$$

3. A transportation metric for Lévy measures

We will use a transportation metric on the space of Lévy measures. This metric takes advantage of an "infinite reservoir" of mass which allows one to handle measures which may not have equal (or finite) total mass. Such a metric was considered by Figalli and Gigli [13], where they studied the basic properties of such a metric and used it to analyze gradient flows with Dirichlet boundary conditions. Our presentation here generally follows that of [13]. This is not the only possible extension of the transport metric to the case of unequal masses: other notions have been considered by Kantorovich and Rubinstein. Another notion of distance for Lévy measures is considered in [15].

We consider the following set of measures,

$$\mathcal{M}_p(\mathbb{R}^d) := \left\{ \mu \in \mathbb{L}_p(\mathbb{R}^d) \;\middle|\; \int_{\mathbb{R}^d \setminus \{0\}} |z|^p d\mu(z) < \infty \right\},$$

that is, the set of Lévy measures with finite p-moment. Note in particular that $\mathbb{L}_p(B_1) \subset \mathcal{M}_p(\mathbb{R}^d)$ due to the measures being supported in B_1.

First, we define the notion of admissible couplings between Lévy measures (see also Definition A.3 for an analogous definition in a more general setting).

Definition 3.1. Let $\mu_1, \mu_2 \in \mathcal{M}_p(\mathbb{R}^d)$. An admissible transport plan between μ_1 and μ_2 is any positive Borel measure on $\mathbb{R}^d \times \mathbb{R}^d$ such that $\gamma(\{0\} \times \{0\}) = 0$ and

$$\pi^1_{\#}\gamma_{|_{\mathbb{R}^d\setminus\{0\}}} = \mu_1, \;\; \pi^2_{\#}\gamma_{|_{\mathbb{R}^d\setminus\{0\}}} = \mu_2,$$

where for $i = 1, 2$, $\pi^i : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is defined by $\pi^i(x_1, x_2) = x_i$. The set of admissible transport plans will be denoted by $\mathrm{Adm}(\mu_1, \mu_2)$.

In particular, if γ is admissible, then, for a Borel set A compactly supported in $\mathbb{R}^d \setminus \{0\}$,

$$\gamma(A \times \mathbb{R}^d) = \mu_1(A), \qquad \gamma(\mathbb{R}^d \times A) = \mu_2(A).$$

The key point in Definition 3.1 which distinguishes it from the notion of optimal transport plans is that the marginals of γ coincide with μ_1 and μ_2 only away from the origin. In particular, the marginals of γ may assign any amount of mass to the origin.

Definition 3.2. Let $1 \le p \le 2$. For a positive Borel measure γ on $\mathbb{R}^d \times \mathbb{R}^d$, we define

$$J_p(\gamma) := \int_{\mathbb{R}^d\times\mathbb{R}^d} |x - y|^p \; d\gamma(x, y).$$

In the appendix we study the problem of minimizing $J_p(\gamma)$ over $\gamma \in \mathrm{Adm}(\mu_1, \mu_2)$ in greater generality. In this section we limit ourselves to stating a few further definitions and a few results needed in later sections.

Definition 3.3. Let $1 \le p \le 2$. The p-distance between measures $\mu, \nu \in \mathcal{M}_p(\mathbb{R}^d)$ is defined by

$$\mathrm{d}_{\mathbb{L}_p}(\mu, \nu) := \left(\inf_{\gamma\in\mathrm{Adm}(\mu,\nu)} J_p(\gamma) \right)^{\frac{1}{p}}.$$

The optimization problem used in the definition of $\mathrm{d}_{\mathbb{L}_p}(\mu_1, \mu_2)$ shares many properties with the usual optimal transportation problem.

Theorem 3.4. *For $\mu_1, \mu_2 \in \mathcal{M}_p(\mathbb{R}^d)$ there is at least one $\gamma \in \mathrm{Adm}(\mu_1, \mu_2)$ that achieves the minimum value of J_p.*

Proof. The theorem is a special case of Theorem A.5 (see the appendix). □

The fact that $\mathrm{d}_{\mathbb{L}_2}$ defines a distance was proved in [13, Theorem 2.2 and Proposition 2.7]. We will need this result for any p.

Theorem 3.5. $\mathrm{d}_{\mathbb{L}_p}$ *defines a metric in* $\mathcal{M}_p(\mathbb{R}^d)$.

Proof. The theorem is a special case of Theorem A.16. □

The main tool at our disposal when estimating $\mathrm{d}_{\mathbb{L}_p}$ is the following duality result.

Lemma 3.6. *For $1 \le p \le 2$ and $\mu_1, \mu_2 \in \mathcal{M}_p(\mathbb{R}^d)$, we have*

$$\mathrm{d}_{\mathbb{L}_p}(\mu_1, \mu_2)^p = \sup\Big\{ \int_{\mathbb{R}^d} \phi(x)\; d\mu_1(x) + \int_{\mathbb{R}^d} \psi(y)\; d\mu_2(y) \;\mid\; (\phi, \psi) \in \mathrm{Adm}^p \Big\}.$$

Here Adm^p *denotes the set*

$$\begin{aligned}\mathrm{Adm}^p := \Big\{(\phi,\psi) \ \mid \ & \phi\in L^1(\mu),\ \psi\in L^1(\nu),\\ & \phi \text{ and } \psi \text{ are upper semicontinuous},\\ & \phi(0)=\psi(0)=0, \text{ and } \phi(x)+\psi(y)\le |x-y|^p \ \ \forall x,y\in\mathbb{R}^d\Big\}.\end{aligned}$$

Proof. The lemma is a special case of Lemma A.14 □

Remark 3.7. In most of the paper we need only to take $\mathrm{d}_{\mathbb{L}_p}(\mu,\nu)$ for $\mu,\nu\in\mathbb{L}_p(B_1)$. In this case we could equivalently define the distance by considering the transport problem in $\overline{B}_1\times\overline{B}_1$, i.e., taking $\Omega=B_1$ instead of $\Omega=\mathbb{R}^d$ (see the appendix). We note that if $\mu,\nu\in\mathbb{L}_p(B_1)$ and $\gamma\in\mathrm{Adm}(\mu,\nu)$, then $\gamma((B_1\times B_1)^c)=0$.

The following proposition (proved in the appendix) will be used in Section 5.

Proposition 3.8. *Let* ψ *be a Lipschitz continuous function with compact support in* $\overline{B}_1\setminus\{0\}$. *If* $\mu,\nu\in\mathbb{L}_p(B_1)$, *then*

$$\left|\int_{B_1}\psi\,d\mu-\int_{B_1}\psi\,d\nu\right|\le\left(\mu(\mathrm{spt}(\psi))+\nu(\mathrm{spt}(\psi))\right)^{\frac{p-1}{p}}[\psi]_{\mathrm{Lip}}\mathrm{d}_{\mathbb{L}_p}(\mu,\nu),$$

where $[\psi]_{\mathrm{Lip}}$ *is the Lipschitz constant of* ψ.

4. Assumptions and main results

In this section we make the necessary assumptions about the measures and various functions appearing in the operator $I(u,x)$ in (1.1). We recall that throughout the whole paper $\mathcal{O}\subset\mathbb{R}^d$ is a bounded domain. The measures $\mu_x^{\alpha\beta}\in\mathbb{L}_p(\mathbb{R}^d)$ for all $x\in\mathcal{O}$, $\alpha\in\mathcal{A}$, $\beta\in\mathcal{B}$ for some index sets $\mathcal{A},\mathcal{B}$. Last but not least, we recall that in (2.3) we introduced the decomposition of a measure μ in terms of measures $\hat{\mu}$ and $\check{\mu}$ supported in $B_1(0)$ and in $\mathbb{R}^d\setminus B_1(0)$, respectively.

Assumption A. There are $p\in[1,2]$ and a constant $C\ge0$ such that

$$\mathrm{d}_{\mathbb{L}_p}(\hat{\mu}_x^{\alpha\beta},\hat{\mu}_y^{\alpha\beta})\le C|x-y| \qquad \forall x,y\in\mathcal{O},\ \forall\alpha,\beta. \tag{4.1}$$

Assumption B. There is a modulus of continuity θ such that

$$\mathrm{d}_{\mathrm{TV}}(\check{\mu}_x^{\alpha\beta},\check{\mu}_y^{\alpha\beta})\le\theta(|x-y|) \quad \forall x,y\in\mathcal{O},\ \forall\alpha,\beta. \tag{4.2}$$

Assumption C. There are a modulus of continuity θ and a constant $C\ge0$ such that

$$|f_{\alpha\beta}(x)-f_{\alpha\beta}(y)|\le\theta(|x-y|) \qquad \forall x,y\in\mathcal{O},\ \forall\alpha,\beta, \tag{4.3}$$

$$|f_{\alpha\beta}(x)|\le C, \qquad \forall x\in\mathcal{O},\ \forall\alpha,\beta. \tag{4.4}$$

Assumption D. There are constants $0<\lambda\le\lambda_1$ such that

$$\lambda\le\inf_{x\in\mathcal{O}}\inf_{\alpha,\beta}c_{\alpha\beta}(x)\le\sup_{x\in\mathcal{O}}\sup_{\alpha,\beta}c_{\alpha\beta}(x)\le\lambda_1, \tag{4.5}$$

and there is a modulus θ such that

$$|c_{\alpha\beta}(x)-c_{\alpha\beta}(y)|\le\theta(|x-y|) \qquad \forall x\in\mathcal{O},\ \forall\alpha,\beta.$$

Assumption E. Let p be from Assumption A. There exist a modulus of continuity θ and a constant $\Lambda \geq 0$ such that

$$\sup_{x\in\mathcal{O}} \sup_{\alpha\beta} \int_{B_r} |z|^p \, d\mu_x^{\alpha\beta}(z) \leq \theta(r), \tag{4.6}$$

$$\sup_{x\in\mathcal{O}} \sup_{\alpha,\beta} \mathcal{N}_p(\mu_x^{\alpha\beta}) \leq \Lambda. \tag{4.7}$$

Assumption B can be weakened; however, we want to keep its simpler form to focus on the main difficulty of dealing with the singular part of the Lévy measures. We leave such generalizations to the interested reader.

We recall two definitions of viscosity solutions of (1.1) which will be used in this paper. To minimize the technicalities we will assume that viscosity sub-/ supersolutions are in $\mathrm{BUC}(\mathbb{R}^d)$. The same results could be obtained assuming that they are just bounded and continuous in $\mathbb{R}^d$.

Definition 4.1. Let $p \in [1,2]$. A function $u \in \mathrm{BUC}(\mathbb{R}^d)$ is a viscosity subsolution of (1.1) if whenever $u - \varphi$ has a global maximum over $\mathbb{R}^d$ at $x \in \mathcal{O}$ for some $\varphi \in C_b^2(\mathbb{R}^d)$ and $\varphi(x) = u(x)$, then $I(\varphi, x) \leq 0$. A function $u \in \mathrm{BUC}(\mathbb{R}^d)$ is a viscosity supersolution of (1.1) if whenever $u - \varphi$ has a global minimum over $\mathbb{R}^d$ at $x \in \mathcal{O}$ for some $\varphi \in C_b^2(\mathbb{R}^d)$ and $\varphi(x) = u(x)$, then $I(\varphi, x) \geq 0$. A function u is a viscosity solution of (1.1) if it is both a viscosity subsolution and a viscosity supersolution of (1.1).

Definition 4.2. Let $p \in [1,2]$. A function $u \in \mathrm{BUC}(\mathbb{R}^d)$ is a viscosity subsolution of (1.1) if whenever $u - \varphi$ has a global maximum over $\mathbb{R}^d$ at $x \in \mathcal{O}$ for some $\varphi \in C^2(\mathbb{R}^d)$, then for every $0 < \delta < 1$

$$\begin{aligned} \sup_{\alpha\in\mathcal{A}} \inf_{\beta\in\mathcal{B}} \Big\{ &- \int_{|z|<\delta} \delta\varphi(x,z) d\mu_x^{\alpha\beta}(z) \\ &- \int_{|z|\geq\delta} \big[u(x+z) - u(x) - \chi_{B_1(0)}(z) D\varphi(x)\cdot z\big] \, d\mu_x^{\alpha\beta}(z) \\ &+ c_{\alpha\beta}(x)u(x) + f_{\alpha\beta}(x)\Big\} \leq 0. \end{aligned}$$

A function $u \in \mathrm{BUC}(\mathbb{R}^d)$ is a viscosity supersolution of (1.1) if whenever $u - \varphi$ has a global minimum over $\mathbb{R}^d$ at $x \in \mathcal{O}$ for some $\varphi \in C^2(\mathbb{R}^d)$, then for every $0 < \delta < 1$

$$\begin{aligned} \sup_{\alpha\in\mathcal{A}} \inf_{\beta\in\mathcal{B}} \Big\{ &- \int_{|z|<\delta} \delta\varphi(x,z) d\mu_x^{\alpha\beta}(z) \\ &- \int_{|z|\geq\delta} \big[u(x+z) - u(x) - \chi_{B_1(0)}(z) D\varphi(x)\cdot z\big] \, d\mu_x^{\alpha\beta}(z) \\ &+ c_{\alpha\beta}(x)u(x) + f_{\alpha\beta}(x)\Big\} \geq 0. \end{aligned}$$

A function u is a viscosity solution of (1.1) if it is both a viscosity subsolution and a viscosity supersolution of (1.1).

We remark that, since the Lévy measures $\mu_x^{\alpha\beta}$ are in $\mathbb{L}_p(\mathbb{R}^d)$, we could use test functions in $C_b^p(\mathbb{R}^d)$ and $C^p(\mathbb{R}^d)$ instead of test functions in $C_b^2(\mathbb{R}^d)$ and $C^2(\mathbb{R}^d)$. However, it is not clear whether such definitions and the standard definitions provided above are equivalent under general assumptions. It is easy to see, however, that they are equivalent for the most common measures considered in Example 5.12.

Below we show that Definitions 4.1 and 4.2 are equivalent to each other.

Proposition 4.3. *Under the assumptions of this paper, Definitions* 4.1 *and* 4.2 *are equivalent.*

Proof. We consider only the case of subsolutions. It is obvious that if u is a viscosity subsolution in the sense of Definition 4.2, then it is a viscosity subsolution in the sense of Definition 4.1. Let now u be a viscosity subsolution in the sense of Definition 4.1. It is easy to see that without loss of generality all maxima/minima in both definitions can be assumed to be strict. So let $u-\varphi$ have a strict global maximum over $\mathbb{R}^d$ at $x\in\mathcal{O}$ for some $\varphi\in C^2(\mathbb{R}^d)$, and we can obviously require that $\varphi(x)=u(x)$. Let $\varphi_n\in C_b^2(\mathbb{R}^d)$ be functions such that $u\le\varphi_n\le\varphi$ on $\mathbb{R}^d$, $\varphi_n(x)=u(x)$, $D\varphi_n(x)=D\varphi(x)$, and $\varphi_n\to u$ as $n\to+\infty$ uniformly on $\mathbb{R}^d$. Then

$$\begin{aligned}
&\sup_{\alpha\in\mathcal{A}}\inf_{\beta\in\mathcal{B}}\Big\{-\int_{|z|<\delta}\delta\varphi(x,z)d\mu_x^{\alpha\beta}(z)\\
&\qquad\qquad-\int_{|z|\ge\delta}\big[u(x+z)-u(x)-\chi_{B_1(0)}(z)D\varphi(x)\cdot z\big]\,d\mu_x^{\alpha\beta}(z)\\
&\qquad\qquad+c_{\alpha\beta}(x)u(x)+f_{\alpha\beta}(x)\Big\}\\
&\quad=\lim_{n\to+\infty}\sup_{\alpha\in\mathcal{A}}\inf_{\beta\in\mathcal{B}}\Big\{-\int_{|z|<\delta}\delta\varphi(x,z)d\mu_x^{\alpha\beta}(z)\\
&\qquad\qquad\qquad-\int_{|z|\ge\delta}\big[\varphi_n(x+z)-\varphi_n(x)-\chi_{B_1(0)}(z)D\varphi_n(x)\cdot z\big]\\
&\qquad\qquad\qquad\cdot d\mu_x^{\alpha\beta}(z)+c_{\alpha\beta}(x)\varphi_n(x)+f_{\alpha\beta}(x)\Big\}\\
&\qquad\qquad\qquad\le I(\varphi_n,x)\le 0.
\end{aligned}$$

□

The main result of the paper is the following theorem.

Theorem 4.4. *Let Assumptions* A–E *hold for* $p\in[1,2]$. *Then the comparison principle holds for equation* (1.1). *That is, if* u *and* v *are, respectively, a viscosity subsolution and a viscosity supersolution of* (1.1) *and* $u(x)\le v(x)$ *for all* $x\notin\mathcal{O}$, *then*

$$u(x)\le v(x)\qquad\forall x\in\mathcal{O}.$$

The following is a special case of Theorem 4.4, which we highlight to illustrate its scope (see Sections 5.1 and 6 for further examples).

Corollary 4.5. *Let Assumptions* C *and* D *be satisfied. Suppose that the measures* $\mu_x^{\alpha\beta}(z)$ *are of the form*

$$d\mu_x^{\alpha\beta}(z)=K_{\alpha\beta}(x,z)dz$$

and that, for some $\sigma\in(0,1)$,

$$\begin{gathered}
0\le K_{\alpha\beta}(x,z)\le K(z):=\Lambda_1|z|^{-(d+\sigma)},\\
|K_{\alpha\beta}(x,z)-K_{\alpha\beta}(y,z)|\le C|x-y|K(z).
\end{gathered}$$

Then in this case Assumption B *holds and Assumptions* A *and* E *hold with* $p=1$. *In particular, the comparison principle holds for equation* (1.1) *in this case.*

Theorem 4.4 and Corollary 4.5 will be proved in the next section. We also note that Theorem 4.4 essentially covers several of the results in [19], where only operators of order less than or equal to 1 are considered. However, it cannot be applied directly to the equations in [19] since the operators considered there had a slightly different form. This is discussed in greater detail in Examples 6.3 and 6.4.

5. Comparison principle

In this section we prove Theorem 4.4. A well-known property of sup-/inf- convolutions is that they produce approximations of viscosity sub- and supersolutions which enjoy one-sided regularity (semiconvexity and semiconcavity), which makes it easier—under the right circumstances—to evaluate the operator $I(\cdot,x)$ in the classical sense.

Remark 5.1. An approach to Theorem 4.4 that does not rely on such approximations can be found in Section 7, where we prove a comparison result (Theorem 7.4) under a different set of assumptions that are not amicable to such approximations. A posteriori, it became clear that the approach in Section 7 leads to a simpler proof of Theorem 4.4; however, we have decided to keep both approaches, as the tools developed in this section are of interest in many other situations. See Remark 7.2 for further comments.

Definition 5.2. Given $u,v \in \mathrm{BUC}(\mathbb{R}^d)$ and $0<\delta<1$, we define the sup-convolution u^δ of u and the inf-convolution v_δ of v by

$$u^\delta(x) = \sup_{y\in\mathbb{R}^d}\left\{u(y) - \frac{1}{\delta}|x-y|^2\right\},$$
$$v_\delta(x) = \inf_{y\in\mathbb{R}^d}\left\{v(y) + \frac{1}{\delta}|x-y|^2\right\}.$$

For the reader's convenience we review some well-known properties of the sup-/inf- convolutions in the following proposition.

Proposition 5.3. *The sup-convolutions and the inf-convolutions have the following properties:*

(1) *If* $\delta_1\le\delta_2$, *then* $u^{\delta_1}\le u^{\delta_2}$ *and* $v_{\delta_1}\ge v_{\delta_2}$. *Moreover,* $\|u^\delta\|_\infty\le\|u\|_\infty, \|v_\delta\|_\infty \le \|v\|_\infty$.
(2) $u^\delta(x)\ge u(x)$ *and* $v_\delta(x)\le v(x)$ *for all* $x\in\mathbb{R}^d$.
(3) $u^\delta\to u$ *and* $v_\delta\to u$ *uniformly on* $\mathbb{R}^d$ *as* $\delta\to 0$.
(4) *The function* u^δ *is semiconvex and for any* $x_0\in\mathbb{R}^d$, u^δ *is touched from below at* x_0 *by a function of the form*

$$u(x_0^*) - \tfrac{1}{\delta}|x-x_0^*|^2 \qquad \text{for some } x_0^*\in\mathbb{R}^d.$$

The function v_δ *is semiconcave, and, for any* $x_0\in\mathbb{R}^d$, v_δ *is touched from above at* x_0 *by a function of the form*

$$v(x_0^*) + \tfrac{1}{\delta}|x-x_0^*|^2 \qquad \text{for some } x_0^*\in\mathbb{R}^d.$$

(5) *Let* ω *be a modulus of continuity of* u. *For any* $x_0\in\mathbb{R}^d$ *and* $x_0^*\in\mathbb{R}^d$ *such that* $u^\delta(x_0)=u(x_0^*)-\frac{1}{\delta}|x_0-x_0^*|^2$, *we have*

$$|x_0-x_0^*|\le(2\delta\|u\|_\infty)^{1/2}$$

and

$$\tfrac{1}{\delta}|x_0 - x_0^*|^2 \le \omega\big((2\delta\|u\|_\infty)^{1/2}\big).$$

The analogous property holds for v_δ.

(6) *Let $\Omega \subset \mathbb{R}^d$, and let $h > 0$. If ω is a modulus of continuity for u in Ω, then for sufficiently small δ $\omega_1(s) = \max(\omega(s), \frac{2}{h}\|u\|_\infty s)$ is a modulus of continuity for u^δ in Ω_{2h}. A similar property holds for v_δ.*

Proof. To prove (1), note that if $\delta_1 \le \delta_2$, then $\frac{1}{\delta_1}|x-y|^2 \ge \frac{1}{\delta_2}|x-y|^2$ for all x and y, and thus $u^{\delta_1}(x) \le u^{\delta_2}(x)$ for all x. The respective statement for v_{δ_1} and v_{δ_2} is proved in the same way. Property (2) is obvious from the definitions. Property (3) follows from (2) and (5).

Regarding (4), we note that the semiconvexity follows from the fact that $u^\delta(x) + \frac{1}{\delta}|x|^2$ is the supremum of affine functions and is hence convex. If we fix x_0 and if x_0^* is such that

$$u^\delta(x_0) = u(x_0^*) - \frac{1}{\delta}|x_0 - x_0^*|^2,$$

then for all other x we have $u^\delta(x) \ge P(x) := u(x_0^*) - \frac{1}{\delta}|x - x_0^*|^2$ by the definition of u^δ, so P is the desired paraboloid. To prove (5), let x_0 and x_0^* be as above. Then

$$\tfrac{1}{\delta}|x_0 - x_0^*|^2 = u(x_0^*) - u^\delta(x_0) \le u(x_0^*) - u(x_0) \le 2\|u\|_\infty,$$

so

$$|x_0 - x_0^*| \le (2\delta\|u\|_\infty)^{1/2}.$$

This means that $u(x_0^*) - u(x_0)$ is in fact bounded from above by $\omega\big((2\delta\|u\|_\infty)^{1/2}\big)$, which gives (5).

Finally, to show (6), we observe that if $x, y \in \Omega_{2h}$ and $u^\delta(x) = u(x^*) - \frac{1}{\delta}|x - x^*|^2$, then for small δ $x \in \Omega_h$. Now if $|y-x| < h$, we have $u^\delta(y) \ge u(x^*+y-x) - \frac{1}{\delta}|x-x^*|^2$, so

$$u^\delta(x) - u^\delta(y) \le u(x^*) - u(x^* + y - x) \le \omega(|x-y|).$$

If $|y-x| \ge h$, then obviously $u^\delta(x) - u^\delta(y) \le \frac{2}{h}\|u\|_\infty|y-x|$. □

Definition 5.4. Given $y \in \mathcal{O}$ and the operator $I(\cdot, x)$ from (1.1), we define

$$I^{(y)}(\phi, x) = \sup_\alpha \inf_\beta \Big\{ -L_{\mu_y^{\alpha\beta}}(\phi, x) + c_{\alpha\beta}(y)\phi(x) + f_{\alpha\beta}(y) \Big\}, \tag{5.1}$$

where

$$L_{\mu_y^{\alpha\beta}}(\phi, x) = \int_{\mathbb{R}^d} [\phi(x+z) - \phi(x) - \chi_{B_1(0)}(z) D\phi(x)\cdot z]\, d\mu_y^{\alpha\beta}(z). \tag{5.2}$$

Note that this last expression is almost identical to $L^{\alpha\beta}(\phi, x)$, except that the Lévy measure used is the one corresponding to the point y. Moreover, the coefficients in (5.1) are evaluated at y.

In the rest of this section, unless stated otherwise, we will always assume that Assumptions A–E are satisfied.

Proposition 5.5. *If u is a viscosity subsolution of $I(u, x) = 0$ in $\mathcal{O}$, then u^δ is a viscosity subsolution of $I_\delta(u^\delta, x) = 0$ in $\mathcal{O}_h$, $h = (2\delta\|u\|_\infty)^{1/2}$, where*

$$I_\delta(\phi, x) := \inf \Big\{ I^{(y)}(\phi, x) : |y - x| \le h \Big\}.$$

If v is a viscosity supersolution of $I(v,x)=0$ in $\mathcal{O}$, then v_δ is a viscosity supersolution of $I^\delta(v_\delta,x)=0$ in $\mathcal{O}_h$, $h=(2\delta\|v\|_\infty)^{1/2}$, where

$$I^\delta(\phi,x) := \sup\left\{ I^{(y)}(\phi,x) : |y-x|\le h\right\}.$$

Proof. Let us prove the statement for u and I_δ (the corresponding one for v and I^δ is entirely analogous, and we omit it). Let ϕ touch u^δ from above at some $x_0\in\mathcal{O}_h$. Let $x_0^*\in\mathbb{R}^d$ be such that

$$u^\delta(x_0) = u(x_0^*) - \frac{1}{\delta}|x_0-x_0^*|^2.$$

It follows from part (5) of Proposition 5.3 that $x_0^*\in\mathcal{O}$. Then, by the definition of u^δ, for any x and y we have

$$u^\delta(x+x_0-x_0^*) \ge u(y) - \frac{1}{\delta}|x+x_0-x_0^*-y|^2.$$

Choosing $y=x$, it follows that for every x we have

$$u^\delta(x+x_0-x_0^*) \ge u(x) - \frac{1}{\delta}|x_0-x_0^*|^2,$$

with equality for $x=x_0^*$. It follows that if define a new test function $\phi^*(x)$ by

$$\phi^*(x) = \phi(x+x_0-x_0^*) + \frac{1}{\delta}|x_0-x_0^*|^2,$$

then ϕ^* touches u from above at x_0^*. Since u is a subsolution, it follows that

$$I(\phi^*,x_0^*)\le 0.$$

Let us rewrite the expression on the left. First, recall

$$I(\phi^*,x_0^*) = \sup_\alpha\inf_\beta\{-L^{\alpha\beta}(\phi^*,x_0^*) + c_{\alpha\beta}(x_0^*)\phi^*(x_0^*) + f_{\alpha\beta}(x_0^*)\}.$$

Next, note that

$$L^{\alpha\beta}(\phi^*,x_0^*) = \int_{\mathbb{R}^d}[\phi^*(x_0^*+z)-\phi^*(x_0^*)-\chi_{B_1(0)}D\phi^*(x_0^*)\cdot z]\,d\mu^{\alpha\beta}_{x_0^*}(z).$$

Since

$$\phi^*(x_0^*+z)-\phi^*(x_0^*)-\chi_{B_1(0)}D\phi^*(x_0^*)\cdot z = \phi(x_0+z)-\phi(x_0)-\chi_{B_1(0)}D\phi(x_0)\cdot z,$$

it follows that

$$L^{\alpha\beta}(\phi^*,x_0^*) = \int_{\mathbb{R}^d}[\phi(x_0+z)-\phi(x_0)-\chi_{B_1(0)}D\phi(x_0)\cdot z]\,d\mu^{\alpha\beta}_{x_0^*}(z) = L_{\mu^{\alpha\beta}_{x_0^*}}(\phi,x_0).$$

In conclusion,

$$\begin{aligned}0\ge I(\phi^*,x_0^*) &= \sup_\alpha\inf_\beta\{-L_{\mu^{\alpha\beta}_{x_0^*}}(\phi,x_0) + c_{\alpha\beta}(x_0^*)(\phi(x_0)+\tfrac{1}{\delta}|x_0-x_0^*|^2) + f_{\alpha\beta}(x_0^*)\}\\ &\ge \sup_\alpha\inf_\beta\{-L_{\mu^{\alpha\beta}_{x_0^*}}(\phi,x_0) + c_{\alpha\beta}(x_0^*)\phi(x_0) + f_{\alpha\beta}(x_0^*)\}\\ &\ge I_\delta(\phi,x_0).\end{aligned}$$

Using part (5) of Proposition 5.3 in this last inequality, the proposition follows. □

Let us also state in a single lemma two basic facts about classical evaluation of Lévy operators and viscosity solutions. The proof of the lemma goes along lines similar to those of the proofs of [9, Lemmas 4.3 and 5.7].

Lemma 5.6. *For any function $u \in \mathrm{BUC}(\mathbb{R}^d)$ that is pointwise-$C^{1,1}$ at a point $x_0 \in \mathcal{O}$ (resp., $x_0 \in \mathcal{O}_h$) the operator $I(u, x_0)$ (resp., $I_\delta(u, x_0)$) is classically defined. If furthermore u is a viscosity subsolution of $I(u, x) = 0$ in $\mathcal{O}$ (resp., $I_\delta(u, x) \leq 0$ in $\mathcal{O}_h$), then also $I(u, x_0) \leq 0$ (resp., $I_\delta(u, x_0) \leq 0$) pointwise. A similar statement is true for viscosity supersolutions.*

Proof. We will prove the statement only for $I(u, x_0)$, as the other statements are proved similarly. Recall that from Assumption E

$$\Lambda = \sup_{x,\alpha,\beta} \{\mathcal{N}_2(\hat{\mu}_x^{\alpha\beta}) + \check{\mu}_x^{\alpha\beta}(B_1^c)\}.$$

From the pointwise-$C^{1,1}$ assumption at x_0, we have

$$\int_{B_1} |u(x_0 + z) - u(x_0) - \chi_{B_1(0)}(z) Du(x_0) \cdot z| \, d\mu_{x_0}^{\alpha\beta}(z) \leq \int_{B_1} C_{u,x_0} |z|^2 \, d\mu_{x_0}^{\alpha\beta}(z),$$
$$\int_{B_1^c} |u(x_0 + z) - u(x_0)| \, d\mu_{x_0}^{\alpha\beta}(z) \leq 2\|u\|_\infty \int_{B_1^c} d\mu_{x_0}^{\alpha\beta}(z),$$

where C_{u,x_0} is from Definition 2.3. It thus follows that each integral defining $L^{\alpha\beta}(u, x_0)$ converges and

$$\sup_{\alpha\beta} |L^{\alpha\beta}(u, x_0)| \leq (C_{u,x_0} + 2\|u\|_\infty)\Lambda < \infty.$$

From here, it is immediate that $I(u, x_0)$ is classically defined. As for the second assertion, define

$$u_r := \begin{cases} \phi & \text{in } B_r(x_0), \\ u & \text{outside of } B_r(x_0), \end{cases}$$

where $\phi(x) = u(x_0) + Du(x_0) \cdot (x - x_0) + C_{u,x_0}|x - x_0|^2$. The function ϕ is touching u from above in a neighborhood of x_0. From Definition 4.2 we have $I(u_r, x_0) \leq 0$ for every $r > 0$. On the other hand,

$$I(u, x_0) \leq I(u_r, x_0) + \mathcal{M}_I^+(u - u_r, x_0),$$

where the operator $\mathcal{M}_I^+$ is given by

$$\mathcal{M}_I^+(u - u_r, x_0) = \sup_{\alpha,\beta} \left\{-L^{\alpha\beta}(u - u_r, x_0)\right\}.$$

Using the special form of u_r, particularly that $u_r = u$ outside of B_r, we have Q4

$$\begin{aligned} \mathcal{M}_I^+(u - u_r, x_0) &= \sup_{\alpha,\beta} \left\{-L^{\alpha\beta}(u - u_r, x_0)\right\} \\ &= \sup_{\alpha,\beta} \left\{ \int_{B_r} [\phi(x + z) - u(x + z)] \, d\mu_{x_0}^{\alpha\beta}(z) \right\} \\ &\leq 2C_{u,x_0} \sup_{\alpha,\beta} \left\{ \int_{B_r} |z|^2 \, d\mu_{x_0}^{\alpha\beta}(z) \right\} \leq \theta(r), \end{aligned}$$

where the last inequality follows from (4.6). Taking the limit as $r \to 0$, we conclude that

$$I(u, x_0) \leq 0.$$

□

We will need smooth approximations of functions $|x-y|^p$ for $p \in [1,2]$. For $\kappa > 0$ we define a function $\tilde{\psi}_\kappa : [0,+\infty) \to [0,+\infty)$ by

$$\tilde{\psi}_\kappa(r) = \left(\kappa + r^2\right)^{\frac{p}{2}} - \kappa^{\frac{p}{2}}.$$

Then the function

$$\psi_\kappa(x) := \tilde{\psi}_\kappa(|x|)$$

is smooth and converges as $\kappa \to 0$ to $|x|^p$ uniformly on $\mathbb{R}^d$. We will be using the following lemma.

Lemma 5.7. *Let $p \in [1,2]$. For every $R > 0$ the function $\psi_\kappa(x)$ is uniformly pointwise-C^p on B_R, i.e., there exists a constant $C_{p,R}$ such that, for every $0 < \kappa < 1$ and every $x_0, x \in B_R$,*

$$|\psi_\kappa(x) - \psi_\kappa(x_0) - D\psi_\kappa(x_0) \cdot (x - x_0)| \leq C_{p,R}|x - x_0|^p \qquad \text{if } 1 < p \leq 2,$$

$$|\psi_\kappa(x) - \psi_\kappa(x_0)| \leq C_{1,R}|x - x_0| \qquad \text{if } p = 1.$$

The following is the main lemma of the paper. We refer the reader to Definition 2.2 for the definition of L_μ.

Lemma 5.8. *Let $u, v \in \mathrm{BUC}(\mathbb{R}^d)$. Let $\alpha > 0$, $p \geq 1$, $0 < \kappa < 1$, and suppose that $(x_*, y_*) \in \mathcal{O} \times \mathcal{O}$ is a global maximum point of the function*

$$w(x,y) := u(x) - v(y) - \alpha\psi_\kappa(x - y).$$

Furthermore, suppose that u and v are pointwise-$C^{1,1}$ at x_* and y_*, respectively. Then, for any two Lévy measures $\mu, \nu \in \mathbb{L}_p(B_1)$, we have the inequality

$$L_\mu(u, x_*) - L_\nu(v, y_*) \leq C_p \alpha \mathrm{d}_{\mathbb{L}_p}(\mu, \nu)^p,$$

where C_p is independent of κ.

Proof. First, note that as (x_*, y_*) is a maximum point of w, we have

$$\begin{aligned} u(x) &\leq \alpha\psi_\kappa(x - y_*) + v(y_*) + (u(x_*) - v(y_*) - \alpha\psi_\kappa(x_* - y_*)), \\ v(y) &\geq -\alpha\psi_\kappa(x_* - y) + u(x_*) - (u(x_*) - v(y_*) - \alpha\psi_\kappa(x_* - y_*)), \end{aligned}$$

with equalities at x_* and y_*, respectively. Second, for any $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$

$$w(x_* + x, y_* + y) - w(x_*, y_*) \leq 0.$$

Let $\gamma \in \mathrm{Adm}(\mu, \nu)$. Using that $\gamma((B_1 \times B_1)^c) = 0$, and since $\delta u(x_*, 0) = 0$ and $\delta v(y_*, 0) = 0$, we thus have

$$\begin{aligned} &L_\mu(u, x_*) - L_\nu(v, y_*) \\ &\quad = \int_{B_1 \times B_1} \Big(u(x_* + x) - v(y_* + y) - (u(x_*) - v(y_*)) \\ &\qquad\qquad - \alpha D\psi_\kappa(x_* - y_*) \cdot (x - y) \Big) d\gamma(x,y). \end{aligned}$$

On the other hand, if $x, y \in B_1$, using Lemma 5.7, we also have

$$\begin{aligned} &u(x_* + x) - v(y_* + y) - (u(x_*) - v(y_*)) - \alpha D\psi_\kappa(x_* - y_*) \cdot (x - y) \\ &\quad \leq \alpha\psi_\kappa(x_* + x - y - y_*) - \alpha\psi_\kappa(x_* - y_*) - \alpha D\psi_\kappa(x_* - y_*) \cdot (x - y) \\ &\quad \leq C_p \alpha |x - y|^p. \end{aligned}$$

Therefore,

$$L_\mu(u,x_*) - L_\nu(v,y_*) \le C_p\alpha \int_{B_1\times B_1} |x-y|^p d\gamma(x,y).$$

Taking the infimum over all $\gamma \in \mathrm{Adm}(\mu,\nu)$, it thus follows that

$$L_\mu(u,x_*) - L_\nu(v,y_*) \le C_p\alpha\, \mathrm{d}_{\mathbb{L}_p}(\mu,\nu)^p. \qquad \square$$

Corollary 5.9. *Let u, v, x_*, and y_* be as in Lemma 5.8, and let $\mu,\nu \in \mathbb{L}_p(\mathbb{R}^d)$. Then*

$$L_\mu(u,x_*) - L_\nu(v,y_*) \le C_p\alpha\, \mathrm{d}_{\mathbb{L}_p}(\hat\mu,\hat\nu)^p + 2\|v\|_\infty \mathrm{d}_{\mathrm{TV}}(\check\mu,\check\nu).$$

Proof. Let us write the difference as follows:

$$L_\mu(u,x_*) - L_\nu(v,y_*) = L_{\hat\mu}(u,x_*) - L_{\hat\nu}(v,y_*) + L_{\check\mu}(u,x_*) - L_{\check\nu}(v,y*).$$

Thanks to Lemma 5.8, the first difference on the right-hand side above is less than or equal to $C_p\alpha\, \mathrm{d}_{\mathbb{L}_p}(\hat\mu,\hat\nu)^p$. For the second one note that

$$\begin{aligned} L_{\check\mu}(u,x_*) - L_{\check\nu}(v,y_*) &= \int_{B_1^c} [u(x_*+z) - u(x_*)]\, d\mu(z) - \int_{B_1^c} [v(y_*+z) - v(y_*)]\, d\nu(z) \\ &= \int_{B_1^c} [u(x_*+z) - v(x_*+z) - (u(x_*) - v(y_*))]\, d\mu(z) \\ &+ \int_{B_1^c} [v(y_*+z) - v(y_*)]\, d(\mu-\nu)(z). \end{aligned}$$

Since w achieves its global maximum at (x_*, y_*), it follows that $u(x_*+z) - v(y_*+z) - (u(x_*) - v(y_*)) \le 0$. Hence we obtain

$$\begin{aligned} L_{\check\mu}(u,x_*) - L_{\check\nu}(v,y_*) &\le \int_{B_1^c} [v(y_*+z) - v(y_*)]\, d(\mu-\nu)(z) \\ &\le 2\|v\|_\infty \mathrm{d}_{\mathrm{TV}}(\check\mu,\check\nu). \end{aligned} \qquad \square$$

We need a variant of a well-known doubling lemma (see, e.g., [12, Lemma 3.1]).

Lemma 5.10. *Let $u, v \in \mathrm{BUC}(\mathbb{R}^d)$ be such that $M = \sup(u-v) > \tau > 0$ and $u(x) - v(x) \le 0$ for $x \in B_R^c$ for some $R > 0$. For any $\varepsilon, \delta, \kappa > 0$ set*

$$\begin{aligned} w(x,y) &:= u^\delta(x) - v_\delta(y) - \frac{1}{\varepsilon}\psi_\kappa(x-y), \\ M_{\varepsilon,\delta,\kappa} &:= \sup_{\mathbb{R}^d\times\mathbb{R}^d} w(x,y). \end{aligned}$$

Then, for sufficiently small $\delta, \varepsilon, \kappa$, there exist $(x_\varepsilon, y_\varepsilon)$ such that

$$M_{\varepsilon,\delta,\kappa} = w(x_\varepsilon, y_\varepsilon).$$

Then we have

$$\begin{aligned} &\frac{|x_\varepsilon - y_\varepsilon|^p}{\varepsilon} \le \omega((C\varepsilon + c_{\kappa,\varepsilon,\delta})^{1/p}) + \frac{c_{\kappa,\varepsilon,\delta}}{\varepsilon}, \\ &\lim_{\varepsilon\to 0}\lim_{\delta\to 0}\lim_{\kappa\to 0} M_{\varepsilon,\delta,\kappa} = M, \end{aligned} \tag{5.3}$$

where above $C = \|u\|_\infty + \|v\|_\infty$ ω is a modulus of continuity of u, and where $c_{\kappa,\varepsilon,\delta}$ is a constant that converges to 0 uniformly in ε and δ as $\kappa \to 0^+$.

If Ω is an open subset of $\mathbb{R}^d$ and in addition $u \in C^{0,r}(\Omega), 0 < r \leq 1$, and if all of the points $x_\varepsilon, y_\varepsilon \in \Omega$, then

$$\limsup_{\kappa\to 0} \frac{|x_\varepsilon - y_\varepsilon|^{p-r}}{\varepsilon} \leq C_1 \tag{5.4}$$

for some constant C_1 independent of $\delta, \varepsilon, \kappa$.

Proof. It is easy to see that the uniform convergence of the u^δ, v_δ to u, v, the uniform convergence of $\psi_\kappa(x-y)$ to $|x-y|^p$, and the uniform continuity of u, v (and hence of u^δ, v_δ, uniform in δ) implies that, for sufficiently small $\delta, \varepsilon, \kappa$, we must have $w(x, y) \leq \tau/2$ when either x or y is in B_R^c. Thus w must attain maximum at some point $(x_\varepsilon, y_\varepsilon) \in B_R \times B_R$.

Denote

$$M_{\varepsilon,\delta} := \sup_{\mathbb{R}^d\times\mathbb{R}^d} (u^\delta(x) - v_\delta(y) - \frac{1}{\varepsilon}|x-y|^p),$$

$$M_\varepsilon := \sup_{\mathbb{R}^d\times\mathbb{R}^d} (u(x) - v(y) - \frac{1}{\varepsilon}|x-y|^p).$$

Again, using the uniform convergence of $u^\delta, v_\delta, \psi_\kappa(x-y)$ and the uniform continuity of u, v, we easily find (see also [12, proof of Lemma 3.1]) that

$$\lim_{\kappa\to 0} M_{\varepsilon,\delta,\kappa} = M_{\varepsilon,\delta}, \quad \lim_{\delta\to 0} M_{\varepsilon,\delta} = M_\varepsilon, \quad \lim_{\varepsilon\to 0} M_\varepsilon = M.$$

We obviously have

$$\frac{1}{\varepsilon}\psi_\kappa(x_\varepsilon - y_\varepsilon) + \frac{c_{\kappa,\varepsilon,\delta}}{\varepsilon} = \frac{1}{\varepsilon}|x_\varepsilon - y_\varepsilon|^p, \tag{5.5}$$

where $c_{\kappa,\varepsilon,\delta}$ is a constant which converges to 0 uniformly in ε and δ as $\kappa \to 0^+$. Now

$$u^\delta(y_\varepsilon) - v_\delta(y_\varepsilon) \leq u^\delta(x_\varepsilon) - v_\delta(y_\varepsilon) - \frac{1}{\varepsilon}\psi_\kappa(x_\varepsilon - y_\varepsilon),$$

which, by (5.5), implies

$$\frac{|x_\varepsilon - y_\varepsilon|^p}{\varepsilon} \leq u^\delta(x_\varepsilon) - u^\delta(y_\varepsilon) + \frac{c_{\kappa,\varepsilon,\delta}}{\varepsilon} \leq \omega(|x_\varepsilon - y_\varepsilon|) + \frac{c_{\kappa,\varepsilon,\delta}}{\varepsilon}.$$

This, together with the fact that we must have

$$\frac{|x_\varepsilon - y_\varepsilon|^p}{\varepsilon} \leq \|u\|_\infty + \|v\|_\infty + \frac{c_{\kappa,\varepsilon,\delta}}{\varepsilon},$$

gives (5.3). The last claim, (5.4), follows by a similar argument since now

$$\begin{aligned}\frac{1}{\varepsilon}(\limsup_{\kappa\to 0}|x_\varepsilon - y_\varepsilon|)^p = \limsup_{\kappa\to 0}\frac{|x_\varepsilon - y_\varepsilon|^p}{\varepsilon} &\leq \limsup_{\kappa\to 0} C|x_\varepsilon - y_\varepsilon|^r \\ &= C(\limsup_{\kappa\to 0}|x_\varepsilon - y_\varepsilon|)^r.\end{aligned}$$ □

Proof of Theorem 4.4. Arguing by contradiction, assume that there is some $\ell > 0$ such that

$$\sup_{x\in\mathbb{R}^d}\{u(x) - v(x)\} = \ell > 0.$$

Step 1 (Taking inf-/sup- convolutions). Let u^δ and v_δ denote the sup- and inf-convolutions of u and v for $\delta > 0$. Then

$$\sup_{x\in\mathbb{R}^d}\{u^\delta(x) - v_\delta(x)\} \geq \ell.$$

We may make δ_0 small enough so that for $\delta < \delta_0$ we have

$$\sup_{x\notin\mathcal{O}}\{u^\delta(x) - v_\delta(x)\} \leq \tfrac{1}{4}\ell.$$

Recall that if ω is a modulus of continuity of u and v, then it is also a modulus of continuity of u^δ and v_δ. Therefore, reducing δ_0 if necessary, we have

$$u^\delta(x) - v_\delta(x) \leq \tfrac{1}{2}\ell \qquad \text{for } x \in \mathcal{O}\setminus\mathcal{O}_{2h_0},$$

as long as $\delta < \delta_0$, where $\mathcal{O}_{h_0} = \{x \in \mathcal{O} \mid d(x,\partial\mathcal{O}) > h_0\}$ and $h_0 > 0$ is some constant. In particular, for such δ the supremum of $u^\delta - v_\delta$ in $\mathbb{R}^d$ can be achieved only within $\overline{\mathcal{O}}_{2h_0}$.

Step 2 (Doubling of variables). For $\varepsilon, \delta, \kappa > 0$ we let w be as in Lemma 5.10 and let $(x_\varepsilon, y_\varepsilon) \in \mathbb{R}^d \times \mathbb{R}^d$ be such that

$$w(x_\varepsilon, y_\varepsilon) = \max_{\mathbb{R}^d\times\mathbb{R}^d} w(x,y).$$

From Step 1 we know that $u^\delta - v_\delta \leq \ell/2$ in $\mathcal{O}\setminus\mathcal{O}_{2h_0}$ and $u^\delta - v_\delta \geq \ell$ somewhere in $\mathcal{O}$. Furthermore, we know that u^δ and v_δ are uniformly continuous in $\mathcal{O}$, and uniformly so with respect to $\delta < 1$. From these facts and (5.3), it follows that $(x_\varepsilon, y_\varepsilon)$ must belong to $\mathcal{O}_{h_0}\times\mathcal{O}_{h_0}$ for all sufficiently small ε, δ, and κ or else it cannot be the maximum point of w^ε.

On the other hand, Proposition 5.5 says that u^δ is a viscosity subsolution of $I_\delta(u^\delta, x) = 0$, and that v_δ is a viscosity supersolution of $I^\delta(v_\delta, x) = 0$ in $\mathcal{O}_{h_0}$ for sufficiently small δ. The function u^δ is touched from above by a smooth function at x_ε, and v_δ is touched from below at y_ε. It follows that u^δ and v_δ are pointwise-$C^{1,1}$ at x_ε and y_ε, respectively (see Definition 2.3). Applying Lemma 5.6, we conclude that $I_\delta(u^\delta, x_\varepsilon)$ and $I^\delta(v_\delta, y_\varepsilon)$ are well defined in the classical sense, with $I_\delta(u^\delta, x_\varepsilon) \leq 0$ and $I^\delta(v_\delta, x_\varepsilon) \leq 0$. It follows from Proposition 5.5 that there are points x^*_δ and y^*_δ such that

$$I^{(x^*_\delta)}(u^\delta, x_\varepsilon) \leq \delta, \qquad I^{(y^*_\delta)}(v_\delta, y_\varepsilon) \geq -\delta,$$

and

$$|x_\varepsilon - x^*_\delta|, |y_\varepsilon - y^*_\delta| \leq h, \tag{5.6}$$

where $h = \big(2\delta(\|u\|_\infty + \|v\|_\infty)\big)^{\frac{1}{2}}$.

Step 3 (Equation structure). Let us use the structure of $I(\cdot, x)$ to bound $I^{(x^*_\delta)}$ $(u^\delta, x_\varepsilon) - I^{(y^*_\delta)}(v^\delta, y_\varepsilon)$ from below. Using the expression in (5.1), we have

$$I^{(x^*_\delta)}(u^\delta, x_\varepsilon) = \sup_\alpha \inf_\beta \left\{-L_{\mu^{\alpha\beta}_{x^*_\delta}}(u^\delta, x_\varepsilon) + c_{\alpha\beta}(x^*_\delta)u^\delta(x_\varepsilon) + f_{\alpha\beta}(x^*_\delta)\right\},$$

$$I^{(y^*_\delta)}(v_\delta, y_\varepsilon) = \sup_\alpha \inf_\beta \left\{-L_{\mu^{\alpha\beta}_{y^*_\delta}}(v_\delta, y_\varepsilon) + c_{\alpha\beta}(y^*_\delta)v_\delta(y_\varepsilon) + f_{\alpha\beta}(y^*_\delta)\right\}.$$

Therefore, for our purposes it suffices to compare the expressions appearing on the right-hand side for each fixed α, β. Let us write

$$\begin{aligned}(\mathrm{I})_{\alpha\beta} &= -L_{\mu^{\alpha\beta}_{x^*_\delta}}(u^\delta, x_\varepsilon) + c_{\alpha\beta}(x^*_\delta)u^\delta(x_\varepsilon) + f_{\alpha\beta}(x^*_\delta),\\ (\mathrm{II})_{\alpha\beta} &= -L_{\mu^{\alpha\beta}_{y^*_\delta}}(v_\delta, y_\varepsilon) + c_{\alpha\beta}(y^*_\delta)v_\delta(y_\varepsilon) + f_{\alpha\beta}(y^*_\delta).\end{aligned}$$

We now look for an upper bound for $(\mathrm{I})_{\alpha\beta} - (\mathrm{II})_{\alpha\beta}$ which is independent of α and β by breaking this difference into parts. First, recall that the function $u^\delta(x) - v_\delta(y) - \frac{1}{\varepsilon}\psi_\kappa(x-y)$ achieves its global maximum at $(x_\varepsilon, y_\varepsilon)$, in which case Corollary 5.9 guarantees that

$$L_{\mu^{\alpha\beta}_{x^*_\delta}}(u^\delta, x_\varepsilon) - L_{\mu^{\alpha\beta}_{y^*_\delta}}(v_\delta, y_\varepsilon) \leq \frac{C}{\varepsilon}\mathrm{d}_{\mathbb{L}_p}(\hat{\mu}^{\alpha\beta}_{x^*_\delta}, \hat{\mu}^{\alpha\beta}_{y^*_\delta})^p + 2\|v_\delta\|_\infty \mathrm{d}_{\mathrm{TV}}(\check{\mu}_{x^*_\delta}, \check{\mu}_{y^*_\delta}).$$

Then, thanks to Assumptions A and B and (5.6), we have

$$\begin{aligned}L_{\mu^{\alpha\beta}_{x^*_\delta}}(u^\delta, x_\varepsilon) - L_{\mu^{\alpha\beta}_{y^*_\delta}}(v_\delta, y_\varepsilon) &\leq \frac{C}{\varepsilon}|x^*_\delta - y^*_\delta|^p + 2\|v\|_\infty \theta(|x^*_\delta - y^*_\delta|)\\ &\leq \frac{C}{\varepsilon}|x_\varepsilon - y_\varepsilon|^p + 2\|v\|_\infty\theta(|x_\varepsilon - y_\varepsilon|) + \rho_\varepsilon(\delta),\end{aligned} \tag{5.7}$$

where for a fixed ε $\lim_{\delta\to 0}\rho_\varepsilon(\delta) = 0$.

Next, we have the elementary inequality

$$\begin{aligned}c_{\alpha\beta}(x^*_\delta)u^\delta(x_\varepsilon) - c_{\alpha\beta}(y^*_\delta)v_\delta(y_\varepsilon) &\geq c_{\alpha\beta}(x^*_\delta)(u^\delta(x_\varepsilon) - v_\delta(y_\varepsilon))\\ &\quad - |c_{\alpha\beta}(x^*_\delta) - c_{\alpha\beta}(y^*_\delta)||v_\delta(y_\varepsilon)|\\ &\geq \lambda\ell - \theta(|x_\varepsilon - y_\varepsilon|)\|v\|_\infty - \rho_\varepsilon(\delta),\end{aligned} \tag{5.8}$$

where $\rho_\varepsilon(\delta)$ is a function as before and we used that $u^\delta(x_\varepsilon) - v_\delta(y_\varepsilon) \geq \ell$.

Finally, by Assumption C

$$|f_{\alpha\beta}(x^*_\varepsilon) - f_{\alpha\beta}(y^*_\varepsilon)| \leq \theta(|x_\varepsilon - y_\varepsilon|) + \rho_\varepsilon(\delta). \tag{5.9}$$

Now, combining (5.7), (5.8), (5.9), we have the estimate

$$(\mathrm{I})_{\alpha\beta} - (\mathrm{II})_{\alpha\beta} \geq \lambda\ell - \frac{C}{\varepsilon}|x_\varepsilon - y_\varepsilon|^p - C\theta\big(|x_\varepsilon - y_\varepsilon|\big) - \rho_\varepsilon(\delta),$$

where C is some absolute constant. Therefore, we conclude that

$$I^{(x^*_\delta)}(u^\delta, x_\varepsilon) - I^{(y^*_\delta)}(v_\delta, y_\varepsilon) \geq \lambda\ell - \frac{C}{\varepsilon}|x_\varepsilon - y_\varepsilon|^p - C\theta\big(|x_\varepsilon - y_\varepsilon|\big) - \rho_\varepsilon(\delta).$$

Step 4 (Using the subsolution and supersolution property). Recalling the way x^*_δ and y^*_δ were selected, we have $I^{(x^*_\delta)}(u^\delta, x_\varepsilon) - I^{(y^*_\delta)}(v_\delta, y_\varepsilon) \leq 2\delta$, and therefore

$$\lambda\ell \leq 2\delta + \frac{C}{\varepsilon}|x_\varepsilon - y_\varepsilon|^p + C\theta\big(|x_\varepsilon - y_\varepsilon|\big) + \rho_\varepsilon(\delta).$$

It now remains to take $\lim_{\varepsilon\to 0}\lim_{\delta\to 0}\limsup_{\kappa\to 0}$ on both sides of the above inequality and use (5.3) to obtain a contradiction. □

5.1. **Estimating $\mathrm{d}_{\mathbb{L}_p}$ in special cases.**

Proposition 5.11. *Let $p \in [1,2]$.*

(i) *Let $\mu, \nu \in \mathbb{L}_p(B_1)$, and let $\phi, \psi \in \mathrm{Adm}^p$. If $\mu - \nu$ is a positive measure, then*

$$\int_{B_1} \phi(x) d\mu(x) + \int_{B_1} \psi(y) d\nu(y) \leq \int_{B_1} |x|^p d(\mu - \nu)(x). \tag{5.10}$$

(ii) *For any $\mu, \nu \in \mathbb{L}_p(B_1)$ we have*

$$\mathrm{d}_{\mathbb{L}_p}(\mu, \nu) \leq 2^{\frac{p-1}{p}} d_{\mathrm{TV}}(\mu_p, \nu_p)^{\frac{1}{p}}, \tag{5.11}$$

where $d\mu_p = |x|^p d\mu, d\mu_p = |x|^p d\mu$.

Proof.

(i) Since $\phi \in L^1(\mu)$, we also have $\phi \in L^1(\nu)$. Then we may write

$$\begin{aligned}
&\int_{B_1} \phi(x) d\mu(x) + \int_{B_1} \psi(y) d\nu(y) \\
&\quad = \int_{B_1} \phi(x) d\mu(x) - \int_{B_1} \phi(x) d\nu(x) + \int_{B_1} \phi(y) d\nu(y) + \int_{B_1} \psi(y) d\nu(y) \\
&\quad = \int_{B_1} \phi(x) d(\mu - \nu)(x) + \int_{B_1} (\phi(y) + \psi(y)) d\nu(y) \\
&\quad \leq \int_{B_1} \phi(x) d(\mu - \nu)(x) \leq \int_{B_1} |x|^p d(\mu - \nu)(x),
\end{aligned}$$

where in the last line we used $\phi(y) + \psi(y) \leq |y - y|^p = 0 \ \ \forall y \in B_1$ and $\phi(x) \leq |x|^p \ \ \forall x \in B_1$.

(ii) Denoting by $(\mu-\nu)^+$ and $(\mu-\nu)^-$, the positive and negative parts of $\mu-\nu$, we have $|\mu-\nu| = (\mu-\nu)^+ + (\mu-\nu)^-$. We also notice that $\mu-(\mu-\nu)^+ = \nu-(\mu-\nu)^-$. It thus follows from (5.10) and Lemma A.14 that

$$\mathrm{d}_{\mathbb{L}_p}(\mu, \mu - (\mu - \nu)^+)^p \leq \int_{B_1} |x|^p d(\mu - \nu)^+(x),$$

$$\mathrm{d}_{\mathbb{L}_p}(\nu, \mu - (\mu - \nu)^+)^p \leq \int_{B_1} |x|^p d(\mu - \nu)^-(x).$$

Moreover, it is obvious that

$$d_{\mathrm{TV}}(\mu_p, \nu_p) = \int_{B_1} |x|^p d(\mu - \nu)^+(x) + \int_{B_1} |x|^p d(\mu - \nu)^-(x).$$

Therefore, using the triangle inequality for the distance and the inequality $a + b \leq 2^{\frac{p-1}{p}} (a^p + b^p)^{\frac{1}{p}}$ for $a, b \geq 0$, we obtain

$$\begin{aligned}
\mathrm{d}_{\mathbb{L}_p}(\mu, \nu) &\leq \mathrm{d}_{\mathbb{L}_p}(\mu, \mu - (\mu - \nu)^+) + \mathrm{d}_{\mathbb{L}_p}(\nu, \mu - (\mu - \nu)^+) \\
&\leq 2^{\frac{p-1}{p}} \left(\int_{B_1} |x|^p d(\mu - \nu)^+(x) + \int_{B_1} |x|^p d(\mu - \nu)^-(x) \right)^{\frac{1}{p}} \\
&= 2^{\frac{p-1}{p}} d_{\mathrm{TV}}(\mu_p, \nu_p)^{\frac{1}{p}}.
\end{aligned}$$

□

Let us now discuss the case when the Lévy measures $\mu_x^{\alpha\beta}$ are absolutely continuous with respect to the Lebesgue measure.

Example 5.12. Let us consider operators whose Lévy measures $d\mu_x^{\alpha\beta}(z)$ are all of the form $K_{\alpha\beta}(x,z)dz$. Assumption E holds, for instance, if

$$0 \le K_{\alpha\beta}(x,z) \le K(z), \tag{5.12}$$

where $K(z)$ is such that for some $p \in [1,2]$

$$\int_{\mathbb{R}^d} \min(1,|z|^p)K(z)dz < +\infty. \tag{5.13}$$

Regarding Assumption A, suppose that there are some $\gamma \in (0,1]$ and $C \ge 0$ such that

$$\int_{B_1} |z|^p |K_{\alpha\beta}(x,z) - K_{\alpha\beta}(y,z)|dz \le C|x-y|^\gamma \qquad \forall x,y \in \mathcal{O},\ \forall \alpha,\beta. \tag{5.14}$$

Condition (5.14) is obviously satisfied if

$$|K_{\alpha\beta}(x,z) - K_{\alpha\beta}(y,z)|dz \le |x-y|^\gamma K(z) \qquad \forall x,y \in \mathcal{O},\ \forall z \in B_1,\ \forall \alpha,\beta. \tag{5.15}$$

Let now $x,y \in \mathcal{O}$. To estimate $\mathrm{d}_{\mathbb{L}_p}(\hat\mu_x^{\alpha\beta},\hat\mu_y^{\alpha\beta})$, we use Proposition 5.11. It follows from (5.11) and (5.14), that

(5.16)
$$\mathrm{d}_{\mathbb{L}_p}(\hat\mu_x^{\alpha\beta},\hat\mu_y^{\alpha\beta}) \le 2^{\frac{p-1}{p}} \left(\int_{B_1} |z|^p \left| K_{\alpha\beta}(x,z) - K_{\alpha\beta}(y,z) \right| dz \right)^{\frac{1}{p}} \le C|x-y|^{\frac{\gamma}{p}}.$$

In particular, (4.1) is satisfied for these measures when $p=1$ and $\gamma=1$.

We can now prove Corollary 4.5.

Proof of Corollary 4.5. We notice that the measures $\mu_x^{\alpha\beta}$ satisfy (5.12), (5.13), and (5.15) with $\gamma = p = 1$ and $K(z) = \Lambda_1|z|^{-d-\sigma}$, and hence they satisfy Assumptions A and E. It is also easy to see that they satisfy Assumption B. Thus the result follows from Theorem 4.4. □

Example 5.13. A well studied subclass of operators which arise in zero-sum two-player stochastic differential games are those of Lévy–Itô form. This corresponds to the situation where the $L^{\alpha\beta}$ appearing in (1.1) have the form

$$L^{\alpha\beta}(u,x) = \int_{U\setminus\{0\}} [u(x+T_x^{\alpha\beta}(z)) - u(x) - Du(x)\cdot T_x^{\alpha\beta}(z)]\, d\mu(z). \tag{5.17}$$

Here U is a separable Hilbert space, and μ is a fixed reference Lévy measure on $U\setminus\{0\}$. The maps $T_x^{\alpha\beta} : U \to \mathbb{R}^d$ are Borel measurable and such that for all $\alpha\in\mathcal{A}, \beta\in\mathcal{B}, x,y\in\mathcal{O}, z\in U\setminus\{0\}$,

$$|T_x^{\alpha\beta}(z) - T_y^{\alpha\beta}(z)| \le C\rho(z)|x-y|, \qquad |T_x^{\alpha\beta}(z)| \le C\rho(z),$$

for some positive Borel function $\rho : U\setminus\{0\} \to \mathbb{R}$ which is bounded on bounded sets, $\inf_{|z|>r}\rho(z) > 0$ for every $r>0$, and

$$\int_{U\setminus\{0\}} \rho(z)^2 d\mu(z) \le C. \tag{5.18}$$

Under these conditions the measures $\mu_x^{\alpha\beta} = (T_x^{\alpha\beta})_\#\mu$ are Lévy measures. The comparison principle for sub-/super- solutions of (1.1) with $L^{\alpha\beta}$ as in (5.17) is known to hold, as discussed in the introduction. Let us revisit it using the transport metric. For every $x_0, y_0 \in \mathcal{O}, \alpha \in \mathcal{A}, \beta \in \mathcal{B}$ we have $\gamma = (T^{\alpha\beta}_{x_0} \times T^{\alpha\beta}_{y_0})_{\#}\mu \in \mathrm{Adm}(\mu^{\alpha\beta}_{x_0}, \mu^{\alpha\beta}_{y_0})$ and, therefore,

$$\begin{aligned} \mathrm{d}_{\mathbb{L}_2}(\mu^{\alpha\beta}_{x_0}, \mu^{\alpha\beta}_{y_0})^2 &\le \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 d\gamma(x,y) \\ &= \int_{U\setminus\{0\}} |T^{\alpha\beta}_{x_0}(z) - T^{\alpha\beta}_{y_0}(z)|^2 \, d\mu(z) \le C|x_0-y_0|^2. \end{aligned}$$

In this case the whole measures μ_{x_0} and μ_{y_0} satisfy Assumption A for $p=2$, and our approach can be applied without the decomposition of the measures $\mu^{\alpha\beta}_x$ into $\hat{\mu}^{\alpha\beta}_x$ and $\check{\mu}^{\alpha\beta}_x$.

If the operators $L^{\alpha\beta}$ in (1.1) have a more common Lévy–Itô form

$$L^{\alpha\beta}(u,x) = \int_{U\setminus\{0\}} [u(x+T^{\alpha\beta}_x(z)) - u(x) - \chi_{B_1(0)} Du(x)\cdot T^{\alpha\beta}_x(z)] \, d\mu(z), \tag{5.19}$$

where instead of (5.18) we now assume only

$$\int_{U\setminus\{0\}} (\rho(z)^2 \chi_{|z|<1} + \chi_{|z|\ge 1}) d\mu(z) \le C,$$

we need to modify this approach. We now do the decomposition

$$\mu^{\alpha\beta}_x = \hat{\mu}^{\alpha\beta}_x + \check{\mu}^{\alpha\beta}_x,$$

where

$$\hat{\mu}^{\alpha\beta}_x = (T^{\alpha\beta}_x)_{\#}\hat{\mu}, \ \hat{\mu} = \mu|_{\{|z|<1\}}, \qquad \check{\mu}^{\alpha\beta}_x = (T^{\alpha\beta}_x)_{\#}\check{\mu}, \ \check{\mu} = \mu|_{\{|z|\ge 1\}},$$

and consider the measures $\hat{\mu}^{\alpha\beta}_x + \check{\mu}^{\alpha\beta}_x$ and $\check{\mu}^{\alpha\beta}_x$ as measures on $\mathbb{R}^d$ by the usual extension. Then the measures $\hat{\mu}^{\alpha\beta}_x \in \mathbb{L}_2(\mathbb{R}^d)$, and they satisfy Assumption A for $p=2$. Unfortunately, the measures $\check{\mu}^{\alpha\beta}_x$ may not satisfy Assumption B now; however, the terms containing them can be handled in a standard way (see, e.g., [16]), and thus our approach can still be implemented (see also the next section and Example 6.2).

6. Variants of the approach

The approach to proving a comparison principle presented so far has been based on the splitting of the measures $\mu^{\alpha\beta}_x$ into $\hat{\mu}^{\alpha\beta}_x$ and $\check{\mu}^{\alpha\beta}_x$, their restrictions to $B_1(0)$ and $B_1^c(0)$, respectively. The reader should think about it as the basic technique. However, in many cases this splitting may not be ideal. When calculating the distance between two measures $\hat{\mu}^{\alpha\beta}_x$ and $\hat{\mu}^{\alpha\beta}_y$, we have only the set $\Gamma = \{0\}$, where we can deposit some excess mass, and moving mass there may be costly. Thus sometimes a much better estimate can be obtained if we allow for a more sophisticated splitting $\mu^{\alpha\beta}_x = \hat{\mu}^{\alpha\beta}_x + \tilde{\mu}^{\alpha\beta}_x + \check{\mu}^{\alpha\beta}_x$, where the measures $\hat{\mu}^{\alpha\beta}_x$ are now supported in some neighborhoods of the origin contained in $B_1(0)$, $\tilde{\mu}^{\alpha\beta}_x$ are bounded measures also supported in $B_1(0)$, and $\check{\mu}^{\alpha\beta}_x$ are as in (2.3). In such a case we may only require that Assumption A (i.e., (4.1)) be satisfied for the new measures $\hat{\mu}^{\alpha\beta}_x$. We will illustrate the advantage of this approach in Example 6.2. Thus the main message is that we should look at the technique of using coupling distance in the proof of comparison principle as flexible, and Assumption A should really be considered to be an assumption about the behavior of Lévy measures for small z, not necessarily for $z \in B_1(0)$.

Suppose then that for every α, β we have a decomposition $\mu_x^{\alpha\beta} = \hat{\mu}_x^{\alpha\beta} + \tilde{\mu}_x^{\alpha\beta} + \check{\mu}_x^{\alpha\beta}$ as described above, and we decompose

$$L^{\alpha\beta}(u,x) = \hat{L}^{\alpha\beta}(u,x) + \tilde{L}^{\alpha\beta}(u,x) + \check{L}^{\alpha\beta}(u,x),$$

where

$$\hat{L}^{\alpha\beta}(u,x) = \int_{\mathbb{R}^d} \delta u(x,z)\, d\hat{\mu}_x^{\alpha\beta}(z), \qquad \tilde{L}^{\alpha\beta}(u,x) = \int_{\mathbb{R}^d} \delta u(x,z)\, d\tilde{\mu}_x^{\alpha\beta}(z),$$
$$\check{L}^{\alpha\beta}(u,x) = \int_{\mathbb{R}^d} [u(x+z) - u(x)]\, d\check{\mu}_x^{\alpha\beta}(z).$$

We can then prove the following variant of Theorem 4.4.

Theorem 6.1. *Let Assumptions* A–E (*with the new measures* $\hat{\mu}_x^{\alpha\beta}$) *hold for* $p \in [1,2]$*, and suppose there are* $L_1, L_2 \geq 0$ *such that*

$$\tilde{\mu}_x^{\alpha\beta}(B_1(0)) \leq L_1 \qquad \forall x \in \mathcal{O},\ \forall \alpha, \beta, \tag{6.1}$$

$$\int_{\mathbb{R}^d} |z|\, d|\tilde{\mu}_x^{\alpha\beta} - \tilde{\mu}_y^{\alpha\beta}|(z) \leq L_2|x-y| \qquad \forall x, y \in \mathcal{O},\ \forall \alpha, \beta. \tag{6.2}$$

Then the comparison principle holds for equation (1.1)*. That is, if* u *and* v *are, respectively, a viscosity subsolution and a viscosity supersolution of* (1.1) *and* $u(x) \leq v(x)$ *for all* $x \notin \mathcal{O}$*, then*

$$u(x) \leq v(x) \qquad \forall x \in \mathcal{O}.$$

Proof. The proof proceeds exactly as the proof of Theorem 4.4 except that now in Step 3 we also need to find an estimate from above for

$$L_{\tilde{\mu}^{\alpha\beta}_{x^*_\delta}}(u^\delta, x_\varepsilon) - L_{\tilde{\mu}^{\alpha\beta}_{y^*_\delta}}(v_\delta, y_\varepsilon),$$

which is independent of α and β, where the operators $L_{\tilde{\mu}^{\alpha\beta}_{x^*_\delta}}(u^\delta, x_\varepsilon)$ and $L_{\tilde{\mu}^{\alpha\beta}_{y^*_\delta}}(v_\delta, y_\varepsilon)$ are defined as in (5.2) for the measures $\tilde{\mu}^{\alpha\beta}_{x^*_\delta}$ and $\tilde{\mu}^{\alpha\beta}_{y^*_\delta}$. We have

$$\begin{aligned} &L_{\tilde{\mu}^{\alpha\beta}_{x^*_\delta}}(u^\delta, x_\varepsilon) - L_{\tilde{\mu}^{\alpha\beta}_{y^*_\delta}}(v_\delta, y_\varepsilon) \\ &\qquad \leq \int_{B_1} \left| u^\delta(x_\varepsilon + z) - u^\delta(x_\varepsilon) - \frac{1}{\varepsilon} D\psi_\kappa(x_\varepsilon - y_\varepsilon)\cdot z \right| d|\tilde{\mu}^{\alpha\beta}_{x^*_\delta} - \tilde{\mu}^{\alpha\beta}_{y^*_\delta}|(z) \\ &\qquad\quad + \int_{B_1} \left(u^\delta(x_\varepsilon + z) - u^\delta(x_\varepsilon) - (v_\delta(y_\varepsilon+z) - v_\delta(y_\varepsilon))\right) d\tilde{\mu}^{\alpha\beta}_{y^*_\delta}(z). \end{aligned}$$

Let ω be the modulus of continuity of u. It is also a modulus of continuity for u^δ. The modulus ω is bounded, and we can assume that it is concave. We notice that the integrand of the second integral above is nonpositive. Therefore, we obtain

$$L_{\tilde{\mu}^{\alpha\beta}_{x^*_\delta}}(u^\delta, x_\varepsilon) - L_{\tilde{\mu}^{\alpha\beta}_{y^*_\delta}}(v_\delta, y_\varepsilon) \leq \int_{B_1} \left(\omega(|z|) + \frac{1}{\varepsilon}|D\psi_\kappa(x_\varepsilon - y_\varepsilon)||z|\right) d|\tilde{\mu}^{\alpha\beta}_{x^*_\delta} - \tilde{\mu}^{\alpha\beta}_{y^*_\delta}|(z).$$

Using (6.1) and (6.2), the concavity of ω, Jensen's inequality, the subadditivity of ω, and $|D\psi_\kappa(x_\varepsilon - y_\varepsilon)| \le p|x_\varepsilon - y_\varepsilon|^{p-1}$, we can now estimate

$$\begin{aligned} L_{\tilde{\mu}^{\alpha\beta}_{x^*_\delta}}(u^\delta, x_\varepsilon) - L_{\tilde{\mu}^{\alpha\beta}_{y^*_\delta}}(v_\delta, y_\varepsilon) &\le C\omega\left(\int_{\mathbb{R}^d} |z|\, d|\tilde{\mu}^{\alpha\beta}_{x^*_\delta} - \tilde{\mu}^{\alpha\beta}_{y^*_\delta}|(z)\right) \\ &\quad + \frac{C}{\varepsilon}|x_\varepsilon - y_\varepsilon|^{p-1}|x^*_\delta - y^*_\delta| \\ &\le C\omega\left(|x^*_\delta - y^*_\delta|\right) + \frac{C}{\varepsilon}|x_\varepsilon - y_\varepsilon|^{p-1}|x^*_\delta - y^*_\delta|. \end{aligned}$$

This allows us to complete the proof by following the rest of the proof of Theorem 4.4. □

The next example illustrates the usefulness of this modified approach and Theorem 6.1.

Example 6.2. Let the measures $\mu^{\alpha\beta}_x$ be such that

$$d\mu^{\alpha\beta}_x(z) = \frac{a_{\alpha\beta}(x)}{|z|^{d+\sigma}}$$

for some $1 < \sigma < 2$. Assume that the functions $a_{\alpha\beta} : \mathcal{O} \to \mathbb{R}$ are nonnegative and such that there exists $L \ge 0$ such that

$$\left|a^{\frac{1}{\sigma}}_{\alpha\beta}(x) - a^{\frac{1}{\sigma}}_{\alpha\beta}(y)\right| \le L|x-y| \qquad \forall x, y \in \mathcal{O},\ \forall \alpha, \beta.$$

Without loss of generality we will also assume that $a_{\alpha\beta} \le 1$. The case $\sigma = 1$ can also be considered similarly, but since calculations are slightly different, it is omitted here, as it is an easy variation. The case $0 < \sigma < 1$ is taken care of by Corollary 4.5.

We decompose the measures $\mu^{\alpha\beta}_x$ in the following way. We set $r^{\alpha\beta}_x := a^{\frac{1}{\sigma}}_{\alpha\beta}(x)$.

$$\hat{\mu}^{\alpha\beta}_x = {\mu^{\alpha\beta}_x}_{|B_{r^{\alpha\beta}_x}}, \quad \tilde{\mu}^{\alpha\beta}_x = {\mu^{\alpha\beta}_x}_{|B_1 \setminus B_{r^{\alpha\beta}_x}}, \quad \check{\mu}^{\alpha\beta}_x = {\mu^{\alpha\beta}_x}_{|B_1^c},$$

and as always all measures are then extended to measures on $\mathbb{R}^d$. We claim that these measures satisfy the assumptions of Theorem 6.1 with $\sigma < p \le 2$.

Assumptions B and E are obvious, so we will focus on Assumption A, (6.1), and (6.2). Regarding Assumption A, we note that, by an elementary calculation, if $a_{\alpha\beta}(x) > 0$, then $\hat{\mu}^{\alpha\beta}_x = (T^{\alpha\beta}_x)_\# \mu$, where

$$T^{\alpha\beta}_x(z) = a^{\frac{1}{\sigma}}_{\alpha\beta}(x) z, \qquad d\mu(z) = \frac{1}{|z|^{d+\sigma}} \chi_{B_1} dz.$$

These types of transformations were used in [2, 10]. If $a_{\alpha\beta}(x) = 0$, then $\hat{\mu}^{\alpha\beta}_x = 0$. Then, if $a_{\alpha\beta}(x) > 0$, $a_{\alpha\beta}(y) > 0$, $\gamma = (T^{\alpha\beta}_x \times T^{\alpha\beta}_y)_\# \mu \in \mathrm{Adm}(\hat{\mu}^{\alpha\beta}_x, \hat{\mu}^{\alpha\beta}_y)$, and

$$\begin{aligned} \mathrm{d}_{\mathbb{L}_p}(\hat{\mu}^{\alpha\beta}_x, \hat{\mu}^{\alpha\beta}_y)^p &\le \int_{\mathbb{R}^d \times \mathbb{R}^d} |z_1 - z_2|^2 d\gamma(z_1, z_2) = \int_{\mathbb{R}^d} |T^{\alpha\beta}_x(z) - T^{\alpha\beta}_y(z)|^p\, d\mu(z) \\ &= \int_{B_1} |a^{\frac{1}{\sigma}}_{\alpha\beta}(x) - a^{\frac{1}{\sigma}}_{\alpha\beta}(y)|^p \frac{|z|^p}{|z|^{d+\sigma}} dz \le C|x-y|^p. \end{aligned}$$

If $a_{\alpha\beta}(x) > 0$ and $\hat\mu_y^{\alpha\beta} = 0$, then also $\gamma = (T_x^{\alpha\beta} \times T_y^{\alpha\beta})_{\#}\mu \in \mathrm{Adm}(\hat\mu_x^{\alpha\beta}, \hat\mu_y^{\alpha\beta})$, where $T_y^{\alpha\beta}(z) = 0$, and

$$\begin{aligned}
\mathrm{d}_{\mathbb{L}_p}(\hat\mu_x^{\alpha\beta}, \hat\mu_y^{\alpha\beta})^p &\le \int_{B_1} |T_x^{\alpha\beta}(z) - 0|^p \, d\mu(z) \\
&= \int_{B_1} |a_{\alpha\beta}^{\frac{1}{\sigma}}(x) - a_{\alpha\beta}^{\frac{1}{\sigma}}(y)|^p \frac{|z|^p}{|z|^{d+\sigma}} dz \le C|x-y|^p.
\end{aligned}$$

Regarding (6.1), we see that

$$\tilde\mu_x^{\alpha\beta}(B_1) = C \int_{r_x^{\alpha\beta}}^1 \frac{a_{\alpha\beta}(x)}{r^{1+\sigma}} dr = \frac{C}{\sigma}(1 - a_{\alpha\beta}(x)).$$

It remains to check condition (6.2). Suppose that $a_{\alpha\beta}(x) < a_{\alpha\beta}(y)$. Then

$$\begin{aligned}
\int_{\mathbb{R}^d} |z| \, d|\tilde\mu_x^{\alpha\beta} - \tilde\mu_y^{\alpha\beta}|(z) = &\int_{r_x^{\alpha\beta} \le |z| < r_y^{\alpha\beta}} |z| \, d\tilde\mu_x^{\alpha\beta}(z) \\
&+ \int_{r_y^{\alpha\beta} \le |z| < 1} \frac{(a_{\alpha\beta}(y) - a_{\alpha\beta}(x))|z|}{|z|^{d+\sigma}} dz.
\end{aligned}$$

We estimate each integral separately:

$$\begin{aligned}
\int_{r_x^{\alpha\beta} \le |z| < r_y^{\alpha\beta}} |z| \, d\tilde\mu_x^{\alpha\beta}(z) &= C a_{\alpha\beta}(x) \frac{1}{r^{\sigma-1}}\Bigg|_{r_x^{\alpha\beta}}^{r_y^{\alpha\beta}} = C a_{\alpha\beta}(x) \frac{a_{\alpha\beta}^{1-\frac{1}{\sigma}}(y) - a_{\alpha\beta}^{1-\frac{1}{\sigma}}(x)}{a_{\alpha\beta}^{1-\frac{1}{\sigma}}(y) a_{\alpha\beta}^{1-\frac{1}{\sigma}}(x)} \\
&\le C a_{\alpha\beta}^{\frac{2}{\sigma}-1}(x) \left(a_{\alpha\beta}^{1-\frac{1}{\sigma}}(y) - a_{\alpha\beta}^{1-\frac{1}{\sigma}}(x) \right).
\end{aligned}$$

By the mean value theorem

$$a_{\alpha\beta}^{1-\frac{1}{\sigma}}(y) - a_{\alpha\beta}^{1-\frac{1}{\sigma}}(x) = c^{\sigma-2} \left(a_{\alpha\beta}^{\frac{1}{\sigma}}(y) - a_{\alpha\beta}^{\frac{1}{\sigma}}(x) \right) \le a_{\alpha\beta}^{1-\frac{2}{\sigma}}(x) \left(a_{\alpha\beta}^{\frac{1}{\sigma}}(y) - a_{\alpha\beta}^{\frac{1}{\sigma}}(x) \right),$$

where c is some number such that $a_{\alpha\beta}^{\frac{1}{\sigma}}(x) < c < a_{\alpha\beta}^{\frac{1}{\sigma}}(y)$. Thus we obtain

$$\int_{r_x^{\alpha\beta} \le |z| < r_y^{\alpha\beta}} |z| \, d\tilde\mu_x^{\alpha\beta}(z) \le C \left(a_{\alpha\beta}^{\frac{1}{\sigma}}(y) - a_{\alpha\beta}^{\frac{1}{\sigma}}(x) \right) \le C_1 |x-y|.$$

For the second integral, by an elementary calculation we obtain

$$\int_{r_y^{\alpha\beta} \le |z| < 1} \frac{(a_{\alpha\beta}(y) - a_{\alpha\beta}(x))|z|}{|z|^{d+\sigma}} dz = C(a_{\alpha\beta}(y) - a_{\alpha\beta}(x)) \left(\frac{1}{a_{\alpha\beta}^{1-\frac{1}{\sigma}}(y)} - 1 \right).$$

Now, again by the mean value theorem,

$$a_{\alpha\beta}(y) - a_{\alpha\beta}(x) = c^{\sigma-1} \left(a_{\alpha\beta}^{\frac{1}{\sigma}}(y) - a_{\alpha\beta}^{\frac{1}{\sigma}}(x) \right) \le a_{\alpha\beta}^{1-\frac{1}{\sigma}}(y) \left(a_{\alpha\beta}^{\frac{1}{\sigma}}(y) - a_{\alpha\beta}^{\frac{1}{\sigma}}(x) \right),$$

where c is some number such that $a_{\alpha\beta}^{\frac{1}{\sigma}}(x) < c < a_{\alpha\beta}^{\frac{1}{\sigma}}(y)$. Therefore, it follows that

$$\begin{aligned}
\int_{r_y^{\alpha\beta} \le |z| < 1} \frac{(a_{\alpha\beta}(y) - a_{\alpha\beta}(x))|z|}{|z|^{d+\sigma}} dz &\le C \left(a_{\alpha\beta}^{\frac{1}{\sigma}}(y) - a_{\alpha\beta}^{\frac{1}{\sigma}}(x) \right) (1 - a_{\alpha\beta}^{1-\frac{1}{\sigma}}(y)) \\
&\le C_2 |x-y|.
\end{aligned}$$

This completes the proof of (6.2).

We remark that if we apply the estimate of Example 5.12 to the kernels

$$K_{\alpha\beta}(x,z) = \frac{a_{\alpha\beta}(x)}{|z|^{d+\sigma}}$$

and use only the information that the functions $a_{\alpha\beta}$ are Lipschitz continuous, we obtain

$$\mathrm{d}_{\mathbb{L}_p}(\hat{\mu}_x^{\alpha\beta}, \hat{\mu}_y^{\alpha\beta}) \leq C|x-y|^{\frac{1}{p}}.$$

Thus Example 6.2 shows that the general estimate of Example 5.12 coming from Proposition 5.11 is not optimal for $1 < p \leq 2$.

The following two examples concern Awatif's comparison results in [19]. Example 6.3 in particular shows Theorem 4.4 implies the main comparison result in [19], in the case where the measures $\mu^{\alpha\beta}$ are all supported in a ball. We note that [19, Theorem III.1] dealt with the case of $\mathcal{O} = \mathbb{R}^d$ and the nonlocal operators there were slightly different, so Theorem 4.4 cannot be applied directly to the case considered in [19]; however, our approach covers the essential difficulties of the proof of the general result of [19]. Example 6.4 is related to an alternative assumption discussed later in the paper [19, Section III.1, p. 1065].

Example 6.3. In this example we explain how the assumption of [19, equation (1.3)] (reproduced below in (6.3)) implies Assumption A with $p = 1$: assume that there is a constant $C > 0$ such that, for any $\phi \in C^{0,1}(B_1)$ with $\phi(0) = 0$ and Lipschitz constant 1, we have the inequality

$$\int_{B_1} \phi(z)\, d\mu_x(z) - \int_{B_1} \phi(z)\, d\mu_y(z) \leq C|x-y| \qquad \forall x, y \in \mathcal{O}, \tag{6.3}$$

where we assume that the measures $\mu_x \in \mathbb{L}_1(B_1)$. We will show that then the measures $\hat{\mu}_x = \mu_x$ satisfy Assumption A with $p = 1$, that is, $\mathrm{d}_{\mathbb{L}_1}(\hat{\mu}_x, \hat{\mu}_y) \leq C|x-y|$ for all x, y. To this end, let ϕ be any function as above. The Lipschitz condition means that for every z_1, z_2 we have $\phi(z_1) - \phi(z_2) \leq |z_1 - z_2|$, and in particular the pair $(\phi, -\phi)$ belongs to the set Adm^1 defined in Lemma 3.6. Then Lemma 3.6 says that

$$\mathrm{d}_{\mathbb{L}_1}(\hat{\mu}_x, \hat{\mu}_y) \leq \int_{B_1} \phi(z)\, d\mu_x(z) + \int_{B_1} (-\phi(z))\, d\mu_y(z),$$

and thus assumption (6.3) implies that $\mathrm{d}_{\mathbb{L}_1}(\hat{\mu}_x, \hat{\mu}_y) \leq C|x-y|$.

Example 6.4. Assume that the measures $\mu_x \in \mathbb{L}_1(B_1)$ and that there is a constant $C > 0$ such that, for all $\tau > 0$ and $x, y \in \mathcal{O}$,

$$\sup_{h \in C(B_1 \setminus \{0\}), \|h\|_\infty \leq 1} \left\{ \int_{\tau \leq |z| < 1} h(z)|z| d\mu_x(z) - \int_{\tau \leq |z| < 1} h(z)|z|\, d\mu_y(z) \right\} \leq C|x-y|. \tag{6.4}$$

Then the measures $\hat{\mu}_x = \mu_x$ satisfy Assumption A with $p = 1$, that is,

$$\mathrm{d}_{\mathbb{L}_1}(\hat{\mu}_x, \hat{\mu}_y) \leq C|x-y|.$$

To show it, we start arguing as in Example 6.3. We take any function ϕ with Lipschitz constant 1 and such that $\phi(0) = 0$. The Lipschitz condition means that for every z_1, z_2 we have $\phi(z_1) - \phi(z_2) \leq |z_1 - z_2|$, and in particular $(\phi, -\phi) \in \mathrm{Adm}^1$. Then Lemma 3.6 guarantees that

$$\mathrm{d}_{\mathbb{L}_1}(\hat{\mu}_x, \hat{\mu}_y) \leq \int_{B_1} \phi(z)\, d\mu_x(z) - \int_{B_1} \phi(z)\, \mu_y(z).$$

Since ϕ has Lipschitz constant 1 and $\phi(0) = 0$, it follows that $|\phi(z)| \leq |z|$. In particular, if $h(z) = |z|^{-1}|\phi(z)|$, then h is continuous in $B_1 \setminus \{0\}$ and $|h(z)| \leq 1$ for all z, so $h(z)$ is an admissible function for the supremum in (6.4). We conclude that for every $\tau > 0$

$$\begin{aligned}&\int_{\tau \leq |z| < 1} \phi(z)\, d\mu_x(z) - \int_{\tau \leq |z| < 1} \phi(z)\, \mu_y(z) \\ &\quad = \int_{\tau \leq |z| < 1} h(z)|z|\, d\mu_x(z) - \int_{\tau \leq |z| < 1} h(z)|z|\, \mu_y(z) \leq C|x - y|.\end{aligned}$$

Letting $\tau \to 0$, the integral on the left converges to $\int_{B_1(0)} \phi(z)\, d\mu_x(z) - \int_{B_1(0)} \phi(z)$ $\mu_y(z)$, so

$$\mathrm{d}_{\mathbb{L}_1}(\hat{\mu}_x, \hat{\mu}_y) \leq C|x - y|.$$

7. Comparison principles under additional assumptions

As in the previous section, throughout this section we consider a fixed bounded domain $\mathcal{O} \subset \mathbb{R}^d$. In this section we prove a few comparison results for more regular viscosity sub-/super- solutions. In return, we are allowed to replace Assumption A with a weaker assumption.

Assumption A1. Let $p \in [1, 2]$. There exist $C > 0$ and $s \in (0, 1)$ such that

$$\mathrm{d}_{\mathbb{L}_p}(\hat{\mu}_x^{\alpha\beta}, \hat{\mu}_y^{\alpha\beta}) \leq C|x - y|^s \qquad \forall x, y \in \mathcal{O},\ \forall \alpha, \beta.$$

Remark 7.1. Consider a Lévy measure $\mu \in \mathbb{L}_p(B_1)$. For $r \in (0, 1)$ we define

$$\mu_r(\cdot) := \mu(\cdot \cap B_r^c).$$

Then we have the estimate

$$\mathrm{d}_{\mathbb{L}_p}(\mu, \mu_r)^p \leq \int_{B_r} |x|^p\, d\mu(x).$$

To see why this is so, simply note that among the admissible plans we have the one that sends all of the mass of μ in $B_r \setminus \{0\}$ to 0 and leaves the rest of the mass fixed in place. To be more precise, define

$$T(x) = \begin{cases} 0 & \text{for } x \in B_r \setminus \{0\}, \\ x & \text{otherwise.} \end{cases}$$

Then $\gamma = (T \times \mathrm{Id})_{\#}\mu \in \mathrm{Adm}(\mu_r, \mu)$ and

$$\mathrm{d}_{\mathbb{L}_p}(\mu, \mu_r)^p \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p d\gamma(x, y) = \int_{\mathbb{R}^d} |T(x) - x|^2\, d\mu(x) = \int_{B_r} |x|^p\, d\mu(x).$$

Remark 7.2. Estimating the distance between μ and μ_r is of interest to us since it can be used to bound the difference between the operators

$$L_{\hat{\mu}}(u, x) := \int_{B_1} \delta u(x, z)\, d\hat{\mu}(z)$$

and

$$L_{\hat{\mu}_r}(u,x) := \int_{B_1 \cap B_r^c} \delta u(x,z)\, d\hat{\mu}_r(z).$$

The second operator can be classically evaluated for any continuous function, while the first one in general cannot. Being able to estimate the difference between them will be an important step in the proof of Theorem 7.4, removing the need for the use of the sup-/inf- convolutions (as mentioned in Remark 5.1).

Theorem 7.3. *Let Assumptions* A1 *and* A–E *be satisfied. Let u be a viscosity subsolution, and let v be a viscosity supersolution of* (1.1)*, and let $u(x) \le v(x)$ for all $x \notin \mathcal{O}$. If either u or v is in $C^{0,r}(\mathcal{O})$ for some $r \in (0,1)$ and we have $1 - \frac{r}{p} < s$ (where s is from Assumption* A1*), then*

$$u(x) \le v(x) \qquad \forall x \in \mathcal{O}.$$

Proof. The proof follows along the lines of the proof of Theorem 4.4. The only difference is that when either u or v is $C^{0,r}$, instead of (5.3) we now have (5.4), i.e.,

$$\limsup_{\kappa \to 0} \frac{|x_\varepsilon - y_\varepsilon|^{p-r}}{\varepsilon} \le C,$$

and in this case, following the original proof, we obtain

$$\lambda \ell \le 2\delta + \frac{C}{\varepsilon}|x_\varepsilon - y_\varepsilon|^{sp} + C\theta\big(|x_\varepsilon - y_\varepsilon|\big) + \rho_\varepsilon(\delta),$$

which produces a contradiction after taking $\lim_{\varepsilon\to 0} \lim_{\delta \to 0} \limsup_{\kappa \to 0}$ if $sp > p - r$, which is the case precisely when $1 - \frac{r}{p} < s$. □

The previous theorem does not cover the limiting situation where $s = 1 - \frac{r}{p}$; however, with extra work one can show that if u or v is of class C^1, then we can choose $s = 1 - \frac{1}{p}$, and we still have a comparison. The proof is different from that of Theorem 4.4 since we do not use the sup-/inf- convolutions.

Theorem 7.4. *Let Assumptions* B–F *hold, and let Assumption* A1 *hold with $p > 1$ and $s = 1 - \frac{1}{p}$. Suppose that u and v are, respectively, a viscosity subsolution and a viscosity supersolution of* (1.1) *and $u(x) \le v(x)$ for all $x \notin \mathcal{O}$. If either u or v is in $C^1(\mathcal{O})$, then*

$$u(x) \le v(x) \qquad \forall x \in \mathcal{O}.$$

Remark 7.5. If we allowed C^1 functions to be test functions in the case $p = 1$, then Theorem 7.4 would trivially hold for $p = 1$ without the need for Assumptions A and B and the continuity of the coefficients, since then either u or v would be a classical sub-/supersolution of (1.1) and could thus be used as a test function.

Proof. Without loss of generality let us say that $u \in C^1(\mathcal{O})$. As before, we argue by contradiction, in which case there is some $\ell > 0$ such that

$$\sup_{x \in \mathbb{R}^d} \{u(x) - v(x)\} = \ell.$$

Step 1 (Doubling of variables and perturbation). Let $K \subset \mathcal{O}$ be a compact neighborhood of the set of maximum points of $u - v$ in $\mathcal{O}$. There exists a sequence of

$C^3(\mathbb{R}^d) \cap \mathrm{BUC}(\mathbb{R}^d)$ functions $\{\phi_n\}_n$, each of which has second and third derivatives bounded in $\mathbb{R}^d$, and such that $|u - \phi_n| \to 0$ uniformly in $\mathbb{R}^d$ and

$$\lim_{n\to\infty} \sup_K |Du - D\phi_n| = 0.$$

Now let $(x_{n,\varepsilon}, y_{n,\varepsilon}) \in \mathbb{R}^d \times \mathbb{R}^d$ be a global maximum point of $w^{n,\varepsilon}$ over $\mathbb{R}^d \times \mathbb{R}^d$, where

$$w^{n,\varepsilon}(x,y) := \big(u(x) - \phi_n(x)\big) - \big(v(y) - \phi_n(y)\big) - \frac{\psi_\kappa(x-y)}{\varepsilon}.$$

As in Lemma 5.10, one can show that for any n

$$\limsup_{\varepsilon\to 0} \limsup_{\kappa\to 0} \frac{|x_{n,\varepsilon} - y_{n,\varepsilon}|^p}{\varepsilon} = 0,$$

$$\limsup_{\varepsilon\to 0} \limsup_{\kappa\to 0} \big(u(x_{n,\varepsilon}) - v(y_{n,\varepsilon})\big) = \ell.$$

Observe that u is touched from above at $x_{n,\varepsilon}$ by

$$\bar{\phi}(x) := u(x_{n,\varepsilon}) - \phi_n(x_{n,\varepsilon}) + \phi_n(x) + \frac{1}{\varepsilon}\left(\psi_\kappa(x - y_{n,\varepsilon}) - \psi_\kappa(x_{n,\varepsilon} - y_{n,\varepsilon})\right),$$

while v is touched from below at $y_{n,\varepsilon}$ by

$$\underline{\phi}(y) := v(y_{n,\varepsilon}) - \phi_n(y_{n,\varepsilon}) + \phi_n(y) - \frac{1}{\varepsilon}\left(\psi_\kappa(x_{n,\varepsilon} - y) - \psi_\kappa(x_{n,\varepsilon} - y_{n,\varepsilon})\right).$$

Since u is C^1, this means first that

$$Du(x_{n,\varepsilon}) - D\phi_n(x_{n,\varepsilon}) = p(\kappa + |x_{n,\varepsilon} - y_{n,\varepsilon}|^2)^{\frac{p}{2}-1} \frac{x_{n,\varepsilon} - y_{n,\varepsilon}}{\varepsilon}.$$

There is some small $c > 0$ such that $B_c(x_{n,\varepsilon}) \cup B_c(y_{n,\varepsilon}) \subset K$ if ε and κ are sufficiently small. Therefore,

$$\limsup_{\varepsilon\to 0} \limsup_{\kappa\to 0} \frac{|x_{n,\varepsilon} - y_{n,\varepsilon}|^{p-1}}{\varepsilon} = o_{\frac{1}{n}}(1). \tag{7.1}$$

On the other hand, since u is a viscosity subsolution and v a viscosity supersolution, for any $0 < r < 1$ we have

$$I(u_r, x_{n,\varepsilon}) \le 0$$

and

$$I(v_r, y_{n,\varepsilon}) \ge 0,$$

where (recall Definition 4.1)

$$u_r(x) := \begin{cases} \bar{\phi}(x) & \text{in } B_r(x_{n,\varepsilon}), \\ u(x) & \text{in } B_r^c(x_{n,\varepsilon}) \end{cases}$$

and

$$v_r(x) := \begin{cases} \underline{\phi}(x) & \text{in } B_r(y_{n,\varepsilon}), \\ v(x) & \text{in } B_r^c(y_{n,\varepsilon}). \end{cases}$$

Step 2 (Equation structure, main term). Using the fact that $I(\cdot,x)$ has the inf-sup representation in (1.1), it follows that

$$
\begin{aligned}
I(v_r,y_{n,\varepsilon})-I(u_r,x_{n,\varepsilon}) &\le \sup_{\alpha,\beta}\Big\{\hat L^{\alpha\beta}(u_r,x_{n,\varepsilon})-\hat L^{\alpha\beta}(v_r,y_{n,\varepsilon})\Big\}\\
&\quad+\sup_{\alpha,\beta}\Big\{\check L^{\alpha\beta}(u_r,x_{n,\varepsilon})-\check L^{\alpha\beta}(v_r,y_{n,\varepsilon})\Big\}\\
&\quad+\sup_{\alpha,\beta}\Big\{c_{\alpha\beta}(y_{n,\varepsilon})v(y_{n,\varepsilon})-c_{\alpha\beta}(x_{n,\varepsilon})u(x_{n,\varepsilon})\Big\}\\
&\quad+\sup_{\alpha,\beta}\Big\{f_{\alpha\beta}(y_{n,\varepsilon})-f_{\alpha\beta}(x_{n,\varepsilon})\Big\}.
\end{aligned}
\tag{7.2}
$$

Let us bound each of the terms on the right-hand side of (7.2). As before, the most delicate term is the first one. Fixing α and β, we note that

$$
\begin{aligned}
\hat L^{\alpha\beta}(u_r,x_{n,\varepsilon}) &= \int_{B_r}\delta u_r(x_{n,\varepsilon},x)\,d\hat\mu^{\alpha\beta}_{x_{n,\varepsilon}}(x)+\int_{B_r^c}\delta u(x_{n,\varepsilon},x)\,d\hat\mu^{\alpha\beta}_{x_{n,\varepsilon}}(x),\\
\hat L^{\alpha\beta}(v_r,y_{n,\varepsilon}) &= \int_{B_r}\delta v_r(y_{n,\varepsilon},y)\,d\hat\mu^{\alpha\beta}_{y_{n,\varepsilon}}(y)+\int_{B_r^c}\delta v(y_{n,\varepsilon},y)\,d\hat\mu^{\alpha\beta}_{y_{n,\varepsilon}}(y).
\end{aligned}
$$

Let us choose $\gamma_r\in\mathrm{Adm}(\hat\mu^{\alpha\beta}_{x_{n,\varepsilon},r},\hat\mu^{\alpha\beta}_{y_{n,\varepsilon},r})$ (using the notation introduced in Remark 7.1), which minimizes the p-cost. Denote

$$
A_r:=(\{0\}\times(B_1\setminus B_r))\cup((B_1\setminus B_r)\times\{0\})\cup((B_1\setminus B_r)\times(B_1\setminus B_r)).
$$

Since $\delta u(x_{n,\varepsilon},0)=\delta v(y_{n,\varepsilon},0)=0$ and

$$
\gamma_r\left((((B_1\setminus B_r)\cup\{0\})^c\times\mathbb{R}^d)\cup(\mathbb{R}^d\times((B_1\setminus B_r)\cup\{0\})^c)\right)=0,\tag{7.3}
$$

we have

$$
\begin{aligned}
\int_{B_r^c}\delta u(x_{n,\varepsilon},x)\,d\hat\mu^{\alpha\beta}_{x_{n,\varepsilon}}(x)&=\int_{B_r^c}\delta u(x_{n,\varepsilon},x)\,d\hat\mu^{\alpha\beta}_{x_{n,\varepsilon},r}(x)\\
&=\int_{\mathbb{R}^d\times\mathbb{R}^d}\delta u(x_{n,\varepsilon},x)\,d\gamma_r(x,y)=\int_{A_r}\delta u(x_{n,\varepsilon},x)\,d\gamma_r(x,y).
\end{aligned}
$$

Similarly,

$$
\int_{B_r^c}\delta v(y_{n,\varepsilon},y)\,d\hat\mu^{\alpha\beta}_{y_{n,\varepsilon}}(y)=\int_{A_r}\delta v(y_{n,\varepsilon},y)\,d\gamma_r(x,y).
$$

Therefore, we obtain

$$
\begin{aligned}
&\hat L^{\alpha\beta}(u_r,x_{n,\varepsilon})-\hat L^{\alpha\beta}(v_r,y_{n,\varepsilon})\\
&\quad=\int_{B_r}\delta u_r(x_{n,\varepsilon},x)\,d\hat\mu^{\alpha\beta}_{x_{n,\varepsilon}}(x)-\int_{B_r}\delta v_r(y_{n,\varepsilon},y)\,d\hat\mu^{\alpha\beta}_{y_{n,\varepsilon}}(y)\\
&\qquad+\int_{A_r}[\delta u(x_{n,\varepsilon},x)-\delta v(y_{n,\varepsilon},y)]\,d\gamma_r(x,y).
\end{aligned}
$$

Using the fact that $(x_{n,\varepsilon},y_{n,\varepsilon})$ is a maximum point of $w^{n,\varepsilon}$, we have the following pointwise bound for pairs $(x,y)\in A_r$:

$$
\delta u(x_{n,\varepsilon},x)-\delta v(y_{n,\varepsilon},y)\le\frac{C}{\varepsilon}|x-y|^p+\delta\phi(x_{n,\varepsilon},x)-\delta\phi(y_{n,\varepsilon},y).
$$

It thus follows (again using $\delta\phi(x_{n,\varepsilon},0)=\delta\phi(y_{n,\varepsilon},0)=0$ and (7.3)) that

$$\begin{aligned}\int_{A_r}&[\delta u(x_{n,\varepsilon},x)-\delta v(y_{n,\varepsilon},y)]\,d\gamma_r(x,y)\\&\le \frac{C}{\varepsilon}\int_{A_r}|x-y|^p\,d\gamma_r(x,y)\\&\quad+\int_{B_r^c}\delta\phi(x_{n,\varepsilon},x)\,d\hat\mu^{\alpha\beta}_{x_{n,\varepsilon}}(x)-\int_{B_r^c}\delta\phi(y_{n,\varepsilon},y)\,d\hat\mu^{\alpha\beta}_{y_{n,\varepsilon}}(y),\end{aligned}$$

and since γ_r is the optimizer in $\mathrm{Adm}(\hat\mu_{x_{n,\varepsilon}},\hat\mu_{y_{n,\varepsilon}})$,

$$\begin{aligned}\int_{A_r}&[\delta u(x_{n,\varepsilon},x)-\delta v(y_{n,\varepsilon},y)]\,d\gamma_r(x,y)\\&\le \frac{C}{\varepsilon}\mathrm{d}_{\mathbb{L}_p}(\hat\mu^{\alpha\beta}_{x_{n,\varepsilon},r},\hat\mu^{\alpha\beta}_{y_{n,\varepsilon},r})^p+\int_{B_r^c}\delta\phi(x_{n,\varepsilon},x)\,d\hat\mu^{\alpha\beta}_{x_{n,\varepsilon}}(x)\\&\quad-\int_{B_r^c}\delta\phi(y_{n,\varepsilon},y)\,d\hat\mu^{\alpha\beta}_{y_{n,\varepsilon}}(y).\end{aligned}$$

As for the integrals over $B_r(0)$, note that

$$\begin{aligned}\int_{B_r}&\delta u_r(x_{n,\varepsilon},x)\,d\hat\mu^{\alpha\beta}_{x_{n,\varepsilon}}(x)-\int_{B_r}\delta v_r(y_{n,\varepsilon},y)\,d\hat\mu^{\alpha\beta}_{y_{n,\varepsilon}}(y)\\&=\int_{B_r}[\delta\phi_n(x_{n,\varepsilon},x)+\frac{1}{\varepsilon}\delta\psi_\kappa(\cdot-y_{n,\varepsilon})(x_{n,\varepsilon},x)]\,d\hat\mu^{\alpha\beta}_{x_{n,\varepsilon}}(x)\\&\quad-\int_{B_r}[\delta\phi_n(y_{n,\varepsilon},y)-\frac{1}{\varepsilon}\delta\psi_\kappa(x_{n,\varepsilon}-\cdot)(y_{n,\varepsilon},y)]\,d\hat\mu^{\alpha\beta}_{y_{n,\varepsilon}}(x).\end{aligned}$$

Putting the last inequality and last equality together, we have

$$\begin{aligned}\hat L^{\alpha\beta}&(u_r,x_{n,\varepsilon})-\hat L^{\alpha\beta}(v_r,y_{n,\varepsilon})\\&\le\frac{C}{\varepsilon}\mathrm{d}_{\mathbb{L}_p}(\hat\mu^{\alpha\beta}_{x_{n,\varepsilon},r},\hat\mu^{\alpha\beta}_{y_{n,\varepsilon},r})^p+\hat L^{\alpha\beta}(\phi_n,x_{n,\varepsilon})-\hat L^{\alpha\beta}(\phi_n,y_{n,\varepsilon})+\frac{C}{\varepsilon}\theta(r),\end{aligned}$$

where we used Assumption E and Lemma 5.7.

Next, we use Remark 7.1 to get $\mathrm{d}_{\mathbb{L}_p}(\hat\mu^{\alpha\beta}_{x_{n,\varepsilon},r},\hat\mu^{\alpha\beta}_{y_{n,\varepsilon},r})^p\le\mathrm{d}_{\mathbb{L}_p}(\hat\mu^{\alpha\beta}_{x_{n,\varepsilon}},\hat\mu^{\alpha\beta}_{y_{n,\varepsilon}})^p+\rho(r)$, where $\rho(r)\to0$ as $r\to0$. Then, using Assumption A1 (recall that $s=1-1/p$), it follows that

$$\mathrm{d}_{\mathbb{L}_p}(\hat\mu^{\alpha\beta}_{x_{n,\varepsilon},r},\hat\mu^{\alpha\beta}_{y_{n,\varepsilon},r})^p\le C|x_{n,\varepsilon}-y_{n,\varepsilon}|^{p-1}+\rho(r).$$

Thus

$$\begin{aligned}\hat L^{\alpha\beta}(u_r,x_{n,\varepsilon})-\hat L^{\alpha\beta}(v_r,y_{n,\varepsilon})&\le\frac{C}{\varepsilon}|x_{n,\varepsilon}-y_{n,\varepsilon}|^{p-1}\\&\quad+\hat L^{\alpha\beta}(\phi_n,x_{n,\varepsilon})-\hat L^{\alpha\beta}(\phi_n,y_{n,\varepsilon})+\frac{C}{\varepsilon}(\theta(r)+\rho(r)).\end{aligned}$$

Letting $r\to0$, it follows that for every κ, n, and ε

$$\begin{aligned}&\limsup_{r\to0}\sup_{\alpha,\beta}\Big\{\hat L^{\alpha\beta}(u_r,x_{n,\varepsilon})-\hat L^{\alpha\beta}(v_r,y_{n,\varepsilon})\Big\}\\&\qquad\le\frac{C}{\varepsilon}|x_{n,\varepsilon}-y_{n,\varepsilon}|^{p-1}+\sup_{\alpha,\beta}\big\{\hat L^{\alpha\beta}(\phi_n,x_{n,\varepsilon})-\hat L^{\alpha\beta}(\phi_n,y_{n,\varepsilon})\big\}.\end{aligned}\tag{7.4}$$

Step 3 (Equation structure, remaining terms). For any $r \in (0,1)$ and any α, β we have

$$\check{L}^{\alpha\beta}(u_r, x_{n,\varepsilon}) - \check{L}^{\alpha\beta}(v_r, y_{n,\varepsilon}) = \check{L}^{\alpha\beta}(u, x_{n,\varepsilon}) - \check{L}^{\alpha\beta}(v, y_{n,\varepsilon}).$$

Furthermore, arguing as in Step 3 of the proof of Theorem 4.4,

$$c_{\alpha\beta}(x_{n,\varepsilon})u(x_{n,\varepsilon}) - c_{\alpha\beta}(y_{n,\varepsilon})v(y_{n,\varepsilon}) \geq \lambda(u(x_{n,\varepsilon}) - v(y_{n,\varepsilon})) - \|v\|_{L^\infty}\theta(|x_{n,\varepsilon} - y_{n,\varepsilon}|), \tag{7.5}$$

$$|f_{\alpha\beta}(x_{n,\varepsilon}) - f_{\alpha\beta}(y_{n,\varepsilon})| \leq \theta(|x_{n,\varepsilon} - y_{n,\varepsilon}|). \tag{7.6}$$

Going back to (7.2) and combining it with (7.4)–(7.6), it follows that for any κ, n, and ε

$$\begin{aligned}\limsup_{r\to 0}\big\{I(v_r, y_{n,\varepsilon}) - I(u_r, x_{n,\varepsilon})\big\} \leq{}& \frac{C}{\varepsilon}|x_{n,\varepsilon} - y_{n,\varepsilon}|^{p-1} + (1 + \|v\|_{L^\infty})\theta(|x_{n,\varepsilon} - y_{n,\varepsilon}) \\ &+ \lambda(v(y_{n,\varepsilon}) - u(x_{n,\varepsilon})) \\ &+ \sup_{\alpha,\beta}\big\{\check{L}^{\alpha\beta}(u, x_{n,\varepsilon}) - \check{L}^{\alpha\beta}(v, y_{n,\varepsilon})\big\} \\ &+ \sup_{\alpha,\beta}\big\{\hat{L}^{\alpha\beta}(\phi_n, x_{n,\varepsilon}) - \hat{L}^{\alpha\beta}(\phi_n, y_{n,\varepsilon})\big\}.\end{aligned} \tag{7.7}$$

Let us handle the last two terms on the right. Using Assumption B and the fact that $\phi_n \in C^{0,1}(\mathbb{R}^d)$, we have for any α and β

$$\begin{aligned}&\check{L}^{\alpha\beta}(u, x_{n,\varepsilon}) - \check{L}^{\alpha\beta}(v, y_{n,\varepsilon}) \\ &\quad= \int_{B_1^c} [\delta u(x_{n,\varepsilon}, x) - \delta v(y_{n,\varepsilon}, x)]\, d\check{\mu}^{\alpha\beta}_{x_{n,\varepsilon}}(x) \\ &\qquad+ \int_{B_1^c} \delta v(y_{n,\varepsilon}, y) d\left[\check{\mu}^{\alpha\beta}_{x_{n,\varepsilon}}(y) - \check{\mu}^{\alpha\beta}_{y_{n,\varepsilon}}(y)\right] \\ &\quad\leq \int_{B_1^c} (\delta\phi_n(x_{n,\varepsilon}, x) - \delta\phi_n(y_{n,\varepsilon}, x))\, d\check{\mu}^{\alpha\beta}_{x_{n,\varepsilon}}(x) + 2\|v\|_{L^\infty} \mathrm{d}_{\mathrm{TV}}(\check{\mu}^{\alpha\beta}_{x_{n,\varepsilon}}, \check{\mu}^{\alpha\beta}_{y_{n,\varepsilon}}) \\ &\quad\leq C(n)|x_{n,\varepsilon} - y_{n,\varepsilon}| + 2\|v\|_{L^\infty}\theta(|x_{n,\varepsilon} - y_{n,\varepsilon}|).\end{aligned}$$

Then we have

$$\sup_{\alpha,\beta}\big\{\check{L}^{\alpha\beta}(u, x_{n,\varepsilon}) - \check{L}^{\alpha\beta}(v, y_{n,\varepsilon})\big\} \leq C(n)|x_{n,\varepsilon} - y_{n,\varepsilon}| + 2\|v\|_{L^\infty}\theta(|x_{n,\varepsilon} - y_{n,\varepsilon}|). \tag{7.8}$$

For the other remaining term we note that

$$\begin{aligned}\hat{L}^{\alpha\beta}(\phi_n, x_{n,\varepsilon}) - \hat{L}^{\alpha\beta}(\phi_n, y_{n,\varepsilon}) ={}& \int_{B_1} [\delta\phi_n(x_{n,\varepsilon}, x) - \delta\phi_n(y_{n,\varepsilon}, x)]\, d\hat{\mu}^{\alpha\beta}_{x_{n,\varepsilon}}(x) \\ &+ \int_{B_1} \delta\phi_n(y_{n,\varepsilon}, x) d\hat{\mu}^{\alpha\beta}_{x_{n,\varepsilon}}(x) \\ &- \int_{B_1} \delta\phi_n(y_{n,\varepsilon}, y) d\hat{\mu}^{\alpha\beta}_{y_{n,\varepsilon}}(y).\end{aligned}$$

Since the third derivatives of ϕ_n are bounded, we have

$$\begin{aligned}&\left|\int_{B_1}[\delta\phi_n(x_{n,\varepsilon},x)-\delta\phi_n(y_{n,\varepsilon},x)]\,d\hat{\mu}^{\alpha\beta}_{x_{n,\varepsilon}}(x)\right|\\&\qquad\le\int_{B_1}\int_0^1|\langle(D^2\phi_n(x_{n,\varepsilon}+sx)-D^2\phi_n(y_{n,\varepsilon}+sx))x,x\rangle(1-s)|ds\,d\hat{\mu}^{\alpha\beta}_{x_{n,\varepsilon}}(x)\\&\qquad\le C(n)|x_{n,\varepsilon}-y_{n,\varepsilon}|.\end{aligned}$$

The remaining integrals are estimated as follows. For $\tau\in(0,1)$ let η_τ be a smooth function such that $0\le\eta_\tau\le 1$, $\eta_\tau\equiv 1$ in $B_\tau(0)$, and $\eta_\tau\equiv 0$ outside of $B_{2\tau}$. Then we may write

$$\begin{aligned}\int_{B_1}\delta\phi_n(y_{n,\varepsilon},x)d\hat{\mu}^{\alpha\beta}_{x_{n,\varepsilon}}(x)&=\int_{B_1}(1-\eta_\tau(x))\delta\phi_n(y_{n,\varepsilon},x)d\hat{\mu}^{\alpha\beta}_{x_{n,\varepsilon}}(x)\\&\quad+\int_{B_1}\eta_\tau(x)\delta\phi_n(y_{n,\varepsilon},x)d\hat{\mu}^{\alpha\beta}_{x_{n,\varepsilon}}(x),\\\int_{B_1}\delta\phi_n(y_{n,\varepsilon},y)d\hat{\mu}^{\alpha\beta}_{y_{n,\varepsilon}}(y)&=\int_{B_1}(1-\eta_\tau(y))\delta\phi_n(y_{n,\varepsilon},y)d\hat{\mu}^{\alpha\beta}_{y_{n,\varepsilon}}(y)\\&\quad+\int_{B_1}\eta_\tau(y)\delta\phi_n(y_{n,\varepsilon},y)d\hat{\mu}^{\alpha\beta}_{y_{n,\varepsilon}}(y).\end{aligned}$$

Applying Proposition 3.8 together with (4.7), and using again Assumption A1, it is straightforward to observe that, for fixed n and τ,

$$\begin{aligned}\lim_{\varepsilon\to 0}\sup_{\alpha,\beta}\Big|&\int_{B_1}(1-\eta_\tau(x))\delta\phi_n(y_{n,\varepsilon},x)d\hat{\mu}^{\alpha\beta}_{x_{n,\varepsilon}}(x)\\&-\int_{B_1}(1-\eta_\tau(y))\delta\phi_n(y_{n,\varepsilon},y)d\hat{\mu}^{\alpha\beta}_{y_{n,\varepsilon}}(y)\Big|=0.\end{aligned}$$

On the other hand, since each ϕ_n is C^3, we have $|\delta\phi_n(y_{n,\varepsilon},x)|\le C_n|x|^2$ for all $x\in B_1$. Therefore,

$$\begin{aligned}\left|\int_{B_1}\eta_\tau(x)\delta\phi_n(y_{n,\varepsilon},x)d\hat{\mu}^{\alpha\beta}_{x_{n,\varepsilon}}(x)\right|&\le\int_{B_1}|\eta_\tau(x)\delta\phi_n(y_{n,\varepsilon},x)|d\hat{\mu}^{\alpha\beta}_{x_{n,\varepsilon}}(x)\\&\le C_n\int_{B_\tau}|x|^2d\hat{\mu}^{\alpha\beta}_{x_{n,\varepsilon}}(x),\\\left|\int_{B_1}\eta_\tau(y)\delta\phi_n(y_{n,\varepsilon},y)d\hat{\mu}^{\alpha\beta}_{y_{n,\varepsilon}}(y)\right|&\le C_n\int_{B_\tau}|y|^2d\hat{\mu}^{\alpha\beta}_{y_{n,\varepsilon}}(y),\end{aligned}$$

and we have

$$\limsup_{\tau\to 0}\sup_{\alpha,\beta}\int_{B_\tau}|x|^2d\hat{\mu}^{\alpha\beta}_{x_{n,\varepsilon}}(x)=\limsup_{\tau\to 0}\sup_{\alpha,\beta}\int_{B_\tau}|y|^2d\hat{\mu}^{\alpha\beta}_{y_{n,\varepsilon}}(y)=0.$$

Gathering these estimates, we conclude that for every n

$$\lim_{\varepsilon\to 0}\sup_{\alpha,\beta}\left|\hat{L}^{\alpha\beta}(\phi_n,x_{n,\varepsilon})-\hat{L}^{\alpha\beta}(\phi_n,y_{n,\varepsilon})\right|=0. \tag{7.9}$$

Step 4 (Using the subsolution and supersolution property). Using Definition 4.2, we obtain from (7.7) that

$$\begin{aligned}\lambda(u(x_{n,\varepsilon})-v(y_{n,\varepsilon})) \leq{}& \frac{C}{\varepsilon}|x_{n,\varepsilon}-y_{n,\varepsilon}|^{p-1}+(1+\|v\|_{L^\infty})\theta(|x_{n,\varepsilon}-y_{n,\varepsilon}|)\\ &+\sup_{\alpha\beta}\left\{\check{L}^{\alpha\beta}(u,x_{n,\varepsilon})-\check{L}^{\alpha\beta}(v,y_{n,\varepsilon})\right\}\\ &+\sup_{\alpha\beta}\left\{\hat{L}^{\alpha\beta}(\phi_n,x_{n,\varepsilon})-\hat{L}^{\alpha\beta}(\phi_n,y_{n,\varepsilon})\right\}.\end{aligned} \tag{7.10}$$

Letting $\kappa\to 0$ first and then $\varepsilon\to 0$ in (7.10), and using (7.1), (7.8), and (7.9), we now obtain

$$0<\lambda l\leq o_{\frac{1}{n}}(1),$$

which gives a contradiction. □

The following is an example of measures satisfying the assumptions of Theorem 7.4.

Example 7.6. Let $d\mu_x^{\alpha\beta}(z):=K_{\alpha\beta}(x,z)dz$ be such that (5.12) and (5.15) hold for some $p\in(1,2]$ and $\gamma\in(0,1]$, where K satisfies (5.13). Using (5.16), Assumption A1 is satisfied with $s=1-\frac{1}{p}$ if $\gamma=p-1$.

Assumption F. There are $\sigma\in(1,2)$ and positive constants $\bar{\lambda}<\bar{\Lambda}$ such that the measures $\mu_x^{\alpha\beta}$ are all of the form $d\mu_x^{\alpha\beta}(z)=K_{\alpha\beta}(x,z)dz$, with $K_{\alpha\beta}(x,z)=K_{\alpha\beta}(x,-z)$, and

$$\frac{\bar{\lambda}}{|z|^{d+\sigma}}\leq K_{\alpha\beta}(x,z)\leq\frac{\bar{\Lambda}}{|z|^{d+\sigma}} \tag{7.11}$$

and

$$|K_{\alpha\beta}(x,z)-K_{\alpha\beta}(y,z)|\leq\frac{\bar{\Lambda}|x-y|^\gamma}{|z|^{d+\sigma}}.$$

Corollary 7.7. *Let the measures $\mu_x^{\alpha\beta}$ be as above. Assume that Assumptions* C, D, *and* F *hold with some $\sigma\in(1,2)$ and $\gamma>\sigma-1$. Then, given a viscosity solution u and a viscosity subsolution* (*resp., supersolution*) *v of* (1.1) *such that $v\leq u$* (*resp., $u\leq v$*) *in $\mathcal{O}^c$, we have*

$$v\leq u \text{ in } \mathcal{O}\ \ (\text{resp.}, u\leq v \text{ in } \mathcal{O}).$$

Proof. The proof is an immediate application of Theorem 7.4 with $p=1+\gamma$, the computation in Example 5.12, and the fact that $u\in C^1(\mathcal{O})$ by [17, Theorem 4.1]. □

Remark 7.8. The comparison result of Corollary 7.7 can be extended to the case $\sigma=1$ if we use Theorem 7.3 instead of Corollary 7.7. The result of Corollary 7.7 can be also extended to the case $\lambda=0$; see [18, Theorem 4.1].

It is worth noting that in [18] the second and third authors obtained uniqueness results under similar assumptions to those of Corollaries 4.5 and 7.7, including Lipschitz-type assumption on the continuity of the kernels with respect to x. However, uniqueness results in [18] cover only σ in the range $(0,3/2)$, whereas the combination of Corollaries 4.5 and 7.7 (see also Remark 7.8) covers all σ up to 2.

Remark 7.9. The assumption (7.11) used in Corollary 7.7 can be relaxed a great deal. This assumption was used merely in order to guarantee that the viscosity solution is $C^{1+\alpha}$ in the interior. Indeed, interior C^α and $C^{1+\alpha}$ regularity estimates are now available for nonlocal equations for a far larger class of kernels, including those $K(x,z)$ which may not be symmetric in z or which vanish even for large sets of directions of z. See the works of Schwab and Silvestre [21, Section 8] and Kriventsov [17].

Appendix A. A variant of the optimal transportation problem

In this appendix, which follows [13], we describe the optimal transport problem with boundary. Throughout we make the following assumptions: Ω is an open subset of $\mathbb{R}^d$, and Γ is a compact subset of $\overline{\Omega}$. We are also given a function $c : \overline{\Omega}\times\overline{\Omega}\to\mathbb{R}$, known as the cost. We impose several assumptions on $c(x,y)$ and Γ, recorded in (1.1) and (1.2). Q5

First, we assume that c satisfies the following:

(1.1)
$$c(x,y) \text{ is continuous};\quad c(x,y)=c(y,x),\ c(x,x)=0,\ c(x,y)>0 \quad \text{if } x\neq y\ \forall x,y.$$

Second, Γ and c must be such that there is a measurable function

$$P:\overline{\Omega}\to\Gamma$$

which plays the role of the "projection" onto Γ, in the sense that

$$c(x,P(x)) = \inf_{y\in\Gamma} c(x,y). \tag{1.2}$$

Definition A.1. Letting E be a Borel subset of $\overline{\Omega}$, we define the function

$$c(x,E) = \inf_{y\in E} c(x,y).$$

Lastly, the following auxiliary cost will be relevant in what follows:

$$\tilde{c}(x,y) = \min\{c(x,y),\ c(x,\Gamma)+c(y,\Gamma)\}.$$

We also consider the set

$$\mathcal{K} = \{(x,y)\in\overline{\Omega}\times\overline{\Omega}\ \mid\ c(x,y)\le c(x,\Gamma)+c(y,\Gamma)\}. \tag{1.3}$$

Definition A.2. Given Ω and Γ, we let $\mathcal{M}_c(\overline{\Omega})$ be the set of positive Borel measures μ on $\overline{\Omega}\setminus\Gamma$ such that

$$\int_{\overline{\Omega}} c(x,\Gamma)\,d\mu(x) < \infty,$$

and

$$\mu(\{x\in\overline{\Omega}\mid d(x,\Gamma)>r\})<\infty \qquad \text{for every } r>0.$$

Definition A.3. Let $\mu,\nu\in\mathcal{M}_c(\overline{\Omega})$. By an admissible coupling of μ and ν, we mean a positive Borel measure γ over $\overline{\Omega}\times\overline{\Omega}$, satisfying $\gamma(\Gamma\times\Gamma)=0$ and

$$\pi^1_{\#}\gamma\mid_{\overline{\Omega}\setminus\Gamma}=\mu, \qquad \pi^2_{\#}\gamma\mid_{\overline{\Omega}\setminus\Gamma}=\nu.$$

The set of admissible couplings will be denoted by $\mathrm{Adm}_\Gamma(\mu,\nu)$.

Note that a measure in $\mathcal{M}_c(\overline{\Omega})$ may fail to have finite mass since $\inf c(x,\Gamma)=0$. We are now ready to state the optimal transport problem with boundary.

Problem A.4. Consider two measures $\mu, \nu \in \mathcal{M}_c(\overline{\Omega})$. Among all admissible measures $\gamma \in \mathrm{Adm}_\Gamma(\mu,\nu)$, find one that minimizes the functional

$$J_c(\gamma) := \int_{\overline{\Omega}\times\overline{\Omega}} c(x,y) d\gamma(x,y).$$

We make no claim as to whether all of the assumptions on the cost and Ω are necessary, but they are sufficiently general for our purposes and make most of the proofs relatively straightforward (for instance, the symmetry assumption on $c(x,y)$ is not necessary but makes the notation simpler). In any case, the costs we care about in the main body of the paper are

$$c_p(x,y) := |x-y|^p, \qquad 1 \leq p \leq 2.$$

Since we are specially concerned with these costs, we shall write $J_p(\gamma)$ to refer to the above functional when the cost is $c_p(x,y)$. At the same time, the main Ω and Γ we care about are

$$\Omega = \mathbb{R}^d \setminus \{0\}$$

and

$$\Gamma = \{0\}.$$

Evidently, these sets, together with the costs c_p, comply with our requirements. The first basic fact about Problem A.4 is the existence of minimizers. The proof is Q6 essentially the same as in the optimal transport case (compactness of the measures and lower semicontinuity of $J_c(\gamma)$) (cf. [4, Theorem 1.5], [13, Section 2]).

Theorem A.5. *Let $\mu, \nu \in \mathcal{M}_c(\overline{\Omega})$. Then $J_c(\gamma) < \infty$ for at least one $\gamma^* \in \mathrm{Adm}_\Gamma(\mu,\nu)$. Moreover, there exists at least one minimizer γ for Problem* A.4.

Proof. With the map P being as in (1.2), we define the measure

$$\gamma^* := (\mathrm{Id} \times P)_{\#}\mu + (P \times \mathrm{Id})_{\#}\nu.$$

It is clear that $\gamma^* \in \mathrm{Adm}_\Gamma(\mu,\nu)$. At the same time,

$$J_c(\gamma^*) = \int_{\overline{\Omega}\times\overline{\Omega}} c(x,y)\, d\gamma^*(x,y) = \int_\Omega c(x,\Gamma)\, d\mu(x) + \int_\Omega c(y,\Gamma)\, d\nu(y),$$

and thus $J_c(\gamma^*) < \infty$ since $\mu, \nu \in \mathcal{M}_c(\overline{\Omega})$.

In order to prove that the infimum is achieved, we will first prove that $\mathrm{Adm}_\Gamma(\mu,\nu)$ is compact with respect to a certain notion of convergence. Let K be any compact subset of $\overline{\Omega}\times\overline{\Omega}\setminus\Gamma\times\Gamma$. Since $\Gamma\times\Gamma$ and K are compact, we have $d(K,\Gamma\times\Gamma) > 0$. Then there exists a compact subset $\tilde{K}$ of $\overline{\Omega}\setminus\Gamma$ such that $K \subset (\tilde{K}\times\overline{\Omega})\cup(\overline{\Omega}\times\tilde{K})$. Since Γ is compact and (1.1) holds, there is an $\varepsilon_0 > 0$ such that $\inf_{x\in\tilde{K}} c(x,\Gamma) > \varepsilon_0$. And thus $\mu(\tilde{K}) < +\infty$ since $\mu \in \mathcal{M}_c(\overline{\Omega})$. Similarly, we have $\nu(\tilde{K}) < +\infty$. Therefore, if $\gamma \in \mathrm{Adm}_\Gamma(\mu,\nu)$, we have

$$\gamma(K) \leq \mu(\tilde{K}) + \nu(\tilde{K}) < \infty.$$

Since $\mu(\tilde{K}) + \nu(\tilde{K})$ is independent of γ, it follows that, given a sequence $\{\gamma_n\}$ in $\mathrm{Adm}_\Gamma(\mu,\nu)$, there is a subsequence γ_{n_k} and a measure γ in $\overline{\Omega}\times\overline{\Omega}$ such that $\gamma_{n_k} \rightharpoonup \gamma$, with the convergence being in the following sense:

(1.4)

$$\int_{\overline{\Omega}\times\overline{\Omega}} \phi(x,y)\, d\gamma(x,y) = \lim_{k\to\infty} \int_{\overline{\Omega}\times\overline{\Omega}} \phi(x,y)\, d\gamma_{n_k}(x,y), \qquad \phi \in C_c^0(\overline{\Omega}\times\overline{\Omega}\setminus\Gamma\times\Gamma).$$

Now we must show that $\gamma \in \mathrm{Adm}_\Gamma(\mu,\nu)$. Observe that for each n

$$\gamma_n \in \mathrm{Adm}_\Gamma(\mu,\nu) \Rightarrow$$
$$\gamma_n(\{(x,y) \mid d(x,\Gamma) \geq r \text{ or } d(y,\Gamma) \geq r\}) \leq \mu(\{d(x,\Gamma) \geq r\}) + \nu(\{d(x,\Gamma) \geq r\}),$$

so from the assumptions on μ and ν (Definition A.2) it follows that the right-hand side goes to 0 as $r \to \infty$ with a rate depending only on μ and ν (note that when $\overline{\Omega}$ is compact, this last assertion holds trivially). From this estimate and the convergence in (1.4) it is not hard to see that

$$\int_{\overline{\Omega}\times\overline{\Omega}} \phi(x)\ d\gamma(x,y) = \lim_{k\to\infty} \int_{\overline{\Omega}\times\overline{\Omega}} \phi(x)\ d\gamma_{n_k}(x,y) \qquad \forall \phi \in C^0_c(\overline{\Omega}\setminus\Gamma);$$

a similar statement holds for functions of y with support away from Γ. In particular,

$$\int_{\overline{\Omega}\times\overline{\Omega}} \phi(x)\ d\gamma(x,y) = \int_{\overline{\Omega}} \phi(x)\ d\mu(x), \qquad \int_{\overline{\Omega}\times\overline{\Omega}} \psi(x)\ d\gamma(x,y) = \int_{\overline{\Omega}} \psi(y)\ d\nu(y),$$

which shows that $\gamma \in \mathrm{Adm}_\Gamma(\mu,\nu)$. In conclusion, the set of admissible couplings $\mathrm{Adm}_\Gamma(\mu,\nu)$ is sequentially compact with respect to the notion of convergence in (1.4).

Let γ_n be a minimizing sequence in $\mathrm{Adm}_\Gamma(\mu,\nu)$, that is, a sequence such that $J_c(\gamma_n) \to \inf J_c$ as $n \to \infty$. At the same time, let c_k be a monotone increasing sequence of continuous functions with compact support in $\overline{\Omega}\times\overline{\Omega}\setminus\Gamma\times\Gamma$ and such that $c_k(x,y) \to c(x,y)$ locally uniformly in $\overline{\Omega}\times\overline{\Omega}\setminus\Gamma\times\Gamma$. Using a diagonal argument, there exist a subsequence, still denoted by γ_n, and $\gamma_* \in \mathrm{Adm}_\Gamma(\mu,\nu)$ such that for every fixed k

$$\lim_{n\to\infty} \int_{\overline{\Omega}\times\overline{\Omega}} c_k(x,y)\ d\gamma_n(x,y) = \int_{\overline{\Omega}\times\overline{\Omega}} c_k(x,y)\ d\gamma_*(x,y).$$

Now, by the monotonicity of the c_k, we have

$$J_c(\gamma_*) = \int_{\overline{\Omega}\times\overline{\Omega}} c(x,y)\ d\gamma_*(x,y) = \sup_k J_{c_k}(\gamma_*),$$

while for any k we have

$$J_{c_k}(\gamma_*) = \lim_{n\to\infty} \int_{\overline{\Omega}\times\overline{\Omega}} c_k(x,y)\ d\gamma_n(x,y) \leq \lim_{n\to\infty} J(\gamma_n) = \inf_{\gamma\in\mathrm{Adm}_\Gamma(\mu,\nu)} J(\gamma).$$

This proves that γ_* achieves the minimum value of J_c among all admissible plans. □

We now characterize minimizers for Problem A.4 using c-concave functions and c-cyclical monotonicity.

Definition A.6. For a function $\phi : \overline{\Omega} \to \mathbb{R}\cup\{\pm\infty\}$ with $\phi \neq -\infty$ for at least some x, its c-transform $\phi^c : \overline{\Omega} \to \mathbb{R}\cup\{-\infty\}$ is the function given by

$$\phi^c(y) = \inf_{x\in\overline{\Omega}} \big\{c(x,y) - \phi(x)\big\}.$$

A function ϕ is said to be c-concave if there is some ψ such that

$$\phi = \psi^c.$$

If ϕ and ψ are two c-concave functions such that $\phi = \psi^c$ and $\psi = \phi^c$, then we say they are c-conjugate to one another. Just the same, we talk about $\tilde{c}$-transforms and $\tilde{c}$-concave functions.

Remark A.7. Since the cost is assumed to be continuous, it follows that ϕ^c is the infimum of a family of continuous functions of y ($\{y \mapsto c(x,y)-\phi(x)\}_x$), accordingly, ϕ^c is upper semicontinuous. In particular, if (ϕ, ψ) is a c-conjugate pair, then both ϕ and ψ are upper semicontinuous functions.

Remark A.8. Suppose that (ϕ, ψ) are c-conjugate. Then for every x and y we have

$$\phi(x) + \psi(y) \le c(x, y).$$

The set of pairs (x, y) for which we have equality will be important in what follows.

Definition A.9. Let ϕ be a c-concave function, and let $\psi = \phi^c$. The c-subdifferential of ϕ, denoted by $\partial^c \phi$, is defined as the set of pairs (x, y) such that

$$\phi(x) + \psi(y) = c(x, y).$$

Moreover, for each x we define $\partial^c\phi(x)$ to be the set of all y such that $(x, y) \in \partial^c\phi$. We define $\partial^{\tilde{c}}\phi$ and $\partial^{\tilde{c}}(x)$ for a $\tilde{c}$-concave ϕ in the same manner.

Definition A.10. A subset of $\overline{\Omega} \times \overline{\Omega}$ is said to be c-cyclically monotone if, given a finite sequence $\{(x_i, y_i)\}_{i=0}^n$ and any permutation σ, we have

$$\sum_{i=1}^n c(x_i, y_i) \le \sum_{i=1}^n c(x_i, y_{\sigma(i)}).$$

If c is replaced by $\tilde{c}$, we have $\tilde{c}$-cyclical monotonicity.

The following proposition is a (minor) modification of a well-known convex analysis result of Rockafellar (previously extended for c-concave functions). This modification pertains the set Γ and the costs $c(x, y)$ and $\tilde{c}(x, y)$.

Proposition A.11. *Let γ be a measure concentrated on $\mathcal{K}$ and such that* $\mathrm{spt}(\gamma) \cup \Gamma \times \Gamma$ *is $\tilde{c}$-cyclically monotone. Then there are c-conjugate functions ϕ and ψ such that*

$$\phi \equiv \psi \equiv 0 \qquad \text{on } \Gamma$$

and

$$\mathrm{spt}(\gamma) \subset \partial^c \phi.$$

Proof. This follows from standard optimal transport theory. Indeed, as shown in [4, proof of Theorem 1.13, (ii) $\Rightarrow$ (iii)], since $\mathrm{spt}(\gamma) \cup \Gamma \times \Gamma$ is $\tilde{c}$-cyclically monotone, there must be a $\tilde{c}$-concave function ϕ such that

$$\mathrm{spt}(\gamma) \cup \Gamma \times \Gamma \subset \partial^{\tilde{c}} \phi.$$

Since any pair $(x, y) \in \Gamma \times \Gamma$ belongs to $\partial^{\tilde{c}}\phi$, it follows that

$$\phi(x) + \phi^{\tilde{c}}(y) = \tilde{c}(x, y) = 0 \qquad \forall x, y \in \Gamma.$$

We emphasize that the above holds for any two points x and y in Γ, which in particular means that ϕ and $\phi^{\tilde{c}}$ are constant on Γ. Adding a constant to ϕ, we can assume without loss of generality that $\phi = 0$ on Γ, which in turn guarantees that $\phi^{\tilde{c}} = 0$ on Γ as well.

We claim that $\partial^{\tilde{c}}\phi \cap \mathcal{K} \subset \partial^c\phi$. Indeed, if $(x_0, y_0) \in \mathcal{K}$ is such that $y_0 \in \partial^{\tilde{c}}\phi(x_0)$, then

$$\phi(x) \le \tilde{c}(x, y_0) - \phi^c(y_0) \le c(x, y_0) - \phi^c(y_0)$$

since $\tilde{c}(x,y) \leq c(x,y)$ for all x and y. Since $(x_0,y_0) \in \mathcal{K}$, we have $\tilde{c}(x_0,y_0) = c(x_0,y_0)$, so

$$\phi(x_0) = \tilde{c}(x,y_0) - \phi^c(y_0) = c(x,y_0) - \phi^c(y_0).$$

It follows from this that $(x_0,y_0) \in \partial^c\phi(x)$, and the claim is proved. The same argument also shows that if $y \in \Gamma$, then $\phi^c(y) = \phi^{\tilde{c}}(y) = 0$. Since ϕ was chosen so that $\mathrm{spt}(\gamma) \subset \partial^{\tilde{c}}\phi$ and γ is supported in $\mathcal{K}$, it follows that $\mathrm{spt}(\gamma) \subset \partial^c\phi$. Therefore, ϕ and ϕ^c are the desired c-conjugate functions. $\square$

As in the usual optimal transport problem, a basic tool for the analysis of Problem A.4 is a *dual problem.* This problem deals with a family of admissible pairs of functions,

$$\begin{aligned}\mathrm{Adm}^c := \Big\{(\phi,\psi) \mid\ & \phi \in L^1(\mu) \text{ and } \psi \in L^1(\nu),\\ & \phi \text{ and } \psi \text{ are upper semicontinuous},\\ & \phi \equiv \psi \equiv 0 \text{ on } \Gamma,\\ & \text{and } \phi(x)+\psi(y) \leq c(x,y) \text{ in } \overline{\Omega}\times\overline{\Omega}\Big\}.\end{aligned} \tag{1.5}$$

We now can state the problem dual to Problem (A.12).

Problem A.12. Among all pairs $(\phi,\psi) \in \mathrm{Adm}^c$, find one that maximizes the functional

$$J^*(\phi,\psi) := \int_{\overline{\Omega}} \phi(x)\, d\mu(x) + \int_{\overline{\Omega}} \psi(y)\, d\nu(y).$$

The characterization of minimizers in Problem A.4 and maximizers for Problem A.12 is the content of Theorem A.13 and Lemma A.14. In the proof we will make use of Proposition A.11, together with the characterization of optimizers for the usual optimal transportation problem [4, Theorem 1.13].

Theorem A.13. *Let $\gamma \in \mathrm{Adm}_\Gamma(\mu,\nu)$ for two measures $\mu,\nu \in \mathcal{M}_c(\Omega)$. Then γ is a minimizer for Problem* A.4 *if and only if γ is concentrated on the set $\mathcal{K}$ defined in* (1.3)*, and* $\mathrm{spt}(\gamma) \cup \Gamma\times\Gamma$ *is a $\tilde{c}$-cyclically monotone set.*

Proof. Assume first that $\gamma \in \mathrm{Adm}_\Gamma(\mu,\nu)$ is optimal. Consider $\tilde{\gamma}$, the plan given by $\tilde{\gamma} = \hat{\gamma}_{|\overline{\Omega}\times\overline{\Omega}\setminus\Gamma\times\Gamma}$, where $\hat{\gamma}$ is defined as

$$\hat{\gamma} := \gamma_{|\mathcal{K}} + (\pi_1, P\circ\pi_1)_\#\left(\gamma_{|\overline{\Omega}\times\overline{\Omega}\setminus\mathcal{K}}\right) + (P\circ\pi_2,\pi_2)_\#\left(\gamma_{|\overline{\Omega}\times\overline{\Omega}\setminus\mathcal{K}}\right);$$

here P is as in (1.2). What the plan $\hat{\gamma}$ is meant to do is to adjust the original plan γ by shifting the transport of some of the mass so that it is sent to Γ whenever it is advantageous to do so (and only for points (x,y) outside of $\mathcal{K}$). The coupling $\tilde{\gamma}$ comes from taking $\hat{\gamma}$ and discarding any potential mass $\Gamma\times\Gamma$; this makes sure we have an admissible coupling. Therefore, $\tilde{\gamma} \in \mathrm{Adm}_\Gamma(\mu,\nu)$. Moreover, we have the formula

$$\int_{\overline{\Omega}\times\overline{\Omega}} c(x,y)\, d\tilde{\gamma}(x,y) = \int_{\mathcal{K}} c(x,y) d\gamma(x,y) + \int_{\overline{\Omega}\times\overline{\Omega}\setminus\mathcal{K}} [c(x,\Gamma)+c(\Gamma,y)]\, d\gamma(x,y).$$

From the definition of $\mathcal{K}$ we have $c(x,y) > c(x,\Gamma)+c(\Gamma,y)$ outside of $\mathcal{K}$; thus

$$\int_{\overline{\Omega}\times\overline{\Omega}\setminus\mathcal{K}} [c(x,\Gamma)+c(\Gamma,y)]\, d\gamma(x,y) \leq \int_{\overline{\Omega}\times\overline{\Omega}\setminus\mathcal{K}} c(x,y)\, d\gamma(x,y).$$

It follows that

$$\int_{\overline{\Omega}\times\overline{\Omega}} c(x,y)\, d\tilde{\gamma}(x,y) \leq \int_{\overline{\Omega}\times\overline{\Omega}} c(x,y)\, d\gamma(x,y),$$

with strict inequality if and only if $\gamma(\overline{\Omega}\times\overline{\Omega}\setminus\mathcal{K})>0$. By the optimality of γ we then conclude that $\gamma(\overline{\Omega}\times\overline{\Omega}\setminus\mathcal{K})=0$; that is, γ is supported in $\mathcal{K}$.

Now we must show that $\mathrm{spt}(\gamma)\cup\Gamma\times\Gamma$ is $\tilde{c}$-monotone. We deal first with the case where γ has finite mass. In this instance, let us write

$$\bar{\mu}=\pi^1_{\#}\gamma, \qquad \bar{\nu}=\pi^2_{\#}\gamma. \tag{1.6}$$

Then, as $\bar{\mu}$ and $\bar{\nu}$ are the marginals of γ (in all of $\overline{\Omega}$), they must have the same total mass which is finite since γ has finite mass. Let γ_0 denote the optimal transport plan between $\bar{\mu}$ and $\bar{\nu}$ according to $\tilde{c}$, and let $\tilde{\gamma}_0$ be constructed from γ_0 in the same way as $\tilde{\gamma}$ was constructed from γ (first, by pushing parts of its mass to the boundary as done above, yielding a measure $\hat{\gamma}_0$, and then restricting to $\overline{\Omega}\times\overline{\Omega}\setminus\Gamma\times\Gamma$). Since $\bar{\mu}_{|\overline{\Omega}}=\mu$ and $\bar{\nu}_{|\overline{\Omega}}=\nu$, we have $\tilde{\gamma}_0$ being a measure in $\mathrm{Adm}_\Gamma(\mu,\nu)$, and as argued above for γ and $\tilde{\gamma}$, if γ_0 were not supported in $\mathcal{K}$, then $\tilde{\gamma}_0$ would be a better coupling. This shows that $c=\tilde{c}$ γ-a.e. and γ_0-a.e., and thus

$$J_{\tilde{c}}(\gamma)=J_c(\gamma), \qquad J_{\tilde{c}}(\gamma_0)=J_c(\gamma_0).$$

Combining these identities with the optimality of γ and γ_0 yields the inequalities

$$J_{\tilde{c}}(\gamma)\geq J_{\tilde{c}}(\gamma_0)=J_c(\gamma_0)\geq J_c(\gamma_{0|\overline{\Omega}\times\overline{\Omega}\setminus\Gamma\times\Gamma})\geq J_c(\gamma)$$

(we used the facts that $\gamma_{0|\overline{\Omega}\times\overline{\Omega}\setminus\Gamma\times\Gamma}\in\mathrm{Adm}_\Gamma(\mu,\nu)$ and that $c(x,y)\geq 0$), and we conclude that

$$J_{\tilde{c}}(\gamma_0)=J_{\tilde{c}}(\gamma).$$

Thus γ is an optimal plan for the usual transport problem with cost $\tilde{c}$. By optimal transport theory, the support set $\mathrm{spt}(\gamma)$ is $\tilde{c}$-cyclically monotone. To prove that $\mathrm{spt}(\gamma)\cup\Gamma\times\Gamma$ is still $\tilde{c}$-cyclically monotone, simply note that if γ_0 is any measure supported in $\Gamma\times\Gamma$, then $\gamma+\gamma_0$ may not belong to $\mathrm{Adm}_\Gamma(\mu,\nu)$, but arguing as above, we can show that it is optimal for the standard optimal transport problem with cost $\tilde{c}$ and marginals $\pi^1_{\#}(\gamma+\gamma_0)$ and $\pi^2_{\#}(\gamma+\gamma_0)$. This shows that $\mathrm{spt}(\gamma+\gamma_0)=\mathrm{spt}(\gamma)\cup\Gamma\times\Gamma$ is $\tilde{c}$-cyclically monotone.

This covers the case where γ has finite mass. For the general case we argue just as in [13, Proposition 2.3]: that the one property from the classical optimal transport problem that we needed was that if the support of γ is not $\tilde{c}$-cyclically monotone, then γ cannot be optimal with respect to $\tilde{c}$. It is worth noting that that even if $\bar{\mu}$ and $\bar{\nu}$ do not have finite mass, they are still the marginals of γ by definition (1.6), so the set of measures with marginals $\bar{\mu}$ and $\bar{\nu}$ is nonempty, so one can proceed with the Kantorovich problem as in standard optimal transport theory. Therefore, the above argument extends to the case of γ with infinite mass, and we conclude that $\mathrm{spt}(\gamma)\cup\Gamma\times\Gamma$ is $\tilde{c}$-cyclically monotone in all cases.

Conversely, assume that γ is supported in $\mathcal{K}$ and that $\mathrm{spt}(\gamma)\cup\Gamma\times\Gamma$ is a $\tilde{c}$-cyclically monotone set. Then Proposition A.11 says that there is a function ϕ which is c-concave, such that ϕ and ϕ^c both vanish on Γ, and

$$\mathrm{spt}(\gamma)\subset\partial^c\phi.$$

In particular, this means that $\phi(x)+\phi^c(y)=c(x,y)$ on $\mathrm{spt}(\gamma)$, so

$$\begin{aligned}\int_{\overline{\Omega}\times\overline{\Omega}} c(x,y)\,d\gamma(x,y) &= \int_{\overline{\Omega}\times\overline{\Omega}}[\phi(x)+\phi^c(y)]\,d\gamma(x,y),\\ &= \int_{(\overline{\Omega}\setminus\Gamma)\times\overline{\Omega}} \phi(x)\,d\gamma(x,y)+\int_{\overline{\Omega}\times(\overline{\Omega}\setminus\Gamma)} \phi^c(y)\,d\gamma(x,y),\\ &= \int_{\overline{\Omega}\setminus\Gamma} \phi(x)\,d\mu(x)+\int_{\overline{\Omega}\setminus\Gamma} \phi^c(y)\,d\nu(y).\end{aligned}$$

This suffices to guarantee the optimality of γ. Indeed, take any $\tilde{\gamma}\in \mathrm{Adm}_\Gamma(\mu,\nu)$; then

$$\begin{aligned}\int_{\overline{\Omega}\times\overline{\Omega}} c(x,y)\,d\tilde{\gamma}(x,y) &\geq \int_{\overline{\Omega}\times\overline{\Omega}}[\phi(x)+\phi^c(y)]\,d\tilde{\gamma}(x,y)\\ &= \int_{\overline{\Omega}\setminus\Gamma}\phi(x)\,d\mu(x)+\int_{\overline{\Omega}\setminus\Gamma}\phi^c(y)\,d\nu(y)=\int_{\overline{\Omega}\times\overline{\Omega}}c(x,y)\,d\gamma(x,y),\end{aligned}$$

and we conclude that γ achieves the minimum value. $\square$

Just as in the usual optimal transport problem, a solution to Problem A.4 corresponds to a solution to Problem A.12, and the corresponding values coincide.

Lemma A.14. *The problems* (A.4) *and* (A.12) *are dual, meaning that*

$$\inf_{\gamma\in\mathrm{Adm}_\Gamma(\mu,\nu)} J(\gamma)=\sup_{(\phi,\psi)\in\mathrm{Adm}^c} J^*(\phi,\psi).$$

Proof. If $(\phi,\psi)\in\mathrm{Adm}^c$, then $\phi(x)+\psi(y)\leq c(x,y)$ for all x and y, and $\phi\equiv\psi\equiv 0$ on Γ. Therefore, for any $\gamma\in\mathrm{Adm}_\Gamma(\mu,\nu)$ we have

$$\begin{aligned}\int_{\overline{\Omega}\times\overline{\Omega}} c(x,y)\,d\gamma(x,y) &\geq \int_{\overline{\Omega}\times\overline{\Omega}}[\phi(x)+\psi(y)]\,d\gamma(x,y)\\ &= \int_{\overline{\Omega}\setminus\Gamma}\phi(x)\,d\mu(x)+\int_{\overline{\Omega}\setminus\Gamma}\psi(y)\,d\nu(y)\\ &= \int_{\overline{\Omega}}\phi(x)\,d\mu(x)+\int_{\overline{\Omega}}\psi(y)\,d\nu(y).\end{aligned}$$

Since $(\phi,\psi)\in\mathrm{Adm}^c$ and $\gamma\in\mathrm{Adm}_\Gamma(\mu,\nu)$ were arbitrary, it follows that

$$\inf_{\gamma\in\mathrm{Adm}_\Gamma(\mu,\nu)}\int_{\overline{\Omega}\times\overline{\Omega}}c(x,y)\,d\gamma(x,y)\geq \sup_{(\phi,\psi)\in\mathrm{Adm}^c}\Big\{\int_{\overline{\Omega}}\phi\,d\mu(x)+\int_{\overline{\Omega}}\psi\,d\nu(y)\Big\}. \tag{1.7}$$

The reverse inequality follows from Theorem A.13. To see why, let $\pi\in\mathrm{Adm}_\Gamma(\mu,\nu)$ be the minimizer. Then the theorem says that $\mathrm{spt}(\gamma)\cup\Gamma\times\Gamma$ is $\tilde{c}$-cyclically monotone, and its support is contained in $\mathcal{K}$, in which case Proposition A.11 says that there are functions ϕ and ψ which are c-conjugate, vanish on Γ, and are such that $\phi(x)+\psi(y)=c(x,y)$ for γ-a.e. (x,y). The functions ϕ,ψ have a couple of extra properties. First, since $\psi(y)=0$ for $y\in\Gamma$, we have $\phi(x)\leq c(x,y)$ for every $y\in\Gamma$, and taking the infimum in y, it follows that

$$\phi(x)\leq c(x,\Gamma)\qquad\forall x\in\Gamma.$$

Likewise, it follows that $\psi(y)\leq c(y,\Gamma)$ for every $y\in\Gamma$. This implies that

$$\max\{\phi,0\}\in L^1(\mu),\qquad \max\{\psi,0\}\in L^1(\nu). \tag{1.8}$$

In particular, the integrals $\int_{\overline{\Omega}} \phi(x)d\mu(x)$ and $\int_{\overline{\Omega}} \psi(y)d\nu(y)$ are well defined. Second, using that $\phi(x)+\psi(y) = c(x,y)$ for γ-a.e. (x,y), that $\phi \equiv \psi \equiv 0$ on Γ, and that $\gamma \in \mathrm{Adm}_\Gamma(\mu,\nu)$, it follows that

$$\begin{aligned}\int_{\overline{\Omega}} \phi(x)\ d\mu(x) + \int_{\overline{\Omega}} \psi(y)\ d\nu(y) &= \int_{\overline{\Omega}\setminus\Gamma} \phi(x)\ d\mu(x) + \int_{\overline{\Omega}\setminus\Gamma} \psi(y)\ d\nu(y)\\ &= \int_{\overline{\Omega}\times\overline{\Omega}} [\phi(x)+\psi(y)]\ d\gamma(x,y)\\ &= \int_{\overline{\Omega}\times\overline{\Omega}} c(x,y)\ d\gamma(x,y).\end{aligned}$$

Since this last integral is finite, it follows that $\int_{\overline{\Omega}} \phi(x)d\mu(x)$ and $\int_{\overline{\Omega}} \psi(y)d\nu(y)$ are finite, and in light of (1.8) it follows that $\phi \in L^1(\mu)$ and $\psi \in L^1(\nu)$. This shows that $(\phi,\psi) \in \mathrm{Adm}^c$, and this yields the reverse inequality of (1.7), proving the lemma. □

The following lemma is a minor modification of [13, Lemma 2.1], and we omit its proof. The lemma itself is a variant of a standard lemma in optimal transport theory [5, Lemma 5.3.2]. We recall that, below, $\mathcal{M}_p(\overline{\Omega}) := \mathcal{M}_c(\overline{\Omega})$ for $c(x,y) = |x-y|^p$.

Lemma A.15. *Let $p \geq 1$, and consider measures $\mu_1,\mu_2,\mu_3 \in \mathcal{M}_p(\overline{\Omega})$, $\gamma^{12} \in \mathrm{Adm}_\Gamma(\mu_1,\mu_2)$, and $\gamma^{23} \in \mathrm{Adm}_\Gamma(\mu_2,\mu_3)$. Then there is a Borel measure in $\overline{\Omega}\times\overline{\Omega}\times\overline{\Omega}$, denoted γ^{123}, whose 2-marginals satisfy*

$$\pi^{12}_{\#}\gamma^{123} = \gamma^{12} + \sigma^{12}, \qquad \pi^{23}_{\#}\gamma^{123} = \gamma^{23} + \sigma^{23}, \tag{1.9}$$

where σ^{12} and σ^{23} are measures concentrated on the set $\{(x,x) \mid x \in \Gamma\}$, and $\pi^{12}(x_1,x_2,x_3) = (x_1,x_2), \pi^{2,3}(x_1,x_2,x_3) = (x_2,x_3)$.

We can now prove that $\mathrm{d}_{\mathbb{L}_p}(\mu,\nu)$ is a metric in $\mathcal{M}_p(\overline{\Omega})$.

Theorem A.16. *The quantity*

$$\mathrm{d}_{\mathbb{L}_p}(\mu,\nu) := \inf_{\gamma\in\mathrm{Adm}_\Gamma(\mu,\nu)} \left(\int_{\overline{\Omega}\times\overline{\Omega}} |x-y|^p d\gamma(x,y)\right)^{\frac{1}{p}}$$

defines a metric in $\mathcal{M}_p(\overline{\Omega})$.

Proof. It is clear that $\mathrm{d}_{\mathbb{L}_p}(\mu,\nu) = \mathrm{d}_{\mathbb{L}_p}(\nu,\mu)$ and that $\mathrm{d}_{\mathbb{L}_p}(\mu,\nu) \geq 0$ for all μ and ν. Moreover, if $\mathrm{d}_{\mathbb{L}_p}(\mu,\nu) = 0$, it means that there is some $\gamma \in \mathrm{Adm}_\Gamma(\mu,\nu)$ such that

$$0 = \int_{\overline{\Omega}\times\overline{\Omega}} |x-y|^p\ d\gamma(x,y) \Rightarrow \mathrm{spt}(\gamma) \subset \{(x,y) \in \overline{\Omega}\times\overline{\Omega}\ \mid x = y\}.$$

This implies that for any $\phi \in C^0_c(\overline{\Omega}\setminus\Gamma)$ we have

$$\int_{\overline{\Omega}\setminus\Gamma} \phi(x)\ d\mu(x) = \int_{\overline{\Omega}\times\overline{\Omega}} \phi(x)\ d\gamma(x,y) = \int_{\overline{\Omega}\times\overline{\Omega}} \phi(y)\ d\gamma(x,y) = \int_{\overline{\Omega}\setminus\Gamma} \phi(y)\ d\nu(y);$$

in other words, $\mu = \nu$. It remains to prove the triangle inequality. Consider measures μ_1,μ_2,μ_3 in $\mathcal{M}_p(\overline{\Omega})$, and let the measures $\gamma^{12} \in \mathrm{Adm}_\Gamma(\mu_1,\mu_2)$ and $\gamma^{23} \in \mathrm{Adm}_\Gamma(\mu_2,\mu_3)$ be optimizers for the respective problems. Then Lemma A.15 guarantees that there is a measure γ^{123} satisfying (1.9).

It will be convenient to denote an element $\overline{\Omega}\times\overline{\Omega}\times\overline{\Omega}$ as (x_1,x_2,x_3). At the same time, the "coordinates" x_1,x_2,x_3 define three functions $\overline{\Omega}\times\overline{\Omega}\times\overline{\Omega} \to \overline{\Omega} \subset \mathbb{R}^d$.

With this in mind, we note that the function $|x_1 - x_3|^p$ is independent of x_2, so (denoting $\pi^{13}(x_1, x_2, x_3) = (x_1, x_3)$)

$$\begin{aligned} \mathrm{d}_{\mathbb{L}_p}(\mu_1, \mu_3)^p &\leq \int_{\overline{\Omega}\times\overline{\Omega}} |x_1 - x_3|^p \, d\pi^{13}_{\#}\gamma^{123}(x_1, x_3) \\ &= \int_{\overline{\Omega}\times\overline{\Omega}\times\overline{\Omega}} |x_1 - x_3|^p \, d\gamma^{123}(x_1, x_2, x_3) \end{aligned} \tag{1.10}$$

On the other hand, applying Minkowski's inequality to $L^p(\overline{\Omega}\times\overline{\Omega}\times\overline{\Omega}, d\gamma^{123})$ for the functions $x_1 - x_2$ and $x_2 - x_3$, we have

$$\begin{aligned} &\left(\int_{\overline{\Omega}\times\overline{\Omega}} |x_1 - x_3|^p \, d\gamma^{123}(x_1, x_2, x_3)\right)^{\frac{1}{p}} \\ &\quad \leq \left(\int_{\overline{\Omega}\times\overline{\Omega}} |x_1 - x_2|^p \, d\gamma^{123}(x_1, x_2, x_3)\right)^{\frac{1}{p}} \\ &\quad\quad + \left(\int_{\overline{\Omega}\times\overline{\Omega}} |x_2 - x_3|^p \, d\gamma^{123}(x_1, x_2, x_3)\right)^{\frac{1}{p}}. \end{aligned}$$

Then, using the optimality of γ^{12} as well as (1.9),

$$\begin{aligned} \int_{\overline{\Omega}\times\overline{\Omega}} |x_1 - x_2|^p \, d\gamma^{123}(x_1, x_2, x_3) &= \int_{\overline{\Omega}\times\overline{\Omega}} |x_1 - x_2|^p \, d(\gamma^{12} + \sigma^{12})(x_1, x_2) \\ &= \int_{\overline{\Omega}\times\overline{\Omega}} |x_1 - x_2|^p(x_1, x_2) \, d\gamma^{12} = \mathrm{d}_{\mathbb{L}_p}(\mu_1, \mu_2)^p, \end{aligned}$$

where the second to last inequality used the fact that σ^{12} is supported on the diagonal so that σ^{12}-a.e. we have $|x_1 - x_2| = 0$. Just the same, we can see that

$$\int_{\overline{\Omega}\times\overline{\Omega}} |x_2 - x_3|^p \, d\gamma^{123}(x_1, x_2, x_3) = \mathrm{d}_{\mathbb{L}_p}(\mu_2, \mu_3)^p.$$

Then, recalling (1.10), we conclude that

$$\mathrm{d}_{\mathbb{L}_p}(\mu_1, \mu_3) \leq \mathrm{d}_{\mathbb{L}_p}(\mu_1, \mu_2) + \mathrm{d}_{\mathbb{L}_p}(\mu_2, \mu_3),$$

which finishes the proof that $\mathrm{d}_{\mathbb{L}_p}(\mu, \nu)$ is a metric. □

Proof of Proposition 3.8. For any $\gamma \in \mathrm{Adm}(\mu, \nu)$ (recall that now $\Gamma = \{0\}$) we have

$$\begin{aligned} \int_{B_1} \psi \, d\mu(x) - \int_{B_1} \psi \, d\nu(y) &= \int_{\mathrm{spt}(\psi)\times\mathbb{R}^d} \psi(x) \, d\gamma(x, y) - \int_{\mathbb{R}^d\times\mathrm{spt}(\psi)} \psi(y) \, d\gamma(x, y) \\ &= \int_{A_\psi} [\psi(x) - \psi(y)] \, d\gamma(x, y), \end{aligned}$$

where $A_\psi := (\mathrm{spt}(\psi) \times \mathbb{R}^d) \cup (\mathbb{R}^d \times \mathrm{spt}(\psi))$. Then

$$\left|\int_{B_1} \psi \, d\mu(x) - \int_{B_1} \psi \, d\nu(y)\right| \leq \int_{A_\psi} |\psi(x) - \psi(y)| \, d\gamma(x, y) \leq \int_{A_\psi} [\psi]_{\mathrm{Lip}} |x - y| \, d\gamma(x, y).$$

Since $\mathrm{spt}(\psi)$ is a positive distance away from Γ, for any admissible γ we have $\gamma(A_\psi) \leq \mu(\mathrm{spt}(\psi)) + \nu(\mathrm{spt}(\psi)) < +\infty$. Thus, by Hölder's inequality,

$$\left|\int_{B_1} \psi \, d\mu - \int_{B_1} \psi \, d\nu\right| \leq [\psi]_{\mathrm{Lip}} \gamma(A_\psi)^{\frac{p-1}{p}} \left(\int_{A_\psi} |x - y|^p \, d\gamma(x, y)\right)^{\frac{1}{p}}.$$

Taking infimum over all $\gamma \in \mathrm{Adm}(\mu,\nu)$, we thus obtain

$$\left| \int_{B_1} \psi \, d\mu - \int_{B_1} \psi \, d\nu \right| \leq \left(\mu(\mathrm{spt}(\psi)) + \nu(\mathrm{spt}(\psi))\right)^{\frac{p-1}{p}} [\psi]_{\mathrm{Lip}} \mathrm{d}_{\mathbb{L}_p}(\mu,\nu). \qquad \square$$

Acknowledgments

We would like to thank the reviewers for a careful reading of the manuscript and for several important observations and suggestions. We also thank Alessio Figalli for helpful conversations regarding the validity or not of Lipschitz estimates for the Wasserstein distance with respect to other metrics.

References

[1] N. Alibaud, *Existence, uniqueness and regularity for nonlinear parabolic equations with non-local terms*, NoDEA Nonlinear Differential Equations Appl. **14** (2007), no. 3-4, 259–289, DOI 10.1007/s00030-007-5029-9. MR2364893

[2] N. Alibaud, S. Cifani, and E. R. Jakobsen, *Optimal continuous dependence estimates for fractional degenerate parabolic equations*, Arch. Ration. Mech. Anal. **213** (2014), no. 3, 705–762, DOI 10.1007/s00205-014-0737-x. MR3218830

[3] O. Alvarez and A. Tourin, *Viscosity solutions of nonlinear integro-differential equations*, Ann. Inst. H. Poincaré Anal. Non Linéaire **13** (1996), no. 3, 293–317, DOI 10.1016/S0294-1449(16)30106-8 (English, with English and French summaries). MR1395674

[4] L. Ambrosio and N. Gigli, *A user's guide to optimal transport*, Modelling and optimisation of flows on networks, Lecture Notes in Math., vol. 2062, Springer, Heidelberg, 2013, pp. 1–155, DOI 10.1007/978-3-642-32160-3_1. MR3050280

[5] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, 2nd ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008. MR2401600

[6] M. Arisawa, *A new definition of viscosity solutions for a class of second-order degenerate elliptic integro-differential equations*, Ann. Inst. H. Poincaré Anal. Non Linéaire **23** (2006), no. 5, 695–711, DOI 10.1016/j.anihpc.2005.09.002 (English, with English and French summaries). MR2259613

[7] G. Barles, R. Buckdahn, and E. Pardoux, *Backward stochastic differential equations and integral-partial differential equations*, Stochastics Stochastics Rep. **60** (1997), no. 1-2, 57–83, DOI 10.1080/17442509708834099. MR1436432

[8] G. Barles and C. Imbert, *Second-order elliptic integro-differential equations: Viscosity solutions' theory revisited*, Ann. Inst. H. Poincaré Anal. Non Linéaire **25** (2008), no. 3, 567–585, DOI 10.1016/j.anihpc.2007.02.007. MR2422079

[9] L. Caffarelli and L. Silvestre, *Regularity theory for fully nonlinear integro-differential equations*, Comm. Pure Appl. Math. **62** (2009), no. 5, 597–638, DOI 10.1002/cpa.20274. MR2494809

[10] E. Chasseigne and E. R. Jakobsen, *On nonlocal quasilinear equations and their local limits*, J. Differential Equations **262** (2017), no. 6, 3759–3804, DOI 10.1016/j.jde.2016.12.001. MR3592657

[11] M. G. Crandall and H. Ishii, *The maximum principle for semicontinuous functions*, Differential Integral Equations **3** (1990), no. 6, 1001–1014. MR1073054

[12] M. G. Crandall, H. Ishii, and P.-L. Lions, *User's guide to viscosity solutions of second order partial differential equations*, Bull. Amer. Math. Soc. (N.S.) **27** (1992), no. 1, 1–67, DOI 10.1090/S0273-0979-1992-00266-5. MR1118699

[13] A. Figalli and N. Gigli, *A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions*, J. Math. Pures Appl. (9) **94** (2010), no. 2, 107–130, DOI 10.1016/j.matpur.2009.11.005 (English, with English and French summaries). MR2665414

[14] N. Guillen and R. W. Schwab, *Min-max formulas for nonlocal elliptic operators*, arXiv:1606.08417 (2016).

[15] J. Gairing, M. Högele, T. Kosenkova, and A. Kulik, *Coupling distances between Lévy measures and applications to noise sensitivity of SDE*, Stoch. Dyn. **15** (2015), no. 2, 1550009, 25, DOI 10.1142/S0219493715500094. MR3332269

[16] E. R. Jakobsen and K. H. Karlsen, *A "maximum principle for semicontinuous functions" applicable to integro-partial differential equations*, NoDEA Nonlinear Differential Equations Appl. **13** (2006), no. 2, 137–165, DOI 10.1007/s00030-005-0031-6. MR2243708

[17] D. Kriventsov, *$C^{1,\alpha}$ interior regularity for nonlinear nonlocal elliptic equations with rough kernels*, Comm. Partial Differential Equations **38** (2013), no. 12, 2081–2106, DOI 10.1080/03605302.2013.831990. MR3169771

[18] C. Mou and A. Święch, *Uniqueness of viscosity solutions for a class of integro-differential equations*, NoDEA Nonlinear Differential Equations Appl. **22** (2015), no. 6, 1851–1882, DOI 10.1007/s00030-015-0347-9. MR3415025

[19] S. Awatif, *Équations d'Hamilton-Jacobi du premier ordre avec termes intégro-différentiels. I. Unicité des solutions de viscosité*, Comm. Partial Differential Equations **16** (1991), no. 6-7, 1057–1074, DOI 10.1080/03605309108820789 (French). MR1116853

[20] S. Awatif, *Équations d'Hamilton-Jacobi du premier ordre avec termes intégro-différentiels. II. Existence de solutions de viscosité*, Comm. Partial Differential Equations **16** (1991), no. 6-7, 1075–1093, DOI 10.1080/03605309108820790 (French). MR1116854

[21] R. W. Schwab and L. Silvestre, *Regularity for parabolic integro-differential equations with very irregular kernels*, Anal. PDE **9** (2016), no. 3, 727–772, DOI 10.2140/apde.2016.9.727. MR3518535

[22] H. M. Soner, *Optimal control with state-space constraint. II*, SIAM J. Control Optim. **24** (1986), no. 6, 1110–1122, DOI 10.1137/0324067. MR861089

[23] H. M. Soner, *Optimal control of jump-Markov processes and viscosity solutions*, Stochastic differential systems, stochastic control theory and applications (Minneapolis, 1986), IMA Vol. Math. Appl., vol. 10, Springer, New York, 1988, pp. 501–511, DOI 10.1007/978-1-4613-8762-6_29. MR934740

Department of Mathematics and Statistics, UMass, Amherst, Massachusetts 01003
Email address: nguillen@math.umass.edu

Department of Mathematics, UCLA, Los Angeles, California 90095
Email address: muchenchen@math.ucla.edu

School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332
Email address: swiech@math.gatech.edu

QUERIES

Q1: The complete article title is too long to serve as a running head in TRAN. As a default, "Coupling Lévy measures and comparison principles" was chosen as a suitable-length version of the title for the running head. Should you wish to use another truncated version of the title for your running, please submit one that consists of fewer than 50 characters. (Please include spaces in your character count.)

Q2: The acronym "PDE" was replaced by its definition, in accordance with TRAN style guidelines, as the acronym appeared only once in the text. Please ensure that the acronym was properly defined as "partial differential equation".

Q3: 2010 Mathematics Subject Classification numbers must be identified as "Primary" or "Secondary" ones. As a default, all of your 2010 MSC numbers have been designated as Primary ones. Please advise if any of these numbers should be designated as Secondary ones.

Q4: There was an unmatched right parentheses on the first line of the unnumbered display equation in the sentence beginning "Using the special form". Please ensure that the removal of this right parentheses conveys your intended meaning.

Q5: Please note that AMS style allows only one use of quotation marks or italics for the emphasis of a given term per paper. When quotation marks or italics were used multiple times for the same term in this paper, only the first mention was set in quotation marks or italics.

Q6: The abbreviation "cf." is taken to mean "compare to" in TRAN papers. Please review your use of this abbreviation in the sentence beginning "The proof is essentially", and change it to "see" if that is your intended meaning.