

ON OPTIMAL MARKOVIAN COUPLINGS OF LÉVY PROCESSES

WEI YANG KANG, TAU SHEAN LIM

ABSTRACT. We study the optimal Markovian coupling problem for two Π -valued Feller processes $\{X_t\}$ and $\{Y_t\}$, which seeks a coupling process $\{(X_t, Y_t)\}$ that minimizes the right derivative at $t = 0$ of the expected cost $\mathbb{E}^{(x,y)}[c(X_t, Y_t)]$, for all initial states $(x, y) \in \Pi^2$ and a given cost function c on Π . This problem was first formulated and solved by Chen [14] for drift–diffusion processes and later extended by Zhang [49] to Markov processes with bounded jumps. In this work, we resolve the case of Lévy processes under the quadratic cost $c(x, y) = \frac{1}{2}|x - y|^2$ by introducing a new formulation of the *Lévy optimal transport problem* between Lévy measures. We show that the resulting optimal coupling process $\{(X_t^*, Y_t^*)\}_{t \geq 0}$ satisfies a minimal growth property: for each $t \geq 0, x, y \in \mathbb{R}^d$, the expectation $\mathbb{E}^{(x,y)}|X_t^* - Y_t^*|^2$ is minimized among all Feller couplings. A key feature of our approach is the development of a dual problem, expressed as a variational principle over test functions of the generators. We prove *strong duality* for this formulation, thereby closing the optimality gap. As a byproduct, we obtain a Wasserstein-type metric on the space of Lévy generators and Lévy measures with finite second moment, and establish several of its fundamental properties.

Keywords: Optimal Markovian couplings, Lévy optimal transport, Lévy processes, strong duality, Wasserstein-type metrics on Lévy generators

CONTENTS

1. Introduction	2
List of symbols and notations	6
2. Framework, Research Problems, and Main Results	8
2.1. Notations and Feller semigroups/generators	9
2.2. Lévy processes, generators, triplets	12
2.3. Optimal transport between Markov flows	15
2.4. Optimal Markovian couplings	17
2.5. Comparison of transport derivative: pointwise, dual, and Markovian	21
2.6. Existing results for optimal Markovian coupling problem	22
2.7. Optimal Markovian coupling and optimality gap problem for Lévy generators	23
2.8. Wasserstein-type metric on the space $\mathcal{G}_2^\Lambda(\mathbb{R}^d)$	25
2.9. Discussion, methods of proof, and organization	26
3. Lévy Optimal Transport Problem	28
3.1. Classical optimal transport problem	29
3.2. Lévy optimal transport problem	30
3.3. Some preliminary results of the Lévy optimal transport cost	34

3.4. Construction of optimal Lévy couplings	36
3.5. The fundamental theorem of Lévy optimal transport	38
3.6. Lévy-Kantorovich duality	40
4. Optimal Markovian Couplings and Strong Duality for Drift, Diffusion and Jump Parts	42
4.1. Affine structure of θ_2, ω'_2	42
4.2. Reduction of global to local minimization	45
4.3. Optimal Markovian couplings and strong duality of drift-diffusion parts	46
4.4. Optimal Markovian coupling and strong duality of jump parts	49
5. Strong Duality and Optimal Markovian Couplings for Lévy Generators	50
5.1. Strong duality: proofs of Theorems 2.31 and 2.32	50
5.2. Optimal Markovian couplings: Proofs of Theorems 2.28(ii, iii)	51
5.3. Additivity of ω'_2 (proof of Lemma 5.1)	52
6. Wasserstein-type Metric on the Space $\mathcal{G}_2^\Lambda(\mathbb{R}^d)$ and $\Lambda_2(\mathbb{R}^d)$	56
6.1. Wasserstein generator metric	57
6.2. Lower semicontinuity, weak compactness and completeness	58
6.3. Lévy-Wasserstein metric and its topological properties	61
6.4. Separability of the Wasserstein generator metric	62
6.5. Pathwise estimate for coupled Lévy processes	63
Appendix A. Approximation of Admissible Pairs	64
A.1. Density of $\mathcal{F}_0 \subset \mathcal{F}$	65
A.2. Density of smooth admissible test functions (proof of Lemma 3.26)	67
A.3. Proof of Lemma 5.3	68
Appendix B. Proofs of Propositions 2.12, 2.21 and 2.27	69
B.1. Equivalent characterization of Feller couplings	69
B.2. Proof of Proposition 2.21	73
B.3. Proof of Proposition 2.27	74
Appendix C. Λ -weak topology in the space $\Lambda(\mathbb{R}^d)$ of Lévy measures	75
C.1. Portmanteau Lemma for Λ -weak convergence	77
C.2. Prokhorov-type Compactness Theorem	79
References	80

1. INTRODUCTION

Coupling is a fundamental probabilistic technique with applications across a wide range of stochastic models. The core idea is as follows: given two probability measures $(\Omega_1, \mathbb{P}_1)$ and $(\Omega_2, \mathbb{P}_2)$, one seeks a joint probability space $(\Omega, \mathbf{P})$, with $\Omega = \Omega_1 \times \Omega_2$, and a probability measure $\mathbf{P}$ on Ω such that the marginals agree with the original measures. That is, for all measurable sets $E_1 \subset \Omega_1$ and $E_2 \subset \Omega_2$,

$$\mathbf{P}(E_1 \times \Omega_2) = \mathbb{P}_1(E_1), \quad \mathbf{P}(\Omega_1 \times E_2) = \mathbb{P}_2(E_2).$$

Such a construction allows one to study the relationship between the two distributions, random variables, or processes by analyzing them on the same probability space. The technique is widely used in areas such as the analysis of ergodicity, convergence rates, and spectral gaps for Markov chains and MCMC [19, 25, 33, 40]; optimal transport theory [2, 46]; Poisson approximation via the Stein–Chen method [4, 13]; interacting particle systems [38, 45, 11, 12]; probabilistic analysis of PDEs [32, 44, 48]; and geometric analysis through couplings of diffusions [27, 47].

In many applications, a central objective is to construct a coupling that satisfies a specific *optimality* criterion. One prominent example is the *maximal coupling* of two probability measures on the sample space, which maximizes the probability that the two coupled random variables are equal. More precisely, under the setting above, $\mathbf{P}$ is a maximal coupling of $\mathbb{P}_1$ and $\mathbb{P}_2$ if

$$\mathbf{P}(X = Y) = \sup_{\mathbf{Q}} \mathbf{Q}(X = Y),$$

where the supremum is taken over all couplings $\mathbf{Q}$ of $\mathbb{P}_1$ and $\mathbb{P}_2$. Equivalently, $\mathbf{P}(X \neq Y) = d_{\text{TV}}(\mathbb{P}_1, \mathbb{P}_2)$, where d_{TV} denotes the total variation distance. Such couplings are widely used in the study of mixing times and convergence rates of Markov chains/MCMC, in the construction of perfect sampling algorithms, and in Poisson approximation via the Stein–Chen method. It is a standard fact that maximal couplings always exist.

A broader class of optimal couplings is given by the viewpoint of *optimal transport theory*. Let μ and ν be probability measures defined on Polish spaces Π and Γ , respectively, equipped with a *cost function* $c : \Pi \times \Gamma \rightarrow \mathbb{R}$. A coupling γ_* , which is a probability measure on $\Pi \times \Gamma$ with marginals μ and ν , is called *c-optimal* if

$$\int_{\Pi \times \Gamma} c(x, y) d\gamma_*(x, y) = \inf_{\gamma} \int_{\Pi \times \Gamma} c(x, y) d\gamma(x, y),$$

where the infimum is taken over all couplings γ of μ and ν . Such *c-optimal* couplings naturally arise when one seeks to match two distributions while minimizing an underlying transportation cost, and they play a central role in probability theory, statistics, and PDEs. Existence of a *c-optimal* coupling is guaranteed under mild conditions, for example when c is lower semicontinuous and bounded from below on $\Pi \times \Gamma$ (see, e.g., [46]). The maximal coupling described above is in fact a special case of this framework, obtained by choosing the cost function $c(x, y) = \mathbf{1}_{\{x \neq y\}}$.

In the context of stochastic processes, one can view a Π -valued process as a probability measure on the corresponding path space, and a coupling can then be defined analogously as a probability measure on the product of two such path spaces with given marginals. For instance, given two processes $\{X_t\}_{t \geq 0}$ and $\{Y_t\}_{t \geq 0}$, one can define their *maximal coupling*—which always exists in this abstract sense. However, such a coupling is typically of limited practical value, as it does not in general preserve the temporal or causal structure inherent to the stochastic evolution of the individual processes. For this reason, when working *at the process level*, couplings are often required to satisfy additional structural constraints that

reflect the dynamic nature of the problem. Common examples include *faithfulness* [41], *co-adaptedness* [18, 8], and *immersion* [28], each imposing progressively stronger compatibility between the joint dynamics and the original processes.

Let us now turn to the notion of *Markovian coupling*, which is the regime of interest in the present work. Broadly speaking, a coupling $\{Z_t = (X'_t, Y'_t)\}_{t \geq 0}$ of two Markov processes $\{X_t\}_{t \geq 0}, \{Y_t\}_{t \geq 0}$ is said to be Markovian if the joint process $\{Z_t\}_t$ itself forms a Markov process on the product state space, with respect to its natural filtration. In this setting, the coupling is determined by a transition kernel or generator on the product space whose marginals coincide with those of the individual processes. This ensures that the joint evolution preserves the memoryless property, which is especially useful in the study of interacting particle systems, where couplings can then be analyzed within the framework of Markov semigroups and generators.

An interesting and widely studied example of coupling processes is the *coalescing coupling*, in which two processes evolve jointly until they first meet, after which they move together identically. In the setting of Markov chains on discrete state spaces, the meeting time often serves as a natural stopping time whose distribution directly informs quantitative mixing properties. This naturally motivates the search for *optimal coalescing couplings*—those that bring the processes together as quickly as possible, for example by minimising the distribution or a moment of the coalescence time. In many discrete settings, such optimal constructions coincide with *maximal couplings* in the sense of total variation distance, yielding the sharpest coupling-based bounds on convergence to equilibrium.

Optimal coalescing couplings have also been studied in the context of *Markovian couplings* on continuous state spaces, often referred to as *Markovian maximal couplings* [9]. A canonical example is the *reflection coupling* of Brownian motions [37], where the processes evolve symmetrically with respect to the hyperplane bisecting their positions until they meet, after which they move together. This construction is both Markovian and optimal, achieving the minimal possible coalescence time among all admissible couplings. We remark also coalescing couplings have also been investigated in the Lévy setting [7, 43, 34].

Another important class of Markovian couplings is the *synchronous coupling*, in which two Markov processes on $\mathbb{R}^d$ —typically described by stochastic differential equations driven by a common source of noise (Brownian or Lévy)—evolve under exactly the same noise realisation. For example, if the two processes $\{X_t\}_t$ and $\{Y_t\}_t$ on $\mathbb{R}^d$ admit the SDE representations

$$dX_t = b(X_t) dt + \sigma(X_t) dW_t, \quad dY_t = b(Y_t) dt + \sigma(Y_t) d\tilde{W}_t,$$

where $\{W_t\}$ and $\{\tilde{W}_t\}$ are d -dimensional Lévy processes with the same law, their synchronous coupling is obtained by setting $W = \tilde{W}$. We note here that reflection couplings of Brownian motions is a special case of synchronous couplings.

In the analysis of interacting particle systems, one is often interested in comparing two such systems, described by Markov processes $\{X_t\}$ and $\{\bar{X}_t\}$ on some state space Π , in terms of their error measured by a cost function—for instance, the expected transport cost

$$\mathbb{E}[c(X_t, \bar{X}_t)], \quad t \geq 0 \tag{1.1}$$

under an appropriate coupling. In the case where the systems are described via SDEs (e.g., diffusion processes), synchronous couplings are particularly useful. By aligning the driving noise across the two systems, synchronous coupling links probabilistic arguments with tools from stochastic calculus and martingale theory, such as Itô's formula and maximal inequalities. This makes it a natural bridge between coupling methods and analytic techniques, particularly in the study of stability, moment bounds, and propagation of chaos in mean-field models.

For general interacting particle systems where an SDE representation is not available, synchronous couplings are no longer directly applicable. Nevertheless, the underlying idea—constructing a coupling process $\{(X_t, \bar{X}_t)\}_{t \geq 0}$ to bound the expected transport cost (1.1)—remains central to the analysis. Within the Markovian coupling framework, as Chen and Li pointed out in [16], one can exploit the semigroup structure together with integral inequalities (for instance, Grönwall's lemma) to obtain global-in-time estimates from bounds on the derivative of the expected cost at time $t = 0$, that is, by analyzing the action of the infinitesimal generator of the coupling on the cost function c .

This motivates the following optimality criterion, introduced by Chen [14]: given two Markov generators $\mathcal{A}$ and $\mathcal{B}$ of two Markov processes on some state space Π , and a cost function $c : \Pi^2 \rightarrow [0, \infty)$, a coupling generator $\mathcal{J}_*$ of $\mathcal{A}$ and $\mathcal{B}$ is said to be *c-optimal* if

$$(\mathcal{J}_*c)(x, y) = \inf_{\mathcal{J}}(\mathcal{J}c)(x, y), \quad \text{for all } (x, y) \in \Pi^2,$$

where the infimum is taken overall all coupling generators of $\mathcal{A}, \mathcal{B}$. A precise formulation of this optimality condition will be presented in the coming section.

The central question is the following: given two Markov generators $\mathcal{A}$ and $\mathcal{B}$ and a cost function c , does there exist a *c-optimal* coupling generator $\mathcal{J}_*$ of $\mathcal{A}$ and $\mathcal{B}$? We refer to this as the *c-optimal Markovian coupling problem*. This problem was first introduced by Chen and Li [16], and was subsequently given a precise formulation by Chen [14, 15] for *bounded Markov jump operators*, who also established the existence of minimizers in the case of drift–diffusion generators under suitable assumptions on the cost function c . Later, Zhang [49] solved the problem for bounded Markov jump operators. For further background, see Section 2.6. The problem has also been studied in the Lévy setting for specific cost functions (e.g., [34, 29]). To the best of our knowledge, however, the *c-optimal* Markovian coupling problem remains open for general unbounded Markov generators.

In the present work, our goal is twofold. First, we rigorously formulate the optimal Markovian coupling problem for general Feller generators. Second, we provide a complete solution to this problem in the case of two Lévy generators on $\mathbb{R}^d$, with respect to the squared cost $c(x, y) = \frac{1}{2}|x - y|^2$. It is well known that Lévy generators can be decomposed into drift–diffusion and jump components, the latter of which may be unbounded. As mentioned earlier, the general drift–diffusion case was addressed in [16], while the bounded jump case was solved in [49]. The main contribution of the present work is therefore the treatment of the unbounded jump component.

One interesting by-product of this research is the formulation and analysis of the *Lévy optimal transport problem*. This is a variation of the classical optimal transport problem,

which seeks an optimal coupling, but now between two *Lévy measures* rather than probability measures. This is not the first time such a problem has appeared; it was already considered in the work of Kolokoltsov [30] (see the paragraph after Corollary 2.34). Nonetheless, the present work provides the first systematic study of this problem. In particular, the link between the Lévy optimal transport problem and the optimal Markovian coupling problem for Lévy jump processes is established here for the first time. This problem will be studied in detail in Section 3, where we shall see its close connection with the optimal Markovian coupling problem for jump generators.

Another theme of this work is the study of the *dual formulation* associated with the optimal Markovian coupling problem and Lévy optimal transport problem. We establish that strong duality holds, and this plays a crucial role in the analysis. In particular, the dual formulation provides a verification principle: once a candidate coupling is constructed, its optimality can be confirmed by showing that its cost coincides with the dual value. The question of whether the dual value equals the primal value is referred to as the *optimality gap problem*, and strong duality ensures that this gap vanishes. This follows the standard approach in optimization theory, where strong duality guarantees that equality of primal and dual values certifies optimality.

Another by-product of this research is the introduction of a *Wasserstein-type* metric on the space of Lévy generators and Lévy measures. This metric arises naturally from the optimal Markovian problem between Lévy generators and Lévy optimal transport problem between Lévy measures, and provides a quantitative way to compare different Lévy-type dynamics. This metric was first introduced in [36], and to our knowledge, this is the first time such a metric has been systematically studied, and it has several potential applications. We present a detailed construction and analysis of this metric in Section 6.

In the coming section, we formulate the two main research problems of this work: the *optimal Markovian coupling problem* and the *optimality gap problem*, both defined with respect to a cost function c in a general abstract framework. Our study focuses on the special case of Lévy generators on $\mathbb{R}^d$ with the squared cost $c(x, y) = \frac{1}{2}|x - y|^2$. The main results, stated as Theorems 2.28, 2.31, 2.32, 2.33, 2.35 and Corollary 2.34, will be presented subsequently. The detailed plan and organization of the remainder of the paper will be outlined at the end of Section 2.

List of symbols and notations.

$\oplus$	: Direct sum of functions, e.g., $f \oplus g = f(x) + g(y)$, or spaces, e.g., $L^1(\mu) \oplus L^1(\nu)$
$\otimes$	: Tensor product (of functions, measures, operators, semigroups)
$\langle \cdot, \cdot \rangle$	: Natural pairing over $\mathcal{M}(\Pi) \times C_b(\Pi)$
$\xrightarrow{\Lambda}$	: Λ -weak convergence of Lévy measures
$(\kappa, \alpha, \mu), (\zeta, \beta, \nu)$	: Lévy triplet
(η, σ, γ)	: Coupled Lévy triplet
$(\eta_*, \sigma_*, \gamma_*)$	: 2-optimal coupled Lévy triplet
α, β	: Diffusion matrices

$\mathcal{A}, \mathcal{B}$	: Feller generators on Π , or Lévy generators on $\mathbb{R}^d$
$\mathcal{A}^\nabla, \mathcal{B}^\nabla$	: Drift parts of $\mathcal{A}, \mathcal{B}$
$\mathcal{A}^\Delta, \mathcal{B}^\Delta$	: Diffusion parts of $\mathcal{A}, \mathcal{B}$
$\mathcal{A}^J, \mathcal{B}^J$	: Jump parts of $\mathcal{A}, \mathcal{B}$
$\mathcal{B}(\Pi)$	: Space of Borel measurable functions on Π
$\mathcal{B}_b(\Pi)$	: Space of bounded Borel measurable functions on Π
c	: Cost function over some state space Π
c_2	: Squared cost on $\mathbb{R}^d$, $c_2(x, y) = \frac{1}{2} x - y ^2$
$\mathcal{C}_c$	: c -optimal transport cost
$\mathcal{C}_2$	: 2-optimal transport cost
$\mathcal{C}_2^\Lambda$	: 2-optimal Lévy transport cost (between Lévy measures)
$C(\mathcal{A}), C(\mathcal{B})$	: Continuity domains of generators $\mathcal{A}, \mathcal{B}$
$C(\Pi)$	: Space of continuous functions on Π
$C_0(\Pi)$	: Space of continuous functions on Π vanishing at infinity
$C_b(\Pi)$	: Space of bounded continuous functions
$C_c(\Pi)$	: Space of continuous functions $f : \Pi \rightarrow \mathbb{R}$ with compact support
$C_2(\mathbb{R}^d)$	: Space of continuous functions $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $ \varphi(x) \leq C(1 + x ^2)$
$C_0^2(\mathbb{R}^d)$	: Space of bounded twice differentiable functions on $\mathbb{R}^d$ vanishing at infinity
$C_b^2(\mathbb{R}^d)$	: Space of bounded twice differentiable functions on $\mathbb{R}^d$
$C_2^2(\mathbb{R}^d)$	: Space of all $\varphi \in C_2(\mathbb{R}^d)$ that is continuously twice differentiable
δ_x, δ_y	: Dirac delta measures at $x, y \in \Pi$
D^-	: Right hand lower Dini derivative operator
$D(\mathcal{A})$	: Domain of the generator $\mathcal{A}$
$\bar{D}(\mathcal{A})$	: Extended domain of the generator $\mathcal{A}$
$\mathcal{D}(x_0, y_0)$	: Set of admissible pairs for the definition of dual transport derivative
$\{e^{t\mathcal{A}}\}_{t \geq 0}, \{e^{t\mathcal{B}}\}_{t \geq 0}$	: Feller semigroups generated by $\mathcal{A}, \mathcal{B}$
η	: $\kappa \oplus \zeta$, coupled drift vector
γ	: Coupling of probability/Lévy measures μ, ν
γ_*	: Lévy 2-optimal coupling of Lévy measures μ, ν
$\Gamma(\mathcal{A}, \mathcal{B})$	: Set of all Feller coupling generators of $\mathcal{A}$ and $\mathcal{B}$
$\Gamma^\Lambda(\mathcal{A}, \mathcal{B})$	: Set of all Lévy coupling generators of $\mathcal{A}$ and $\mathcal{B}$
$\Gamma(\mu, \nu)$	: Set of all (classical) couplings of bounded measures μ and ν with equal mass
$\Gamma^\Lambda(\mu, \nu)$	: Set of all (Lévy) couplings of Lévy measures μ and ν
$\mathcal{G}(\Pi)$	: Space of probability generators on Π
$\mathcal{G}^\Lambda(\mathbb{R}^d)$	: Space of Lévy generators on $\mathbb{R}^d$
$\mathcal{G}_2^\Lambda(\mathbb{R}^d)$	: Space of Lévy generators with finite second moments

$\mathcal{J}$	: Coupling generator of $\mathcal{A}, \mathcal{B}$
$\mathcal{J}^-$	: Lower generator of the generator $\mathcal{J}$
$\mathcal{J}_*$	: Optimal coupling generator of $\mathcal{A}, \mathcal{B}$
κ, ζ	: Drift vectors
Λ	: Lévy, particularly any notions related to Lévy settings
$\Lambda(\mathbb{R}^d)$	: Class of Lévy measures on $\mathbb{R}^d$
$\Lambda_2(\mathbb{R}^d)$	: Class of Lévy measures on $\mathbb{R}^d$ with finite second moment
$\Lambda_{\text{glob}}(\kappa, \alpha, \mu)$	: Lévy generator associated with the triplet (κ, α, μ) expressed in global form
μ, ν	: Probability measures on Π , or Lévy measures on $\mathbb{R}^d$
$m_{\mathcal{A}}, m_{\mathcal{B}}$	: Mean vectors of Lévy generator $\mathcal{A}, \mathcal{B}$
$M(\mathbb{R}^d)$	: Space of squared $d \times d$ real matrices
$\mathcal{M}(\Pi)$	: Space of bounded (Borel) measures on Π
$\omega_c^\pm$	: Pointwise upper/lower transport derivative
ω_c^*	: Dual transport derivative
Π	: State (Polish) space
$\mathcal{P}(\Pi)$	: Space of probability (Borel) measures on Π
$\mathcal{P}_2(\mathbb{R}^d)$	: Space of probability measures μ with finite second moments
$Q_{\mathcal{A}}, Q_{\mathcal{B}}$	: covariance matrices of Lévy generators $\mathcal{A}, \mathcal{B}$
$\mathcal{S}_{\geq 0}(\mathbb{R}^d)$	: Space of nonnegative definite $(d \times d)$ -matrices with real entries
σ	: Coupled diffusion matrix
$\{T_t\}_{t \geq 0}$	: Markov or Feller semigroup
θ_c	: Markovian transport derivative
$\theta_2, \omega_2^\pm, \omega_2^*$	: Transport derivatives w.r.t. the squared cost c_2
ω_2'	: Restricted dual transport derivative
$\mathcal{W}_2$	: Wasserstein 2-metric between probability measures
$\mathcal{W}_{\mathcal{G}}$	: Wasserstein generator metric
$\mathcal{W}_{\Lambda}$	: $(\mathcal{C}_2^{\Lambda})^{1/2}$, Lévy-Wasserstein metric
$\mathcal{W}_{\mathcal{S}}$	: Bures-Wasserstein distance
$\{X_t\}_{t \geq 0}, \{Y_t\}_{t \geq 0}$	: Markov, Feller or Lévy processes

2. FRAMEWORK, RESEARCH PROBLEMS, AND MAIN RESULTS

In this section, we formulate the optimal Markovian coupling problem and the associated optimality gap problem in the abstract setting of Feller generators, and present the background required for their analysis. We begin with notations and the general framework of the study, followed by a brief review of Feller and Lévy processes and their generators. Key notions, particularly transport derivatives, are introduced in Section 2.3. The two research problems are then stated in Sections 2.4 and 2.5, and related results from the literature are

discussed in Section 2.6. Our main contributions, concerning the case of Lévy processes, are presented in Sections 2.7 and 2.8. We conclude with a discussion of the results, proof strategies, and the organization of the remainder of the article.

2.1. Notations and Feller semigroups/generators. Throughout this paper, we denote Π a state space, which is a *Polish* (completely metrizable separable) space. Let us introduce also the following notations for spaces of functions, measures, generators, pairings, and actions:

- $\mathcal{B}(\Pi)$: the space of Borel functions $f : \Pi \rightarrow \mathbb{R}$.
- $\mathcal{B}_b(\Pi)$: the space of bound Borel functions $f : \Pi \rightarrow \mathbb{R}$.
- $C(\Pi)$: the space of continuous functions $f : \Pi \rightarrow \mathbb{R}$.
- $C_b(\Pi)$: the space of bounded continuous functions $f : \Pi \rightarrow \mathbb{R}$, with the supremum norm $\|f\| = \sup_{x \in \Pi} |f(x)|$.
- $C_0(\Pi)$: the space of continuous functions $f : \Pi \rightarrow \mathbb{R}$ vanishing at infinity.
- $C_c(\Pi)$: the space of continuous functions $f : \Pi \rightarrow \mathbb{R}$ with compact support.
- $\mathcal{M}(\Pi)$: the space of bounded (Borel) measures on Π .
- $\mathcal{P}(\Pi)$: the space of probability (Borel) measures on Π .
- $\mathcal{G}(\Pi)$: the space of all Feller generators on Π , see the following subsection.
- $\langle \cdot, \cdot \rangle : \mathcal{M}(\Pi) \times C_b(\Pi) \rightarrow \mathbb{R}$: the natural pairing between bounded measures and bounded continuous functions

$$\langle \mu, f \rangle = \int_{\Pi} f d\mu.$$

This notion of pairing may potentially extend to unbounded functions f or unbounded measures μ (e.g., Lévy measures), so long as the integral makes sense.

- *Left action:* Given an operator $T : C_b(\Pi) \rightarrow C_b(\Pi)$ (for example, a Markov kernel, a Feller semigroup, or a generator) and a measure $\mu \in \mathcal{M}(\Pi)$, we define the *left action* of T on μ , denoted μT , by

$$\langle \mu T, f \rangle := \langle \mu, T f \rangle, \quad \text{for all } f \in C_b(\Pi).$$

In particular, if T is a Markov kernel or a Feller semigroup, then μT is again a measure.

In the present work, although the research problems are formulated in the abstract framework introduced above, we will mainly concentrate on the case $\Pi = \mathbb{R}^d$. Accordingly, the spaces of continuous functions, measures, and generators will be understood with $\Pi = \mathbb{R}^d$. For later use, we also introduce the following spaces:

- $C_2(\mathbb{R}^d)$: the space of continuous functions satisfying the quadratic growth bound for some constant $C > 0$:

$$|\varphi(x)| \leq C(1 + |x|^2), \quad x \in \mathbb{R}^d, \quad (2.1)$$

- $C^2(\mathbb{R}^d)$: the space of twice continuously differentiable functions;
- $C_b^2(\mathbb{R}^d)$: the space of bounded twice continuously differentiable functions;
- $C_2^2(\mathbb{R}^d) = C_2(\mathbb{R}^d) \cap C^2(\mathbb{R}^d)$: the space of twice continuously differentiable functions φ satisfying the quadratic growth bound (2.1).

- $\mathcal{P}_2(\mathbb{R}^d)$: the space of probability measures on $\mathbb{R}^d$ with finite second moment, i.e.

$$\int_{\mathbb{R}^d} |x|^2 d\mu(x) < \infty.$$

2.1.1. *Markov/Feller semigroups, generators, and processes.* We begin with a brief review of Markov and Feller semigroups, their generators, and the associated processes, as these form the central framework of the present work. For a comprehensive treatment, we refer the reader to [22, 31, 35].

A *Markov semigroup* is a family of bounded linear operators $\{T_t\}_{t \geq 0}$ on the space $\mathcal{B}_b(\Pi)$ of bounded Borel measurable functions, satisfying:

- Identity: $T_0 = I$.
- Semigroup property: for all $\varphi \in \mathcal{B}_b(\Pi)$ and $s, t \geq 0$,

$$T_{t+s}\varphi = T_t(T_s\varphi).$$

- Positivity: if $\varphi \geq 0$, then $T_t\varphi \geq 0$ for all $t \geq 0$.
- Conservativeness: $T_t 1 = 1$ for all $t \geq 0$.

A canonical example is the semigroup induced by a *Markov process* $\{X_t\}_{t \geq 0}$, given by

$$T_t\varphi(x) = \mathbb{E}^x[\varphi(X_t)], \quad \varphi \in \mathcal{B}_b(\Pi), \quad x \in \Pi. \quad (2.2)$$

In this case, $\{T_t\}_{t \geq 0}$ admits a *transition kernel* $\{\kappa_t(x, dy)\}_{t \geq 0, x \in \Pi}$ such that

$$T_t\varphi(x) = \int_{\Pi} \varphi(y) \kappa_t(x, dy) = \langle \kappa_t(x), \varphi \rangle.$$

This kernel representation also permits extending $T_t\varphi$ to certain unbounded functions whenever the integral is well defined. In terms of the left action notation introduced earlier, we may write

$$\kappa_t(x) = \delta_x T_t.$$

A *Feller semigroup*¹ is a Markov semigroup with the additional *Feller property*:

- $T_t(C_0(\Pi)) \subset C_0(\Pi)$ for all $t \geq 0$;
- $\lim_{t \searrow 0} T_t\varphi = \varphi$ in $C_0(\Pi)$ for all $\varphi \in C_0(\Pi)$.

The *infinitesimal generator* $\mathcal{A}$ of a Feller semigroup and its domain $D(\mathcal{A})$ is defined by

$$D(\mathcal{A}) := \left\{ \varphi \in C_0(\Pi) : \lim_{t \searrow 0} \frac{T_t\varphi - \varphi}{t} \text{ exists in } C_0(\Pi) \right\}, \quad \mathcal{A}\varphi := \lim_{t \searrow 0} \frac{T_t\varphi - \varphi}{t}. \quad (2.3)$$

The domain $D(\mathcal{A})$ is a dense subspace of $C_0(\Pi)$, and $(\mathcal{A}, D(\mathcal{A}))$ is a (typically unbounded) closed operator on $C_0(\Pi)$.

To highlight the connection between the generator $\mathcal{A}$ and its semigroup $\{T_t\}_{t \geq 0}$, it is customary to use the formal notation

$$T_t = e^{t\mathcal{A}}, \quad t \geq 0,$$

¹In [35], it is referred to as a *probability semigroup*.

which expresses the fact that the semigroup is generated by $\mathcal{A}$ in the sense of (2.3). As introduced earlier, the collection of all such generators on Π is denoted by $\mathcal{G}(\Pi)$ and referred to as the set of *Feller generators*. Every Feller semigroup arises from a Markov process $\{X_t\}_{t \geq 0}$ via (2.2), called a *Feller process*. Classically, there is a one-to-one correspondence between Feller processes, their semigroups, and their generators, providing a powerful analytic framework for their study; see, for example, [22, 35].

2.1.2. Extended domains of generators. For a Feller generator $\mathcal{A} \in \mathcal{G}(\Pi)$, the domain $D(\mathcal{A})$ consists of all $\varphi \in C_0(\Pi)$ for which the limit in (2.3) exists in the uniform topology. This definition, though natural in semigroup theory, is often too restrictive: for example, the quadratic cost $c(x, y) = \frac{1}{2}|x - y|^2$ does not lie in $C_0(\mathbb{R}^{2d})$. Yet $e^{t\mathcal{A}}\varphi$ can still be defined for such unbounded φ whenever the integral against the transition kernel is well-defined. To handle these cases, one introduces the *extended domain* of the generator.

Definition 2.1 (Continuity domain and extended domain of a generator). Let $\mathcal{A} \in \mathcal{G}(\Pi)$. The *continuity domain* $C(\mathcal{A})$ of $\mathcal{A}$ is the vector space of all $\varphi \in C(\Pi)$ such that

$$e^{t\mathcal{A}}\varphi \in C(\Pi) \quad \forall t \geq 0, \quad \text{and} \quad e^{t\mathcal{A}}\varphi \rightarrow e^{s\mathcal{A}}\varphi \text{ locally uniformly on } \Pi \text{ as } t \rightarrow s.$$

The *extended domain* $\bar{D}(\mathcal{A})$ of $\mathcal{A}$ consists of all $\varphi \in C(\mathcal{A})$ such that the following holds: for each $t \geq 0$, there exists $g_t \in C(\mathcal{A})$ such that

$$\frac{e^{(t+h)\mathcal{A}}\varphi - e^{t\mathcal{A}}\varphi}{h} \rightarrow g_t \quad \text{locally uniformly on } \Pi \text{ as } h \searrow 0.$$

In this case, we continue to write $\mathcal{A}\varphi := g_0$ for the limit in (2.3), now understood in the local uniform sense.

Clearly, we have $C_b(\Pi) \subset C(\mathcal{A})$ and $D(\mathcal{A}) \subset \bar{D}(\mathcal{A})$. Moreover, for any $\varphi \in \bar{D}(\mathcal{A})$ and $t \geq 0$, one has $e^{t\mathcal{A}}\varphi \in \bar{D}(\mathcal{A})$, and it holds

$$\mathcal{A}e^{t\mathcal{A}}\varphi = \lim_{h \searrow 0} \frac{e^{(t+h)\mathcal{A}}\varphi - e^{t\mathcal{A}}\varphi}{h} = e^{t\mathcal{A}}\mathcal{A}\varphi.$$

By the fundamental theorem of calculus, for any $\varphi \in \bar{D}(\mathcal{A})$ it holds

$$e^{t\mathcal{A}}\varphi = \varphi + \int_0^t e^{s\mathcal{A}}\mathcal{A}\varphi ds. \quad (2.4)$$

In the present work, our focus lies on the *lower right-hand Dini derivative* of trajectories of the form $t \mapsto e^{t\mathcal{A}}\varphi$, where $\varphi \in \mathcal{B}(\Pi)$ is a measurable test function without any regularity assumptions. This leads naturally to the following notion.

Definition 2.2 (Lower generator). Let $\mathcal{A} \in \mathcal{G}(\Pi)$. The *lower generator* $\mathcal{A}^-$ is the operator acting on $\mathcal{B}(\Pi)$ by

$$\mathcal{A}^-\varphi(x) := \liminf_{h \searrow 0} \frac{e^{h\mathcal{A}}\varphi(x) - \varphi(x)}{h}, \quad \varphi \in \mathcal{B}(\Pi).$$

Clearly, if $\varphi \in \bar{D}(\mathcal{A})$, then $\mathcal{A}^-\varphi = \mathcal{A}\varphi$. One may also define the *upper generator* $\mathcal{A}^+$ analogously, by replacing $\liminf$ with $\limsup$.

2.2. Lévy processes, generators, triplets. Our primary interest lies in Lévy processes. We now provide a brief overview of their definition and generators; for a comprehensive treatment, see [6, 42]

2.2.1. Basics of Lévy processes. A Lévy process $\{X_t\}_{t \geq 0}$ on $\mathbb{R}^d$ is a Feller process with *stationary* and *independent increments*. By the *Lévy–Khintchine formula*, the associated Feller generator $\mathcal{A}$, defined for test functions $\varphi \in C_c^2(\mathbb{R}^d)$, is given by

$$\mathcal{A}\varphi = \kappa \cdot \nabla \varphi + \frac{1}{2} \text{tr}(\alpha D^2 \varphi) + \int_{\mathbb{R}^d \setminus \{0\}} [\varphi(x+x') - \varphi(x) - \chi_{B_1(0)}(x') x' \cdot \nabla \varphi] d\mu(x'). \quad (2.5)$$

where the triple (κ, α, μ) is called the *Lévy triplet* and consists of:

- $\kappa \in \mathbb{R}^d$, a vector representing the *drift* component;
- $\alpha \in \mathcal{S}_{\geq 0}(\mathbb{R}^d)$, a nonnegative definite matrix representing the *diffusion* component;
- μ , a *Lévy measure*, i.e., a Borel measure on $\mathbb{R}^d$ satisfying

$$\mu(\{0\}) = 0, \quad \int_{\mathbb{R}^d \setminus \{0\}} \min\{1, |x|^2\} d\mu(x) < \infty. \quad (2.6)$$

The three terms in (2.5) are referred to as the *drift*, *diffusion*, and *jump* parts of the generator $\mathcal{A}$, respectively.

2.2.2. Lévy generators with finite second moments. Let $\Lambda(\mathbb{R}^d)$ denote the family of all Lévy measures on $\mathbb{R}^d$, that is, measures μ satisfying Condition (2.6). As before, $\mathcal{G}(\mathbb{R}^d)$ denotes the space of all Feller generators on $\mathbb{R}^d$. We write $\mathcal{G}^\Lambda(\mathbb{R}^d) \subset \mathcal{G}(\mathbb{R}^d)$ for the subspace of all *Lévy generators* on $\mathbb{R}^d$, i.e., Feller generators of the form (2.5). Specifically, *the Greek letter Λ is reserved for any notions that is related to “Lévy.”*

For $p \in [1, \infty)$, denote by $\Lambda_p(\mathbb{R}^d) \subset \Lambda(\mathbb{R}^d)$ the class of Lévy measures with finite p -th moment outside the unit ball:

$$\int_{B_1(0)^c} |x|^p d\mu(x) < \infty. \quad (2.7)$$

A Lévy process $\{X_t\}_{t \geq 0}$ has finite p -th moments for $p \geq 1$, i.e., $\mathbb{E}[|X_t|^p] < \infty$ for all $t \geq 0$, if and only if its generator (given by (2.5)) is associated with a Lévy measure $\mu \in \Lambda_p(\mathbb{R}^d)$, see [42]. We note also the inclusion holds: if $p \leq q$, then

$$\Lambda_q(\mathbb{R}^d) \subset \Lambda_p(\mathbb{R}^d).$$

In this work, we focus on generators whose associated Lévy processes that admit finite second moments, which corresponds to Lévy measures μ with finite second moment. Combining the conditions (2.6) and (2.7), $\mu \in \Lambda_2(\mathbb{R}^d)$ if and only if μ admits (globally) finite second moment:

$$\int_{\mathbb{R}^d} |x|^2 d\mu(x) < \infty. \quad (2.8)$$

Similarly, let $\mathcal{G}_2^\Lambda(\mathbb{R}^d) \subset \mathcal{G}^\Lambda(\mathbb{R}^d)$ denote the class of all Lévy generators $\mathcal{A}$ taking the form (2.5) with $\mu \in \Lambda_2(\mathbb{R}^d)$. As discussed in the preceding paragraph, $\mathcal{A} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$ if and only if the semigroup admits a finite second moment, that is, $e^{t\mathcal{A}}[|x|^2](0) < \infty$ for all $t \geq 0$.

Recall the spaces $C_2(\mathbb{R}^d)$, $C_2^2(\mathbb{R}^d)$ introduced in Section 2.1, together with the continuity domain $C(\mathcal{A})$ and the extended domain $\bar{D}(\mathcal{A})$ of a generator $\mathcal{A}$ from Definition 2.1. If $\mathcal{A} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$, then

$$C_2(\mathbb{R}^d) \subset C(\mathcal{A}), \quad C_2^2(\mathbb{R}^d) \subset \bar{D}(\mathcal{A}).$$

In particular, the extended domain $\bar{D}(\mathcal{A})$ of a Lévy generator with finite second moments contains all quadratic functions. This fact is crucial, since the quadratic cost function used in our framework must belong to the domain of the generator.

2.2.3. Global cutoff functions. We briefly remark on the role of *cutoff functions* in the definition of the jump operator in (2.5). The standard choice is the characteristic function $\chi_{B_1(0)}$ of the unit ball, which yields the compensated expression

$$\varphi(x + x') - \varphi(x) - \chi_{B_1(0)}(x') x' \cdot \nabla \varphi(x),$$

designed to regularize the potential singularity near the origin when the Lévy measure μ is infinite there. More generally, as noted in Sato [42], any compactly supported measurable function $\chi(x')$ satisfying $\chi(x') = 1 + o(1)$ as $x' \rightarrow 0$ may be used in place of $\chi_{B_1(0)}$. Different choices of cutoff result in equivalent generators on $C_c^2(\mathbb{R}^d)$, with corresponding adjustments to the drift vector κ . Thus, the cutoff is a matter of representation rather than intrinsic structure.

When the Lévy measure μ admits a finite p -th moment (2.7) for some $p \geq 1$, the first moment outside the unit ball (as a vector in $\mathbb{R}^d$) is also finite (componentwise):

$$\kappa' := \int_{B_1(0)^c} x' d\mu(x') < \infty.$$

In this case, it is natural to adopt the *global form* of the jump operator by taking $\chi(x') \equiv 1$, which leads to

$$\int_{\mathbb{R}^d} [\varphi(x + x') - \varphi(x) - x' \cdot \nabla \varphi(x)] d\mu(x').$$

This integral remains well-defined for $\varphi \in C_c^2(\mathbb{R}^d)$ due to the cancellation near zero and the integrability of x' at infinity. The standard cutoff-based form (2.5) and the global form are related by

$$\mathcal{A}\varphi(x) = -\kappa' \cdot \nabla \varphi(x) + \int_{\mathbb{R}^d} [\varphi(x + x') - \varphi(x) - \chi_{B_1(0)}(x') x' \cdot \nabla \varphi(x)] d\mu(x'),$$

so switching to the global form simply absorbs the tail drift κ' into the total drift.

While the global form is less common in the literature, it proves useful when studying couplings of processes or generators (see, e.g., [30]), as it simplifies both computations and comparisons. For notational consistency—especially in our later treatment of coupling generators—we adopt this form throughout the paper.

Convention 2.3. Given a Lévy triplet (κ, α, μ) with $\mu \in \Lambda_p(\mathbb{R}^d)$, $p \geq 1$, we denote the associated generator by

$$\mathcal{A} = \Lambda_{\text{glob}}(\kappa, \alpha, \mu), \tag{2.9}$$

defined for $\varphi \in C_0^2(\mathbb{R}^d)$ by the *global (untruncated) form* (in contrast to the *local (truncated) form* in (2.5)):

$$\begin{aligned} (\mathcal{A}\varphi)(x) &:= \mathcal{A}^\nabla \varphi + \mathcal{A}^\Delta \varphi + \mathcal{A}^J \varphi \\ &:= \kappa \cdot \nabla \varphi + \frac{1}{2} \text{tr}[\alpha D^2 \varphi] + \int_{\mathbb{R}^d \setminus \{0\}} [\varphi(x+x') - \varphi(x) - x' \cdot \nabla \varphi(x)] d\mu(x'). \end{aligned} \quad (2.10)$$

This form is well-defined under (2.7) and will be used as the default representation throughout the paper. Unless otherwise stated, all Lévy measures μ considered in this paper are assumed to have finite second moment (2.8), i.e., $\mu \in \Lambda_2(\mathbb{R}^d)$.

2.2.4. Mean vectors and covariance matrices of Lévy processes. Given a Lévy process $\{X_t\}_{t \geq 0}$ on $\mathbb{R}^d$ with generator $\mathcal{A}$, suppose that the process—equivalently, the Lévy measure μ restricted on $B_1(0)^c$ —admits a finite first moment. Then there exists a vector $m_{\mathcal{A}} \in \mathbb{R}^d$, called the *mean vector* of the process (or of the generator $\mathcal{A}$), such that the expectation of the process starting from the origin satisfies

$$e^{t\mathcal{A}}[x](0) = \mathbb{E}^0[X_t] = tm_{\mathcal{A}}.$$

Here, the expression $e^{t\mathcal{A}}[x]$ denotes the action of the semigroup $e^{t\mathcal{A}}$ on the coordinate function $x = (x_1, x_2, \dots, x_d)$, applied componentwise. Taking the time derivative at $t = 0$, we obtain the identity

$$m_{\mathcal{A}} = \mathcal{A}[x](0) = \lim_{t \searrow 0} \frac{e^{t\mathcal{A}}[x](0)}{t}, \quad (2.11)$$

where $\mathcal{A}[x]$ is interpreted componentwise. Alternatively, for any vector $\theta \in \mathbb{R}^d$, define the linear functional $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$, $\varphi(x) := x^\top \theta$. Then $\varphi \in \bar{D}(\mathcal{A})$, and the generator acts on φ as

$$(\mathcal{A}\varphi)(x) = \theta^\top (x + tm_{\mathcal{A}}).$$

The mean vector $m_{\mathcal{A}}$ can be explicitly expressed in terms of the Lévy triplet (κ, α, μ) , but its precise form depends on the choice of cutoff function used in the definition of the jump operator. For example, under the standard (localized) representation (2.5), which uses the cutoff function $\chi_{B_1(0)}$, the mean vector is given by

$$m_{\mathcal{A}} = \kappa + \int_{B_1(0)^c} x' d\mu(x').$$

In contrast, for the global (untruncated) form (2.10), where the compensation term is applied over the entire domain $\mathbb{R}^d \setminus \{0\}$, the mean vector reduces to

$$m_{\mathcal{A}} = \kappa.$$

Both expressions follow directly from the general identity (2.11).

The second-order structure of a Lévy process can be described in terms of its *covariance matrix*. Assume that the process admits a finite second moment, equivalently, the Lévy

measure $\mu \in \Lambda_2(\mathbb{R}^d)$ or the Lévy generator $\mathcal{A} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$. Then there exists a symmetric non-negative definite matrix $Q_{\mathcal{A}}$, called the *covariance matrix* of the process (or of the generator $\mathcal{A}$), such that the centered second moment satisfies

$$\text{Cov}^0(X_t) = \mathbb{E}^0[(X_t - tm_{\mathcal{A}})(X_t - tm_{\mathcal{A}})^\top] = t Q_{\mathcal{A}}.$$

Equivalently, for any quadratic function $\varphi(x) = x^\top U x$, where $U \in M(\mathbb{R}^d)$ is a $(d \times d)$ -matrix, we have $\varphi \in \bar{D}(\mathcal{A})$, and

$$(\mathcal{A}\varphi)(x) = \text{tr}(U Q_{\mathcal{A}}) + 2x^\top U m_{\mathcal{A}}.$$

Particularly the following identity holds:

$$(e^{t\mathcal{A}}\varphi)(x) = (x + tm_{\mathcal{A}})^\top U (x + tm_{\mathcal{A}}) + t \text{tr}(U Q_{\mathcal{A}}).$$

In terms of the Lévy triplet (κ, α, μ) , the covariance matrix takes the explicit form

$$Q_{\mathcal{A}} = \alpha + \int_{\mathbb{R}^d} x x^\top d\mu(x),$$

when expressed in the global representation (2.10). Under the localized representation (2.5), the integral is restricted to $B_1(0)$, so that

$$Q_{\mathcal{A}} = \alpha + \int_{B_1(0)} x x^\top d\mu(x).$$

As in the case of the mean vector, the precise decomposition depends on the chosen cutoff function, but the covariance matrix itself is intrinsic to the process.

2.3. Optimal transport between Markov flows. Let us return to the abstract general setting. Let Π be a state space equipped with a cost function $c : \Pi^2 \rightarrow [0, \infty)$. Given two probability measures $\mu, \nu \in \mathcal{P}(\Pi)$, we denote by $\Gamma(\mu, \nu)$ the set of all *couplings* of μ and ν , that is, the set of probability measures $\gamma \in \mathcal{P}(\Pi^2)$ such that

$$\gamma(E \times \Pi) = \mu(E), \quad \gamma(\Pi \times E) = \nu(E), \quad \text{for all measurable } E \subset \Pi.$$

The *c-optimal transport cost* between μ and ν is defined as

$$\mathcal{C}_c(\mu, \nu) := \inf_{\gamma \in \Gamma(\mu, \nu)} \int_{\Pi^2} c(x, y) d\gamma(x, y). \quad (2.12)$$

Any coupling $\gamma \in \Gamma(\mu, \nu)$ that attains this infimum is called a *c-optimal coupling*. The *c-optimal transport problem* asks whether such an optimal coupling exists. Existence is guaranteed under mild assumptions—for example, if c is lower semicontinuous.²

The *Kantorovich dual formulation* of (2.12) is given by

$$\sup \left[\int_{\Pi} \varphi(x) d\mu(x) + \int_{\Pi} \psi(y) d\nu(y) \right], \quad (2.13)$$

where the supremum is taken over all continuous functions $\varphi, \psi : \Pi \rightarrow \mathbb{R}$ satisfying the pointwise inequality

$$\varphi(x) + \psi(y) \leq c(x, y) \quad \text{for all } x, y \in \Pi.$$

²A finite *c*-moment condition on μ and ν is often imposed to ensure that $\mathcal{C}_c(\mu, \nu) < \infty$.

The Kantorovich duality theorem asserts that, under appropriate conditions on c , μ , and ν , the infimum in (2.12) coincides with the supremum in (3.10).

In many applications, a central question is how the c -optimal transport cost between two probability measures evolves under the action of Markov flows. Let $\mathcal{A}, \mathcal{B} \in \mathcal{G}(\Pi)$ be Feller generators on Π , and μ, ν be two probability measures. One then studies the evolution of the function

$$t \mapsto \rho(t) = \rho(t; \mu, \nu) := \mathcal{C}_c(\mu e^{t\mathcal{A}}, \nu e^{t\mathcal{B}}),$$

which tracks the c -optimal transport cost between the time-evolved measures $\mu e^{t\mathcal{A}}$ and $\nu e^{t\mathcal{B}}$. A key goal is to bound the growth of $\rho(t)$ in terms of the generators $\mathcal{A}, \mathcal{B}$ and the initial measures μ, ν . Leveraging the semigroup property of the flows, such bounds can often be derived by analyzing the derivative of $\rho(t; \mu, \nu)$ at time $t = 0$ for $\mu = \delta_x, \nu = \delta_y, x, y \in \Pi$. This motivates the following definition.

Definition 2.4 (Pointwise transport derivatives between Markov flows). Let $\mathcal{A}, \mathcal{B} \in \mathcal{G}(\Pi)$. The *pointwise upper transport derivative* is the map

$$\omega_c^+(x, y; \mathcal{A}, \mathcal{B}) := \limsup_{t \searrow 0} \frac{\mathcal{C}_c(\delta_x e^{t\mathcal{A}}, \delta_y e^{t\mathcal{B}}) - c(x, y)}{t}, \quad (x, y) \in \Pi^2. \quad (2.14)$$

Similarly, the *pointwise lower transport derivative* is given by

$$\omega_c^-(x, y; \mathcal{A}, \mathcal{B}) := \liminf_{t \searrow 0} \frac{\mathcal{C}_c(\delta_x e^{t\mathcal{A}}, \delta_y e^{t\mathcal{B}}) - c(x, y)}{t}, \quad (x, y) \in \Pi^2.$$

Remark 2.5. By definition, one always has $\omega_c^-(x, y; \mathcal{A}, \mathcal{B}) \leq \omega_c^+(x, y; \mathcal{A}, \mathcal{B})$. A natural question is whether these two bounds coincide, that is, whether the map $t \mapsto \mathcal{C}_c(\delta_x e^{t\mathcal{A}}, \delta_y e^{t\mathcal{B}})$ is differentiable at $t = 0$. This is related to the *optimality gap problem*. See Problem 2.22.

The notion of the *pointwise transport derivative* provides a local-in-time measure of how the optimal transport cost evolves under the action of two Markov semigroups. It was first appeared in [1] (in the case $\Pi = \mathbb{R}^d$ and $c(x, y) = |x - y|^p$) and further developed in [36], where it is used to study the stability of optimal transport costs along Markov flows. In particular, it is shown that if there exist constants $\alpha, \beta > 0$ such that

$$\omega_c^+(x, y; \mathcal{A}, \mathcal{B}) \leq \alpha + \beta c(x, y) \quad \text{for all } x, y \in \Pi,$$

then the following exponential stability estimate holds: for any pair of probability measures $\mu, \nu \in \mathcal{P}(\Pi)$ with finite c -moment,

$$\mathcal{C}_c(\mu e^{t\mathcal{A}}, \nu e^{t\mathcal{B}}) \leq \mathcal{C}_c(\mu, \nu) e^{\beta t} + \frac{\alpha}{\beta} (e^{\beta t} - 1).$$

To estimate or control $\omega_c^\pm$ in concrete applications, it is often convenient to consider its dual formulation. This leads to the definition of the *dual transport derivative*, defined as follows:

Definition 2.6 (Dual transport derivative). Given $\mathcal{A}, \mathcal{B} \in \mathcal{G}(\Pi)$ and points $x_0, y_0 \in \Pi$, define the *dual value* of (2.14) by

$$\omega_c^*(x_0, y_0; \mathcal{A}, \mathcal{B}) := \sup_{(\varphi, \psi) \in \mathcal{D}(x_0, y_0)} [\mathcal{A}\varphi(x_0) + \mathcal{B}\psi(y_0)], \quad (2.15)$$

where $\mathcal{D}(x_0, y_0)$ is the set of all pairs of functions $\varphi, \psi : \Pi \rightarrow \mathbb{R}$ such that:

- $\varphi \in \bar{D}(\mathcal{A})$, $\psi \in \bar{D}(\mathcal{B})$;
- $(x, y) \mapsto c(x, y) - \varphi(x) - \psi(y)$ achieves a global minimum at (x_0, y_0) .

Remark 2.7. Since both $\mathcal{A}$ and $\mathcal{B}$ annihilate constants, the functional

$$(\varphi, \psi) \mapsto (\mathcal{A}\varphi)(x_0) + (\mathcal{B}\psi)(y_0)$$

is invariant under constant shifts of the form $\varphi \mapsto \varphi + b$ and $\psi \mapsto \psi + b'$ with $b, b' \in \mathbb{R}$. Consequently, by adding suitable constants, we may assume without loss of generality that $(\varphi, \psi) \in \mathcal{D}(x_0, y_0)$ in (2.15) satisfies the *touching condition*

$$\varphi(x) + \psi(y) \leq c(x, y), \quad \text{with equality at } (x_0, y_0).$$

Remark 2.8. In Definition 2.6, the choice of domains $\bar{D}(\mathcal{A})$ and $\bar{D}(\mathcal{B})$ is not essential. One may instead work with smaller domains such as $D(\mathcal{A})$, $D(\mathcal{B})$ (as in [36]), or with other cores or dense subspaces. For example, when $\Pi = \mathbb{R}^d$, a natural choice is the space $C_2^2(\mathbb{R}^d)$, which is the main domain adopted in the present work. In many situations these domains are dense in one another, so the value of the supremum in (2.15) remains unchanged.

The notion of dual transport derivative was also first explored in [1], and further investigated in [36]. As hinted in the definition, the notion is inspired by the theory of *viscosity solutions*. In particular, the relationship between (2.15) and (2.14) closely parallels the classical duality between the Kantorovich formulation (2.13) and the primal optimal transport problem (2.12). Just as Kantorovich duality equates the infimum over couplings with a supremum over admissible potentials, the dual formulation (2.15) offers a variational perspective on the infinitesimal growth of the transport cost in (2.14). This analogy suggests that, under suitable conditions, the primal and dual transport derivatives coincide: $\omega_c^* = \omega_c^\pm$. Determining whether this equality holds is precisely the content of the optimality gap problem, Problem 2.22, which will also be addressed in the present work.

2.4. Optimal Markovian couplings. In this subsection, we formulate our first main problem—the optimal coupling problem between two Feller processes, and equivalently, between their generators. To set the stage, we first introduce the notion of *couplings between processes and generators*.

2.4.1. Feller couplings of generators. Let $\{X_t\}_{t \geq 0}$ and $\{Y_t\}_{t \geq 0}$ be two Π -valued stochastic processes defined on a common state space Π , and let $\{\mu_t\}_{t \geq 0}$ and $\{\nu_t\}_{t \geq 0}$ denote their respective laws:

$$\mu_t = \text{law}(X_t), \quad \nu_t = \text{law}(Y_t).$$

A *pointwise (in time) coupling* of the two processes is a family $\{\gamma_t\}_{t \geq 0}$ of probability measures on Π^2 such that

$$\gamma_t \in \Gamma(\mu_t, \nu_t) \quad \text{for all } t \geq 0.$$

This notion provides a coupling between the marginal distributions at each fixed time, but it does not encode any joint temporal behavior of the processes. In particular, the family $\{\gamma_t\}_{t \geq 0}$ is not required to arise from the law of a stochastic process on Π^2 , and thus lacks the dynamical structure needed to study time-dependent interactions between X_t and Y_t .

To overcome this limitation, one often considers a (*general*) *coupling of stochastic processes*, defined as a Π^2 -valued stochastic process $\{Z_t = (X'_t, Y'_t)\}_{t \geq 0}$ such that for all $t \geq 0$, the marginals of Z_t agree with those of the original processes:

$$\text{law}(X'_t) = \text{law}(X_t), \quad \text{law}(Y'_t) = \text{law}(Y_t). \quad (2.16)$$

If both $\{X_t\}$ and $\{Y_t\}$ are Feller (or more generally, Markov) processes, it is natural to restrict attention to couplings $\{Z_t\}_{t \geq 0}$ that also preserve the Markovian structure. That is, we seek couplings that themselves form a Markov process on Π^2 . In the Feller setting, such processes are fully characterized by their generators, which leads to the following definition of *Feller couplings*. Recall the definition of continuity domain $C(\mathcal{A})$ and extended domain $\bar{D}(\mathcal{A})$ of a generator $\mathcal{A}$ from Definition 2.1.

Definition 2.9 (Markovian/Feller couplings). Let $\mathcal{A}, \mathcal{B} \in \mathcal{G}(\Pi)$ be two Feller generators.

(i) A Markov semigroup $\{T_t\}_{t \geq 0}$ is a *Markovian coupling semigroup* of $\mathcal{A}, \mathcal{B}$ if it holds for all $\varphi \in C(\mathcal{A}), \psi \in C(\mathcal{B})$ and $t \geq 0$

$$T_t[\varphi \otimes 1] = e^{t\mathcal{A}}\varphi \otimes 1, \quad T_t[1 \otimes \psi] = 1 \otimes e^{t\mathcal{B}}\psi.$$

Here, $\varphi \otimes 1$ and $1 \otimes \psi$ denote the functions on Π^2 defined by

$$(\varphi \otimes 1)(x, y) := \varphi(x), \quad (1 \otimes \psi)(x, y) := \psi(y).$$

(ii) A *Feller coupling generator* of $\mathcal{A}, \mathcal{B}$ is a Feller generator $\mathcal{J} \in \mathcal{G}(\Pi^2)$ such that $\{e^{t\mathcal{J}}\}_{t \geq 0}$ is a Markovian coupling semigroup of $\mathcal{A}, \mathcal{B}$. That is, for all $\varphi \in C(\mathcal{A}), \psi \in C(\mathcal{B})$ and all $t \geq 0$, we have

$$e^{t\mathcal{J}}[\varphi \otimes 1] = e^{t\mathcal{A}}\varphi \otimes 1, \quad e^{t\mathcal{J}}[1 \otimes \psi] = 1 \otimes e^{t\mathcal{B}}\psi.$$

We denote by $\Gamma(\mathcal{A}, \mathcal{B})$ the set of all Feller coupling generators of $\mathcal{A}$ and $\mathcal{B}$.

Remark 2.10. This definition coincides with that in [39]. The notion of a coupling generator can be extended beyond the Feller setting by replacing “Feller generators” with their general *Markovian* counterparts. Furthermore, a broader concept of coupling—formulated in terms of martingale problems—is also available, e.g., see [29]. In the present work, however, we will remain within the Feller framework.

Remark 2.11. Given two Feller generators $\mathcal{A}, \mathcal{B}$, the set of couplings $\Gamma(\mathcal{A}, \mathcal{B})$ is always nonempty. For instance, the operator

$$\mathcal{J} := \mathcal{A} \otimes I + I \otimes \mathcal{B}$$

belongs to $\Gamma(\mathcal{A}, \mathcal{B})$ and corresponds to the generator of the *independent coupling* of the processes generated by $\mathcal{A}$ and $\mathcal{B}$.

We now present an equivalent characterization of Feller coupling generators, stated as a proposition below. The proof is elementary and will be given in Appendix B.1.

Proposition 2.12. Let $\mathcal{A}, \mathcal{B} \in \mathcal{G}(\Pi)$ and $\mathcal{J} \in \mathcal{G}(\Pi^2)$. The following are equivalent.

- (i) $\mathcal{J}$ is a Feller coupling generator of $\mathcal{A}, \mathcal{B}$;
- (ii) for all $\varphi \in \bar{D}(\mathcal{A}), \psi \in \bar{D}(\mathcal{B})$, it holds $\varphi \otimes 1, 1 \otimes \psi \in \bar{D}(\mathcal{J})$ and

$$\mathcal{J}[\varphi \otimes 1] = \mathcal{A}\varphi \otimes 1, \quad \mathcal{J}[1 \otimes \psi] = 1 \otimes \mathcal{B}\psi;$$

(iii) for all $\varphi \in D(\mathcal{A}), \psi \in D(\mathcal{B})$, it holds $\varphi \otimes 1, 1 \otimes \varphi \in \bar{D}(\mathcal{J})$, and the marginal condition from (ii) holds.

2.4.2. Optimal Markovian couplings. Let $\{X_t\}$ and $\{Y_t\}$ be Feller processes on a Polish space (Π, c) , with generators $\mathcal{A}$ and $\mathcal{B}$, respectively. We are interested in constructing a coupling of these processes that is *optimal* with respect to the transport cost c . A natural approach is to consider a family of *pointwise* couplings $\{\gamma_t\}_{t \geq 0}$, where each $\gamma_t \in \mathcal{P}(\Pi^2)$ is a c -optimal coupling of the marginals $\mu_t = \text{law}(X_t)$ and $\nu_t = \text{law}(Y_t)$. While this yields the minimal transport cost at each time, as mentioned earlier it lacks temporal structure for stochastic processes: there is no guarantee that $\{\gamma_t\}_{t \geq 0}$ corresponds to the time marginals of any joint process $\{(X'_t, Y'_t)\}_{t \geq 0}$ on Π^2 .

To incorporate temporal dynamics, one instead seeks a coupling at the level of Markov processes: a Π^2 -valued Markov/Feller process $\{Z_t^* = (X_t^*, Y_t^*)\}_{t \geq 0}$ whose marginals coincide with those of $\{X_t\}$ and $\{Y_t\}$. The aim is to construct such a process-level (Markovian) coupling that satisfies an optimality criterion with respect to the cost function c , for example:

$$\mathbb{E}^{(x,y)}[c(X_t^*, Y_t^*)] = \inf_{\{(X'_t, Y'_t)\}} \mathbb{E}^{(x,y)}[c(X'_t, Y'_t)], \quad \text{for all } (t, x, y) \in [0, \infty) \times \Pi^2,$$

where the infimum is taken over all Markovian/Feller couplings $\{(X'_t, Y'_t)\}_{t \geq 0}$ of $\{X_t\}$ and $\{Y_t\}$. Alternatively, in the semigroup formulation, we are to identify the Feller generator $\mathcal{J}_* \in \Gamma(\mathcal{A}, \mathcal{B})$ such that the following holds for all $\mathcal{J} \in \Gamma(\mathcal{A}, \mathcal{B})$

$$e^{t\mathcal{J}_*}c(x, y) \leq e^{t\mathcal{J}}c(x, y), \quad \forall (t, x, y) \in [0, \infty) \times \Pi^2. \quad (2.17)$$

However, this global-in-time optimality may be too strong: the infimum need not be attained or even well-defined at each time $t \geq 0$. To circumvent this, one instead considers couplings that minimize the *initial rate of change* of the cost—that is, the time derivative of $\mathbb{E}[c(X'_t, Y'_t)]$ at $t = 0$:

$$\left. \frac{d}{dt} \right|_{t=0} \mathbb{E}^{(x,y)}[c(X'_t, Y'_t)] = \left. \frac{d}{dt} \right|_{t=0} (e^{t\mathcal{J}}c)(x, y) = (\mathcal{J}c)(x, y),$$

where $\mathcal{J}$ is the generator of the coupling process $\{(X'_t, Y'_t)\}$, which is a Feller coupling generator of $\mathcal{A}, \mathcal{B}$. This leads to the following optimality condition, as formulated by Chen in [14]:

Definition 2.13 (c -optimal coupling generator and Markovian transport derivative). Let Π be a state space, $c : \Pi^2 \rightarrow [0, \infty)$ be a cost function, and let $\mathcal{A}, \mathcal{B} \in \mathcal{G}(\Pi)$ be two Feller generators on Π . The *Markovian transport derivative* is the function $\theta_c(\cdot, \cdot; \mathcal{A}, \mathcal{B}) : \Pi^2 \rightarrow [-\infty, \infty]$ defined by

$$\theta_c(x, y; \mathcal{A}, \mathcal{B}) := \inf_{\mathcal{J} \in \Gamma(\mathcal{A}, \mathcal{B})} (\mathcal{J}^- c)(x, y), \quad (2.18)$$

where $\mathcal{J}^-$ is the lower generator given in Definition 2.2. We say that $\mathcal{J}_* \in \Gamma(\mathcal{A}, \mathcal{B})$ is a *c-optimal coupling* of $\mathcal{A}$ and $\mathcal{B}$ if $c \in \bar{D}(\mathcal{J}_*)$ and it holds for all $(x, y) \in \mathbb{R}^{2d}$:

$$(\mathcal{J}_* c)(x, y) = \theta_c(x, y; \mathcal{A}, \mathcal{B}). \quad (2.19)$$

Remark 2.14. A mild technical requirement in the definition of θ_c is that $e^{t\mathcal{J}}c$ be well defined for small times $t \geq 0$. This is guaranteed, for example, if c admits a subadditive bound of the form

$$c(x, y) \leq \Phi(x) + \Psi(y), \quad \Phi \in C(\mathcal{A}), \Psi \in C(\mathcal{B}), \quad (2.20)$$

where $C(\mathcal{A})$ and $C(\mathcal{B})$ denote the continuity domains of $\mathcal{A}$ and $\mathcal{B}$.

Remark 2.15. In (2.18) we use the lower generator $\mathcal{J}^-$ to avoid the domain issue of whether c belongs to $\bar{D}(\mathcal{J})$. In the jump-process setting of [14], the situation is simpler: $\mathcal{A}$ and $\mathcal{B}$ are bounded Markov jump operators with jump kernels $\{\lambda(x, dx')\}_{x \in \Pi}$, and under mild assumptions, e.g., (2.20), the cost c always lies in the extended domain of every coupling $\mathcal{J} \in \Gamma(\mathcal{A}, \mathcal{B})$. There the infimum in (2.18) is taken over all Markovian couplings (not necessarily Feller), and the generator itself can be used directly. For unbounded generators, however, c may fail to belong to $\bar{D}(\mathcal{J})$, so the lower generator $\mathcal{J}^-$ provides a robust formulation.

Remark 2.16. An alternative notion of the Markovian transport derivative, denoted by $\tilde{\theta}_c$, can be introduced in integral form, motivated by the semigroup identity (2.4). Specifically, $\tilde{\theta}_c$ is defined as the largest function $\theta : \Pi^2 \rightarrow [-\infty, \infty]$ such that for every $t \geq 0$, $(x, y) \in \Pi^2$, and every coupling $\mathcal{J} \in \Gamma(\mathcal{A}, \mathcal{B})$,

$$(e^{t\mathcal{J}}c)(x, y) \geq c(x, y) + \int_0^t (e^{s\mathcal{J}}\theta)(x, y) ds.$$

Letting $\mathcal{H}$ denote the collection of all such admissible functions θ , one has $\tilde{\theta}_c = \sup_{\theta \in \mathcal{H}} \theta$ pointwise. It is immediate from the definitions that $\theta_c \leq \tilde{\theta}_c$. The main advantage of this integral formulation lies in its natural extension to general Markovian (not necessarily Feller) couplings, or even coupling processes. On the other hand, it also introduces certain technical challenges, such as in establishing weak duality (cf. Proposition 2.21). In the Lévy setting considered here, however, the two notions coincide.

Given two Feller generators $\mathcal{A}, \mathcal{B} \in \mathcal{G}(\Pi)$, the problem of determining whether there exists a c -optimal Markovian coupling, that is, $\mathcal{J}_* \in \Gamma(\mathcal{A}, \mathcal{B})$ that attains the infimum in (2.19), is referred to as the *c-optimal Markovian coupling problem of $\mathcal{A}, \mathcal{B}$* .

Problem 2.17 (*c-optimal Markovian coupling problem*). Given $\mathcal{A}, \mathcal{B} \in \mathcal{G}(\Pi)$, does there exist a c -optimal coupling $\mathcal{J}_* \in \Gamma(\mathcal{A}, \mathcal{B})$ in the sense of Definition 2.13?

We emphasize that the above is a *global optimization* problem: the quantity $(\mathcal{J}_*c_2)(x, y)$ is minimized simultaneously for all $(x, y) \in \Pi^2$, rather than at a fixed point.

Remark 2.18. In the literature, the *optimal Markovian coupling problem* is often formulated for general *Markov generators*. In this work, we adopt the *Feller generator* setting. This restriction is deliberate: it provides a natural topological structure on the state space, avoids certain technical difficulties arising in the absence of topology, and aligns with the functional analytic approach in our work. Consequently, all “Markovian couplings” considered in this paper are, in fact, *Feller* couplings. We retain the term “Markovian” in the problem name for consistency with the established literature, but the analysis is carried out entirely within the Feller framework.

Before proceeding, we record a simple observation showing how the global-in-time optimality condition (2.17) immediately yields the pointwise (infinitesimal) optimality condition (2.18) at the level of generators. We note it was also proved in [14] for the case of bounded jump operators.

Proposition 2.19. *Let $\mathcal{A}, \mathcal{B} \in \mathcal{G}(\Pi)$ and $\mathcal{J}, \mathcal{L} \in \Gamma(\mathcal{A}, \mathcal{B})$. If*

$$(e^{t\mathcal{L}}c)(x, y) \leq (e^{t\mathcal{J}}c)(x, y), \quad \forall x, y \in \Pi, \quad \forall t \geq 0,$$

then

$$(\mathcal{L}^-c)(x, y) \leq (\mathcal{J}^-c)(x, y), \quad \forall x, y \in \Pi.$$

In particular, if $\mathcal{J}_$ satisfies (2.17) for all $\mathcal{J} \in \Gamma(\mathcal{A}, \mathcal{B})$ and $c \in \bar{D}(\mathcal{J})$, then $\mathcal{J}_*$ is a c -optimal coupling of $\mathcal{A}$ and $\mathcal{B}$.*

Proof. Fix $x, y \in \Pi$ and $t > 0$. Subtract $c(x, y)$ from both sides of the assumed inequality, divide by t , and let $t \searrow 0$. The semigroup definition of $\mathcal{J}$ and $\mathcal{L}$ then yields the claim. $\square$

Remark 2.20. It is natural to ask whether the converse of Proposition 2.19 holds: namely, if $\mathcal{J}_*$ is a c -optimal coupling of $\mathcal{A}$ and $\mathcal{B}$, does it necessarily satisfy the global minimality condition (2.17) for all $\mathcal{J} \in \Gamma(\mathcal{A}, \mathcal{B})$? Preliminary considerations suggest that this implication is *not* valid in general, and constructing an explicit counterexample would be of independent interest.

2.5. Comparison of transport derivative: pointwise, dual, and Markovian. Given two Feller generators $\mathcal{A}, \mathcal{B} \in \mathcal{G}(\Pi)$, generating Feller processes $\{X_t\}$ and $\{Y_t\}$, we have introduced three notions of transport derivative associated with the pair: the *pointwise* $\omega_c^\pm$, the *dual* ω_c^* , and the *Markovian* transport derivatives θ_c , see Definitions 2.4, 2.6 and 2.13. Particularly, the pointwise version captures the (least upper bound) rate of change of the optimal transport cost between the marginal laws under *pointwise couplings* of $\{X_t\}$ and $\{Y_t\}$. In contrast, the Markovian version measures the derivative of the cost under the most efficient *Feller coupling* of the two processes. Since every Feller coupling induces a pointwise coupling, the pointwise infinitesimal transport cost provides a lower bound on the Markovian counterpart. Hence we have the following proposition, whose proof is presented in Appendix B.

Proposition 2.21. *Let Π be a state space, $c : \Pi^2 \rightarrow \mathbb{R}^+$ be a cost function, and let $\mathcal{A}, \mathcal{B} \in \mathcal{G}(\Pi)$. Let $\omega_c^\pm, \omega_c^*$ and θ_c be the transport derivative functions defined by Definitions 2.4, 2.6, and 2.13, respectively. Then for all $x, y \in \Pi$, it holds that*

$$\omega_c^*(x, y; \mathcal{A}, \mathcal{B}) \leq \omega_c^-(x, y; \mathcal{A}, \mathcal{B}) \leq \omega_c^+(x, y; \mathcal{A}, \mathcal{B}) \leq \theta_c(x, y; \mathcal{A}, \mathcal{B}).$$

Our second research problem addresses the potential gap between the pointwise and Markovian notions of transport derivative at the level of individual states. This leads to the following question, which we refer to as the *optimality gap problem*.

Problem 2.22 (Optimality gap problem). *Given a state space Π , a cost function $c : \Pi^2 \rightarrow \mathbb{R}^+$, and two Feller generators $\mathcal{A}, \mathcal{B} \in \mathcal{G}(\Pi)$, does the following identities hold:*

$$\omega_c^*(x, y; \mathcal{A}, \mathcal{B}) = \omega_c^-(x, y; \mathcal{A}, \mathcal{B}) = \omega_c^+(x, y; \mathcal{A}, \mathcal{B}) = \theta_c(x, y; \mathcal{A}, \mathcal{B}), \quad \text{for all } x, y \in \Pi?$$

Remark 2.23. If the equality holds, then $\omega_c^- = \omega_c^+$ implies the limit (2.14), i.e., the derivative of $t \mapsto \mathcal{C}_c(\delta_x e^{tA}, \delta_y e^{tA})$ at $t = 0$, exists.

Remark 2.24. The equality in Problem 2.22, if it holds, has several noteworthy implications. In particular, the identity $\theta_c = \omega_c^\pm$ would mean that for any $\varepsilon > 0$, one can construct a coupling process $\{(X_t, Y_t)\}_{t \geq 0}$ between the processes generated by $\mathcal{A}$ and $\mathcal{B}$ such that the infinitesimal growth rate of the transportation cost $\mathbb{E}^{(x,y)}[c(X_t, Y_t)]$ at $t = 0$ is within ε of the optimal value given by $\omega_c^\pm$, i.e., the derivative of the optimal transport cost between $\delta_x e^{tA}$ and $\delta_y e^{tB}$, which serves as the lower bound for $c(X_t, Y_t)$ over all possible (generic) couplings. Moreover, if a c -optimal coupling process exists, the equality would ensure that the instantaneous cost growth at the level of the coupling process coincides with the pointwise optimal value obtained from the dual formulation. Such a property can be particularly relevant in the analysis of interacting particle systems, where the short-time behavior of couplings between generators plays a crucial role in quantifying convergence rates, stability, or propagation of chaos.

2.6. Existing results for optimal Markovian coupling problem. Let us briefly review the existing results for the optimal Markovian coupling problem.

Prior to the development of the optimal Markovian coupling problem with transport-type cost criteria, coupling methods were extensively studied in the context of maximizing the probability that two processes meet by a fixed time. The study of such maximal couplings dates back at least to the pioneering works of Griffeath [25] and Goldstein [24], who constructed maximal couplings for Markov chains and more general probability spaces, respectively. In the setting of continuous-time diffusions, Lindvall and Rogers [37] established that the reflection coupling of Brownian motion achieves this maximal meeting probability, thus providing an explicit optimal coupling construction. Related problems have also been investigated in the setting of general Lévy processes; see [43, 7, 34].

The idea of constructing Markovian couplings to bound the Wasserstein distance between the laws of two Markov processes was, to the best of our knowledge, first introduced by Chen and Li [16]. They showed that one can obtain estimates in the Wasserstein p -metric, $p = 1, 2$, between the laws of two processes by constructing a Markovian coupling that satisfies certain appropriate bounds. Furthermore, they established an optimality criterion and proved the existence of such couplings for the case $p = 2$, which includes the drift–diffusion setting considered in the present work.

Subsequently, Chen [14] formulated this optimality criterion (2.18), along with the optimal Markovian coupling problem in a general abstract setting of Markov processes on an arbitrary state space. He also solved the optimal coupling problem for diffusion processes on $\mathbb{R}^d$ with a concave cost $c(x, y) = f(|x - y|)$, where f satisfies $f(0) = 0$, $f' > 0$, and $f'' \leq 0$, together with several applications [15]. Under a similar setting in cost, the optimal Markovian coupling problem was also studied in the Lévy setting [34, 29]. In a related development, Zhang [49] proved the existence of c -optimal Markovian couplings for bounded Markov jump generators on an abstract state space Π , where c is a nonnegative lower semicontinuous cost function.

2.7. Optimal Markovian coupling and optimality gap problem for Lévy generators. We now return to the two research problems posed earlier, and specialize to the setting of a quadratic cost on the Euclidean space:

$$\Pi = \mathbb{R}^d, \quad c(x, y) = c_2(x, y) = \frac{1}{2}|x - y|^2.$$

In this case, we adopt the subscript “2” in place of the general cost function c in the notation for transport derivatives, e.g. $\theta_c = \theta_2$, $\omega_c^\pm = \omega_2^\pm$, and $\omega_c^* = \omega_2^*$.

In principle, the problems of optimal Markovian coupling and the associated optimality gap can be posed for arbitrary Feller generators $\mathcal{A}, \mathcal{B} \in \mathcal{G}(\mathbb{R}^d)$, or more specific class, *Lévy-type generators*. Such a level of generality, however, appears presently out of reach due to the analytic complexity. As a tractable yet structurally rich first step, we therefore focus on the translation-invariant case, namely on Lévy generators $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$. This restriction preserves many of the essential features of the general problem, while allowing a more transparent development of the ideas introduced above.

Recall the notion of the dual transport derivative ω_c from Definition 2.6. Since $\mathcal{A}, \mathcal{B}$ are Lévy generators, the space $C_2^2(\mathbb{R}^d)$ of quadratically bounded continuous functions serves as a common core for both. In this case it is more convenient to work with (2.15) restricted to this domain. This leads to the following definition.

Definition 2.25 (Restricted dual supremum). Let $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$ and $(x_0, y_0) \in \mathbb{R}^{2d}$. We define the *restricted dual transport derivative*

$$\omega_2'(x_0, y_0; \mathcal{A}, \mathcal{B}) := \sup_{(\varphi, \psi) \in \mathcal{D}'(x_0, y_0)} \left[\mathcal{A}\varphi(x_0) + \mathcal{B}\psi(y_0) \right], \quad (2.21)$$

where $\mathcal{D}'(x_0, y_0)$ consists of all pairs $(\varphi, \psi) \in [C_2^2(\mathbb{R}^d)]^2$ satisfying the *touching condition*

$$\varphi(x) + \psi(y) \leq \frac{1}{2}|x - y|^2 \quad \text{for all } (x, y), \text{ with equality at } (x_0, y_0). \quad (2.22)$$

Equivalently, the definition in (2.21) is unchanged if condition (2.22) is replaced by requiring that $c_2 - \varphi \oplus \psi$ attains a global minimum at (x_0, y_0) ; see Remark 2.7.

Remark 2.26. Since for $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$, $C_2^2(\mathbb{R}^d) \subset \bar{D}(\mathcal{A}) \cap \bar{D}(\mathcal{B})$, clearly it holds for all $x_0, y_0 \in \mathbb{R}^d$:

$$\omega_2'(x_0, y_0; \mathcal{A}, \mathcal{B}) \leq \omega_2^*(x_0, y_0; \mathcal{A}, \mathcal{B}).$$

Before stating our main results, we address a technical question: does the quadratic cost c_2 belong to the continuity domain of a coupling generator $\mathcal{J}$ associated with two Lévy generators? The following proposition gives an affirmative answer. Its proof is elementary and is included in Appendix B.

Proposition 2.27. Let $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$ be Lévy generators with mean vectors $m_{\mathcal{A}}, m_{\mathcal{B}}$ and covariance matrices $Q_{\mathcal{A}}, Q_{\mathcal{B}}$. Define

$$U(t, x, y) := (x - y + t(m_{\mathcal{A}} - m_{\mathcal{B}}))^\top (m_{\mathcal{A}} - m_{\mathcal{B}}),$$

and let $\{T_t\}_{t \geq 0}$ be a Markovian coupling semigroup of $\mathcal{A}, \mathcal{B}$ and $\mathcal{J} \in \Gamma(\mathcal{A}, \mathcal{B})$ be a Feller coupling generator.

(i) It holds for all $t, h \geq 0$ and $(x, y) \in \mathbb{R}^{2d}$:

$$\begin{aligned} \frac{h}{2} U(t, x, y) &\leq (T_{t+h} c_2)(x, y) - (T_t c_2)(x, y) \\ &\leq h^2(|m_{\mathcal{A}}|^2 + |m_{\mathcal{B}}|^2) + h[\text{tr}(Q_{\mathcal{A}} + Q_{\mathcal{B}}) + U(t, x, y)]. \end{aligned}$$

In particular, the map $t \mapsto (e^{t\mathcal{J}} c_2)(x, y)$ is locally Lipschitz.

(ii) One has $c_2 \in C(\mathcal{J})$.

(iii) If $\mathcal{J}$ itself is a Lévy generator, then $c_2 \in \bar{D}(\mathcal{J})$.

Our first main theorem, addressing to the optimal Markovian coupling problem between two Lévy generators, is stated as follows.

Theorem 2.28 (Existence of 2-optimal Markovian coupling). *Let $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$ be two Lévy generators with finite second moments, and let $\theta_2(\cdot, \cdot; \mathcal{A}, \mathcal{B})$ denote the Markovian transport derivative of $\mathcal{A}, \mathcal{B}$ defined in Definition 2.13.*

(i) It holds that

$$\theta_2(x, y; \mathcal{A}, \mathcal{B}) = \theta_2(0, 0; \mathcal{A}, \mathcal{B}) + (m_{\mathcal{A}} - m_{\mathcal{B}})^\top (x - y),$$

where $m_{\mathcal{A}}, m_{\mathcal{B}} \in \mathbb{R}^d$ are the mean vectors associated with $\mathcal{A}$ and $\mathcal{B}$, respectively.

(ii) A 2-optimal Markovian coupling $\mathcal{J}_* \in \Gamma(\mathcal{A}, \mathcal{B})$ exists. In particular, if the generators are in global form with Lévy triplets $\mathcal{A} = \Lambda_{\text{glob}}(\kappa, \alpha, \mu)$ and $\mathcal{B} = \Lambda_{\text{glob}}(\zeta, \beta, \nu)$, then $\mathcal{J}_* = \Lambda_{\text{glob}}(\eta_*, \sigma_*, \gamma_*) \in \mathcal{G}_2^\Lambda(\mathbb{R}^{2d})$ is a 2-optimal coupling, where:

- $\eta_* = \kappa \oplus \zeta \in \mathbb{R}^{2d}$;
- $\sigma_* = \begin{bmatrix} \alpha & K_* \\ K_*^\top & \beta \end{bmatrix}$, where $K_* \in M(\mathbb{R}^d)$ solves the semidefinite program (4.6);
- $\gamma_* \in \Lambda_2(\mathbb{R}^{2d})$ is a 2-optimal Lévy coupling of μ and ν , given by Definition 3.13.

(iii) For any 2-optimal coupling $\mathcal{J}_*$ of $\mathcal{A}, \mathcal{B}$, it holds that

$$(e^{t\mathcal{J}_*} c_2)(x, y) = c_2(x, y) + t\theta_2(x, y; \mathcal{A}, \mathcal{B}) + \frac{t^2}{2}|m_{\mathcal{A}} - m_{\mathcal{B}}|^2. \quad (2.23)$$

In particular, the semigroup $\{e^{t\mathcal{J}_*}\}_{t \geq 0}$ satisfies the following minimal growth condition: for every Markovian coupling semigroup $\{T_t\}_{t \geq 0}$ of $\mathcal{A}, \mathcal{B}$,

$$(e^{t\mathcal{J}_*} c_2)(x, y) \leq (T_t c_2)(x, y), \quad \forall x, y \in \mathbb{R}^d, t \geq 0. \quad (2.24)$$

Remark 2.29. We emphasize that the minimal growth property in (2.24) holds among *all* Markovian coupling semigroups of $\mathcal{A}, \mathcal{B}$, not only the Feller ones.

Remark 2.30. If $\mathcal{A}, \mathcal{B}$ are of global form with $\mathcal{A} = \Lambda_{\text{glob}}(\kappa, \alpha, \mu)$, $\mathcal{B} = \Lambda_{\text{glob}}(\zeta, \beta, \nu)$, the identity from (i) becomes

$$\theta_2(x, y; \mathcal{A}, \mathcal{B}) = \theta_2(0, 0; \mathcal{A}, \mathcal{B}) + (\kappa - \zeta)^\top (x - y).$$

Our second main result addresses the *optimality gap problem*.

Theorem 2.31 (Strong duality). *Let $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$, and let $\omega_2^\pm, \omega_2^*, \theta_2, \omega_2'$ be defined as in Definitions 2.4, 2.6, 2.13 and 2.25. Then, for all $x, y \in \mathbb{R}^d$, it holds that*

$$\omega_2'(x, y; \mathcal{A}, \mathcal{B}) = \omega_2^*(x, y; \mathcal{A}, \mathcal{B}) = \omega_2^\pm(x, y; \mathcal{A}, \mathcal{B}) = \theta_2(x, y; \mathcal{A}, \mathcal{B}).$$

As a byproduct of Theorem 2.31, we obtain an additivity property of the transport derivative with respect to the Lévy–Khintchine decomposition.

Theorem 2.32 (Additivity w.r.t Lévy–Khintchine decomposition). *Let $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$ be in the global form, and $\mathcal{A}^\bullet, \mathcal{B}^\bullet$, $\bullet \in \{\nabla, \Delta, J\}$ be the drift, diffusion, and jump part of $\mathcal{A}, \mathcal{B}$ respectively. It holds for all $x, y \in \mathbb{R}^d$:*

$$\theta_2(x_0, y_0; \mathcal{A}, \mathcal{B}) = \sum_{\bullet \in \{\nabla, \Delta, J\}} \theta_2(x, y; \mathcal{A}^\bullet, \mathcal{B}^\bullet).$$

2.8. Wasserstein-type metric on the space $\mathcal{G}_2^\Lambda(\mathbb{R}^d)$. An interesting application of Theorems 2.28 and 2.31 is the introduction of a Wasserstein-type metric on the space of Lévy generators $\mathcal{G}_2^\Lambda(\mathbb{R}^d)$ and the space $\Lambda_2(\mathbb{R}^d)$ of Lévy measures with finite second moments.

In what follows, we define the corresponding functional, termed the *Wasserstein generator metric*, as

$$\mathcal{W}_G(\mathcal{A}, \mathcal{B})^2 := \frac{1}{2} |m_{\mathcal{A}} - m_{\mathcal{B}}|^2 + \theta_2(0, 0; \mathcal{A}, \mathcal{B}), \quad (2.25)$$

where $m_{\mathcal{A}}$ and $m_{\mathcal{B}}$ denote the mean vectors associated with the generators $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$, respectively. Since any Lévy generator in $\mathcal{G}_2^\Lambda(\mathbb{R}^d)$ is associated with a Lévy triplet (in the global form (2.10)), the metric (2.25) can be expressed in term of the triplet. As seen in Theorem 2.32, this functional admits a natural decomposition into separate contributions from the drift, diffusion, and jump parts (in the global form). Specifically, we have the drift part $\theta_2(0, 0; \mathcal{A}^\nabla, \mathcal{B}^\nabla) = 0$, and thus

$$\mathcal{W}_G(\mathcal{A}, \mathcal{B})^2 = \frac{1}{2} |m_{\mathcal{A}} - m_{\mathcal{B}}|^2 + \theta_2(0, 0; \mathcal{A}^\Delta, \mathcal{B}^\Delta) + \theta_2(0, 0; \mathcal{A}^J, \mathcal{B}^J).$$

Moreover, if $\mathcal{A} = \Lambda_{\text{glob}}(\kappa, \alpha, \mu)$, $\mathcal{B} = \Lambda_{\text{glob}}(\eta, \beta, \nu)$, then the above reduces to

$$\mathcal{W}_G(\mathcal{A}, \mathcal{B})^2 = \frac{1}{2} (|\kappa - \eta|^2 + \mathcal{W}_S(\alpha, \beta)^2 + \mathcal{W}_\Lambda(\mu, \nu)^2),$$

where: $|\kappa - \eta|$ is the standard Euclidean norm, $\mathcal{W}_S$ denotes the *Bures–Wasserstein distance* between symmetric nonnegative definite matrices, $\mathcal{W}_\Lambda := (\mathcal{C}_2^\Lambda)^{1/2}$ is the square root of the *Lévy transport cost* between Lévy measures, which will be called the *Lévy–Wasserstein metric*. These functional will be defined in Sections 3 and 4.

The main result of this subsection can be stated as follows.

Theorem 2.33. *The functional $\mathcal{W}_G$ from (2.25) defines a complete and separable metric on the space of Lévy generators $\mathcal{G}_2^\Lambda(\mathbb{R}^d)$.*

As a direct consequence, $\mathcal{W}_\Lambda$ also induces a complete and separable metric on the space $\Lambda_2(\mathbb{R}^d)$ of Lévy measures with finite second moments.

Corollary 2.34. *Let $\mathcal{W}_\Lambda := (\mathcal{C}_2^\Lambda)^{1/2}$, where $\mathcal{C}_2^\Lambda(\mu, \nu)$, $\mu, \nu \in \Lambda_2(\mathbb{R}^d)$, is the Lévy optimal transport cost defined in (3.8). Then $\mathcal{W}_\Lambda$ is a complete and separable metric on $\Lambda_2(\mathbb{R}^d)$.*

Finally, we state our last theorem in this subsection, which establishes an L^2 maximal-type estimate on the pathwise discrepancy between two Lévy processes under an optimal coupling of their generators.

Theorem 2.35. *Let $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$, and let $\mathcal{J}_* \in \Gamma(\mathcal{A}, \mathcal{B})$ be a 2-optimal Lévy coupling of $\mathcal{A}$ and $\mathcal{B}$. Denote by $\{Z_t = (X_t, Y_t)\}_{t \geq 0}$ the Lévy process on $\mathbb{R}^{2d}$ with generator $\mathcal{J}_*$. It holds for all $T \geq 0$:*

$$\mathbb{E}^{(0,0)} \left[\sup_{t \in [0, T]} |X_t - Y_t|^2 \right] \leq 8 \max\{T, T^2\} \mathcal{W}_G(\mathcal{A}, \mathcal{B})^2.$$

If, in addition, both processes have zero mean (i.e. $m_{\mathcal{A}} = m_{\mathcal{B}} = 0$), then the sharper bound holds:

$$\mathbb{E}^{(0,0)} \left[\sup_{t \in [0, T]} |X_t - Y_t|^2 \right] \leq 4T \mathcal{W}_G(\mathcal{A}, \mathcal{B})^2.$$

We remark that the metrics $\mathcal{W}_G$ and $\mathcal{W}_\Lambda$ were recently introduced in [36], while the latter already appeared earlier in [30]. In both works, these metrics were used to establish important results, such as propagation of chaos and the generation of Lévy-type generators, under Lipschitz continuity assumptions w.r.t. these metrics. Although it was suggested in [36] that $\mathcal{W}_G$ and $\mathcal{W}_\Lambda$ indeed define valid metrics, a formal proof was not given. Theorem 2.33 completes this picture by rigorously verifying their metric properties.

We emphasize that Theorem 2.33 and Corollary 2.34 provide the first rigorous construction of a transport-type metric on the space of Lévy generators $\mathcal{G}_2^\Lambda(\mathbb{R}^d)$ and the space of Lévy measures $\Lambda_2(\mathbb{R}^d)$. While, as mentioned earlier, Wasserstein-type metrics on Lévy measures have appeared earlier in the literature, our work further establishes their fundamental metric properties. This advancement opens new avenues for quantitative analysis in stochastic processes, including stability estimates, propagation of chaos for Lévy-type mean-field systems, and potential applications in optimal transport and probabilistic numerics.

Further structural and topological properties of the space $\mathcal{G}_2^\Lambda(\mathbb{R}^d)$ under the metric $\mathcal{W}_G$ and the space $\Lambda_2(\mathbb{R}^d)$ under the metric $\mathcal{W}_\Lambda$, including compactness criteria and characterization of convergence—will be developed in Section 6.

2.9. Discussion, methods of proof, and organization. Let us highlight our contributions in this work. As mentioned earlier, the optimal Markovian coupling problem for diffusion processes, including the drift–diffusion generator considered in Theorem 2.28, was established by [16]. In fact, as we will see later, in this case the optimal coupling can be explicitly constructed using the results of [23] on the optimal coupling between Gaussian measures. On the other hand, the case of bounded jump operators has been settled by Zhang [49]. Consequently, our main contribution lies in addressing the jump component for generators with *unbounded* Lévy measures, where previous techniques do not directly apply.

Let us next discuss the difference in our proof strategy compared to Zhang [49]. There, bounded jump operators are considered, so any coupling $\mathcal{J}$ is also bounded and can be

represented via a jump kernel $\lambda(x, y, dx', dy')$:

$$(\mathcal{J}\Phi)(x, y) = \int_{\Pi^2} [\Phi(x', y') - \Phi(x, y)] d\lambda(x, y, dx', dy').$$

Zhang constructed the jump kernel corresponding to an optimal Markovian coupling by taking the pointwise (in $(x, y) \in \Pi$) limit along a subsequence as $t \searrow 0$:

$$\lambda(x, y, dx', dy') = \lim_{t \searrow 0} \frac{1}{t} [\gamma_t - \delta_{(x, y)}],$$

where γ_t is a c -optimal coupling between the laws $\delta_x e^{tA}$ and $\delta_y e^{tB}$. The existence of this limit along a subsequence follows from the compactness of measures. A key technical challenge is establishing the measurability of the kernel $\lambda(x, y)$ with respect to (x, y) , which is achieved using a measurable selection theorem.

Our construction strategy is to tackle the optimal Markovian coupling problem for Lévy generators by considering each component—drift, diffusion, and jump—separately, and then combining the results using the additivity property of the problem and the strong duality (to be established later). The main challenge, of course, lies in constructing the optimal coupling for the jump components. While the pointwise limit approach used by Zhang [49] could potentially be adapted to this setting, we adopt a different strategy. Specifically, we construct an optimal coupling by solving an *optimal transport problem between Lévy measures*, analogous to the classical optimal transport problem for probability measures. Given two Lévy measures μ, ν on $\mathbb{R}^d$, we seek a coupling (in an appropriate sense) γ of μ and ν , which is itself a Lévy measure on $\mathbb{R}^{2d}$, that minimizes the transport cost between these two measures. To the best of our knowledge, this is the first time an optimal transport problem has been studied systematically for Lévy measures. We refer to this as the *Lévy optimal transport problem*. Once a minimizer is identified, it directly yields the optimal Markovian jump operator for the pair of processes.

Let us remark on the connection between the optimal Markovian coupling problem and the optimality gap problem. At first sight these may appear to be independent questions, but in fact they are closely intertwined. Specifically, once the optimal coupling generators have been constructed for each component — drift, diffusion, and jump — it is in fact straightforward to verify that their superposition yields an optimal coupling, though only within the class of *Lévy* (or even Lévy-type) coupling generators. To extend this optimality to the larger class of all Feller coupling generators, or even Markovian coupling semigroups, one needs to appeal to the duality: the dual problem provides the verification principle that confirms the superposition is indeed minimal.

Future directions. Let us also outline some possible future directions of this research:

- (1) **General** $p \in [1, \infty)$. The optimal Markovian coupling problem for Lévy generators can be formulated for the general p -cost $c_p(x, y) = \frac{1}{p}|x - y|^p$. In this work, we focus on the case $p = 2$, primarily because of the self-duality of $\frac{1}{2}|x|^2$. This property simplifies certain arguments (e.g., Proposition 4.3), since the Legendre transform of $\frac{1}{2}|x|^2$ is itself. For general p , this convenience is lost, and the problem may become

substantially more challenging. For instance, an optimal Markovian coupling between $\mathcal{A}, \mathcal{B}$, if exists, may not be translation-invariant (i.e., a Lévy generator).

- (2) **General Lévy-type generators.** One may pose the optimal coupling problem for a pair of Lévy-type generators, i.e., operators of the form (2.5) where the triplet (κ, α, μ) depends on $x \in \mathbb{R}^d$. A Lévy-type generator with bounded second moment can be viewed as a family of Lévy generators $\{\mathcal{A}(x)\}_x$ with $\mathcal{A}(x) \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$. A natural candidate for an optimal coupling is then $\{\mathcal{J}_*(x, y)\}_{x, y}$, where $\mathcal{J}_*(x, y)$ is an optimal coupling of $\mathcal{A}(x)$ and $\mathcal{B}(y)$ guaranteed by Theorem 2.28. However, several challenges remain: the optimal coupling may not be unique, which can create measurability issues, and the family $\{\mathcal{J}_*(x, y)\}_{x, y}$ may not generate a Feller semigroup. These problems warrant further investigation.
- (3) **Applications of the metric $\mathcal{W}_G$.** Given the Wasserstein generator metric $\mathcal{W}_G$ defined in (2.25), one can explore a variety of potential applications, as illustrated in [30, 36]. For instance, consider a Lévy-type generator $\{\mathcal{A}(x)\}_x$ that is Lipschitz with respect to $\mathcal{W}_G$, meaning that

$$\mathcal{W}_G(\mathcal{A}(x), \mathcal{A}(y)) \leq C|x - y| \quad \text{for some constant } C \geq 0.$$

A natural question arising from this observation is whether such a Lipschitz condition guarantees that $\mathcal{A}$ indeed generates a Feller semigroup.

- (4) **Optimal Markovian coupling problem for generalized couplings.** In the present work, the optimal coupling is sought within the class of *Feller generators/processes*. In more general settings, however, the Feller property may be too restrictive. It is therefore natural to enlarge the framework and search for optimal couplings within broader notions, such as general Markov semigroups, Markov processes, or solutions to martingale problems. This calls for a suitably generalized formulation of the coupling problem; see Remark 2.16.

Organization. The remainder of this article is organized as follows. Section 3 introduces and solves the Lévy optimal transport problem between Lévy measures. The development closely parallels the classical theory of optimal transport, including the existence of minimizers (optimal couplings), fundamental structural properties, and a Kantorovich-type duality. Section 4 addresses the optimal Markovian coupling problem and the associated optimality gap problem for each component of a Lévy generator — drift, diffusion, and jump — as an intermediate step. Building on this, Section 5 establishes the additivity property, which allows us to combine the componentwise results and thereby prove the main theorems stated in Section 2.7 (Theorems 2.28, 2.31, and 2.32). Section 6 is devoted to the study of the Wasserstein-type metrics $\mathcal{W}_G, \mathcal{W}_\Lambda$, where we establish Theorem 2.33, 2.35, Corollary 2.34, and further investigate their topological properties. Finally, the appendix collects the proofs of several technical lemmas used throughout Sections 3–6.

3. LÉVY OPTIMAL TRANSPORT PROBLEM

In this section, we formulate and solve a variation of the classical optimal transport problem between two Lévy measures μ and ν , including cases where μ and ν have different, or even

infinite, total mass. We refer to this formulation as the *Lévy optimal transport problem*. As will be seen in the next section, this problem is closely related to the optimal coupling problem of pure jump Lévy processes.

We begin this section with a brief overview of the classical optimal transport problem on $\mathbb{R}^d$ with quadratic cost. The Lévy optimal transport problem is then introduced in the subsequent subsection, along with motivating examples and preliminary results. In the final two subsections, we establish the existence of minimizers for the Lévy transport problem and develop its dual formulation, drawing parallels with the classical theory.

In the proof, to simplify notation for computations we set $c_2(x, y) := \frac{1}{2}|x - y|^2$ for the squared cost, and use the natural pairing

$$\langle \mu, f \rangle := \int f d\mu,$$

where the integral is taken over the space ($\mathbb{R}^d$ or $\mathbb{R}^{2d}$) on which the (Lévy) measure μ is defined. This notation can also be extended to certain unbounded functions f provided that the integral is well-defined.

3.1. Classical optimal transport problem. Let us begin our discussion with the *classical optimal transport problem* on $\mathbb{R}^d$ with respect to the squared cost.

Definition 3.1 (Coupling of measures and 2-optimality). Let $\mu, \nu \in \mathcal{M}(\mathbb{R}^d)$ be a pair of bounded measures with equal mass: $\mu(\mathbb{R}^d) = \nu(\mathbb{R}^d)$.

(i) A *coupling measure* is a (bounded) measure $\gamma \in \mathcal{M}(\mathbb{R}^{2d})$ such that it holds for all measurable $E \subset \mathbb{R}^d$

$$\gamma(E \times \mathbb{R}^d) = \mu(E), \quad \gamma(\mathbb{R}^d \times E) = \nu(E).$$

Denote $\Gamma(\mu, \nu)$ the family of all coupling.

(ii) The *2-optimal transport cost* between μ, ν is defined by

$$\mathcal{C}_2(\mu, \nu) := \inf_{\gamma \in \Gamma(\mu, \nu)} \langle \gamma, c_2 \rangle = \inf_{\gamma \in \Gamma(\mu, \nu)} \int_{\mathbb{R}^{2d}} \frac{1}{2} |x - y|^2 d\gamma(x, y).$$

(iii) A *2-optimal coupling* of the pair μ, ν is a coupling measure $\gamma_* \in \Gamma(\mu, \nu)$ such that

$$\langle \gamma_*, c_2 \rangle = \int_{\mathbb{R}^{2d}} \frac{1}{2} |x - y|^2 d\gamma_*(x, y) = \mathcal{C}_2(\mu, \nu).$$

(iv) A bounded measure $\gamma \in \mathcal{M}(\mathbb{R}^{2d})$ is *2-optimal* if it is 2-optimal w.r.t. its marginal.

Remark 3.2. In the literature, (e.g., [2, 46]), the optimal transport problem is typically studied in the setting of *probability measures*, that is, when $\mu(\mathbb{R}^d) = \nu(\mathbb{R}^d) = 1$. The problem and many of the associated results to be presented later, e.g., existence of minimizer, duality, can be naturally extended to the more general setting of bounded measures with equal mass, i.e., $\mu(\mathbb{R}^d) = \nu(\mathbb{R}^d) < \infty$.

In the classical theory of optimal transport, the fundamental theorem of optimal transport establishes a deep connection between optimal coupling measures, cyclical monotonicity, and Kantorovich duality. The following definition and theorem formalize this relationship. While

they can be stated in the general setting of a Polish space Π and cost c , we restrict our attention to the case $\Pi = \mathbb{R}^d$ and $c(x, y) = c_2(x, y) = \frac{1}{2}|x - y|^2$.

Definition 3.3 (2-cyclical monotonicity). A measurable set $E \subset \mathbb{R}^d \times \mathbb{R}^d$ is *2-cyclically monotone* if for any finite sequences $\{(x_k, y_k)\}_{k=1}^N \subset E$, and any permutations σ on $\{1, 2, \dots, N\}$, it holds

$$\sum_{k=1}^N \frac{1}{2}|x_k - y_k|^2 \leq \sum_{k=1}^N \frac{1}{2}|x_k - y_{\sigma(k)}|^2.$$

Theorem 3.4 (The fundamental theorem of optimal transport [2]). *Let $E \subset \mathbb{R}^{2d}$ be measurable. The following are equivalent:*

- (i) *E is 2-cyclically monotone;*
- (ii) *there are lower semicontinuous functions $\varphi, \psi : \mathbb{R}^d \rightarrow [-\infty, \infty)$ with $\varphi \oplus \psi \leq c_2$ such that*

$$\varphi(x) + \psi(y) = c_2(x, y), \quad \text{for all } (x, y) \in E;$$

- (iii) *for any bounded measures γ with $\text{supp}(\gamma) \subset E$, γ is classically 2-optimal.*

3.2. Lévy optimal transport problem. Let $\mu, \nu \in \Lambda(\mathbb{R}^d)$ be two Lévy measures. In this work, we are interested in a *coupling* Lévy measure $\gamma \in \Lambda(\mathbb{R}^{2d})$ in the following sense: if $\{X_t\}$ and $\{Y_t\}$ are pure jump Lévy processes on $\mathbb{R}^d$ with Lévy measures μ and ν , respectively, and $\{Z_t\}$ is a pure jump Lévy process on $\mathbb{R}^{2d}$ with Lévy measure γ , then $\{Z_t\}$ is a *Feller coupling process* of $\{X_t\}$ and $\{Y_t\}$. This motivates the following definition.

Definition 3.5 (Coupling of Lévy measures). Let $\mu, \nu \in \Lambda(\mathbb{R}^d)$ be two Lévy measures on $\mathbb{R}^d$. A *Lévy coupling* between μ, ν is a Lévy measure γ on $\mathbb{R}^{2d}$ such that for all measurable sets $E \subset \mathbb{R}^d$ that *does not contain a neighbourhood of 0*, it holds

$$\gamma(E \times \mathbb{R}^d) = \mu(E), \quad \gamma(\mathbb{R}^d \times E) = \nu(E). \quad (3.1)$$

We denote the family of all Lévy couplings of μ, ν as $\Gamma^\Lambda(\mu, \nu)$

Remark 3.6. Unlike the classical coupling of probability measures, the marginal condition (3.1) for Lévy couplings is only required to hold on measurable sets that *do not* contain a neighborhood of the origin. At first glance, this restriction may seem unusual, but it is the natural choice in the context of Lévy measures, see Proposition 3.12 below.

Remark 3.7. For $\mu, \nu \in \Lambda(\mathbb{R}^d)$, the set $\Gamma^\Lambda(\mu, \nu)$ is nonempty. For instance, let $\gamma \in \Lambda(\mathbb{R}^{2d})$ be defined by

$$\gamma = \mu \otimes \delta_0 + \delta_0 \otimes \nu,$$

where $\otimes$ denotes the product (tensor) of measures. In particular, γ is a Lévy measure supported on $\mathbb{R}^d \times \{0\} \cup \{0\} \times \mathbb{R}^d$. One can readily verify that γ is a Lévy coupling of μ and ν . Indeed, if μ and ν are the Lévy measures associated with processes $\{X_t\}$ and $\{Y_t\}$, respectively, then γ corresponds to the Lévy measure of the process $\{Z_t\} = \{(X_t, Y_t)\}$, assuming that $\{X_t\}$ and $\{Y_t\}$ are independent.

In the coming proposition and lemma, let us provide some equivalent conditions for Lévy couplings of μ, ν .

Proposition 3.8. *Let $\mu, \nu \in \Lambda(\mathbb{R}^d)$ and $\gamma \in \Lambda(\mathbb{R}^{2d})$. The following are equivalent:*

- (i) $\gamma \in \Gamma^\Lambda(\mu, \nu)$;
- (ii) *for all bounded measurable functions f that vanish on a neighbourhood of 0, we have*

$$\int_{\mathbb{R}^{2d}} f(x) d\gamma(x, y) = \int_{\mathbb{R}^d} f(x) d\mu(x), \quad \int_{\mathbb{R}^{2d}} f(y) d\gamma(x, y) = \int_{\mathbb{R}^d} f(y) d\nu(x); \quad (3.2)$$

- (iii) *for all bounded twice differentiable functions $f \in C_b^2(\mathbb{R}^d)$ that vanish on a neighbourhood of 0, (3.2) holds;*

- (iv) *for all bounded continuous functions $f \in C_b(\mathbb{R}^d)$ that vanish on a neighbourhood of 0, (3.2) holds;*

- (v) *for all bounded continuous functions $f \in C_b(\mathbb{R}^d)$ such that $|f(x)| \leq C|x|^2$ for some $C \geq 0$, (3.2) holds.*

Remark 3.9. It is instructive to contrast this with the classical setting of probability measure couplings. In that case, the equivalence stated in Proposition 3.8 holds without the restriction that the functions vanish on a neighborhood of the origin.

Remark 3.10. Using the natural pairing notation, the marginal integral condition (3.2) can be rewritten in the more compact form:

$$\langle \gamma, f \otimes 1 \rangle = \langle \mu, f \rangle, \quad \langle \gamma, 1 \otimes f \rangle = \langle \nu, f \rangle. \quad (3.3)$$

Proof. (i) $\rightarrow$ (ii). It is straightforward from Definition 3.5 that (3.2) holds for all characteristic functions f of sets that does not contain a neighborhood of 0. One may then approximate any bounded measurable functions vanishing around 0 as a sequence of simple functions. The dominated convergence theorem implies (3.2) holds for all such functions.

(ii) $\rightarrow$ (iii). Trivial.

(iii) $\rightarrow$ (iv). Let $f \in C_b(\mathbb{R}^d)$ vanish on $B_r(0)$ for some $r > 0$. Then there exists a sequence $\{f_n\}_n \subset C_b^2(\mathbb{R}^d)$, each vanishing on a smaller ball (e.g., $B_{r/2}(0)$), such that $f_n \rightarrow f$ locally uniformly. Since (3.2) holds for all f_n , letting $n \rightarrow \infty$ and applying the bounded convergence theorem yields the desired equality for f .

(iv) $\rightarrow$ (v). Given a bounded continuous functions $f \in C_b(\mathbb{R}^d)$ such that $|f(x)| \leq C|x|^2$ for some $C \geq 0$, we may define a sequence of bounded continuous functions $\{f_n\}$ such that f_n vanishes in $B_{1/n}(0)$, the sequence converges pointwise to f , and $|f_n| \leq C|x|^2$. Since (3.2) holds for all f_n and $\min\{1, |x|^2\} \in L^1(\mu), L^1(\nu)$, invoking the dominated convergence theorem, (3.2) holds for f as well.

(v) $\rightarrow$ (i). Given an open set $E \subset \mathbb{R}^d \setminus B_\varepsilon(0)$ for some $\varepsilon > 0$, by Urysohn lemma we may find a sequence of continuous functions $\{f_n\}$ vanishing on $B_{\varepsilon/2}(0)$ such that $f_n \searrow \chi_E$ pointwise. We note that for each n , f_n satisfies Condition (iii) and thus (3.2) holds. Applying the dominated convergence theorem, passing $n \rightarrow \infty$, (3.2) holds for $f = \chi_E$, that is, (3.1) holds for all open sets that does not contain a neighborhood of origin. Applying a standard argument involving π - λ theorem, (3.1) extends to all measurable sets E that does not contain a neighborhood of origin. This proves (i). $\square$

Let us record the following simple lemma, which will be useful later.

Lemma 3.11. *Suppose $\gamma \in \Gamma^\Lambda(\mu, \nu)$ and let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be measurable such that $f^+ \in L^1(\mathbb{R}^d, d\mu)$. Then $f^+ \otimes 1 \in L^1(\mathbb{R}^{2d}, d\gamma)$ and*

$$\int_{\mathbb{R}^{2d}} f(x) d\gamma(x, y) = \int_{\mathbb{R}^d} f(x) d\mu(x). \quad (3.4)$$

The integrals above are understood to take values in $[-\infty, \infty)$.

Proof. For $\varepsilon > 0$, define the cutoff functions $f_\varepsilon := f \chi_{B_\varepsilon^c(0)}$. We then have

$$f_\varepsilon^\pm = \max\{\pm f, 0\} = f^\pm \chi_{B_\varepsilon^c(0)},$$

and the sequences $\{f_\varepsilon^\pm\}$ increase pointwise to $f^\pm$ as $\varepsilon \nearrow 0$.

Since $\gamma \in \Gamma^\Lambda(\mu, \nu)$, by Proposition 3.8, (3.4) holds with $f_\varepsilon^\pm$ in place of f . By the monotone convergence theorem, (3.4) also holds with $f^\pm$ in place of f . In particular, since $f^+ \in L^1(\mu)$, the integral of f^+ with respect to γ is finite, i.e., $f^+ \otimes 1 \in L^1(\gamma)$. Finally, since $f^+ - f^- = f$, (3.4) follows by subtracting the integrals of positive and negative parts. $\square$

As mentioned earlier, the motivation behind the above definition is to relate a coupling of pure jump Lévy processes with their corresponding Lévy measures. The proposition below establishes this connection.

Proposition 3.12. *Let $\mu, \nu \in \Lambda_2(\mathbb{R}^d)$, $\gamma \in \Lambda_2(\mathbb{R}^{2d})$, $\mathcal{A} = \Lambda_{\text{glob}}(0, 0, \mu)$, $\mathcal{B} = \Lambda_{\text{glob}}(0, 0, \nu)$ and $\mathcal{J} = \Lambda_{\text{glob}}(0, 0, \gamma)$. Then $\mathcal{J} \in \Gamma(\mathcal{A}, \mathcal{B})$ if and only if $\gamma \in \Gamma^\Lambda(\mu, \nu)$.*

Remark. The assumption that $\mathcal{A}, \mathcal{B}$ are given in the *global form* is essential for this proposition. In particular, the identification $\mathcal{J} = \Lambda_{\text{glob}}(0, 0, \gamma)$ relies on the compensated jump operator being expressed without a cutoff function. In the localized form (e.g., with a cutoff such as $\chi_{B_1(0)}$), the generator depends on the choice of cutoff, and the marginal condition would no longer correspond directly to $\gamma \in \Gamma^\Lambda(\mu, \nu)$.

Proof. Before the proof, let us introduce some notations to simplify the computation. Fix $f \in C^2(\mathbb{R}^d)$ and $x_0 \in \mathbb{R}^d$. Write the (second-order) Taylor remainder in the direction x as

$$\mathcal{R}_{x_0} f(x) := f(x_0 + x) - f(x_0) - x \cdot \nabla f(x_0).$$

For $F \in C^2(\mathbb{R}^{2d})$ and $(x_0, y_0) \in \mathbb{R}^{2d}$ define analogously

$$\mathcal{R}_{(x_0, y_0)} F(x, y) := F(x_0 + x, y_0 + y) - F(x_0, y_0) - (x, y)^\top \cdot \nabla F(x_0, y_0).$$

With this notation, the jump-type Lévy generators act as

$$\mathcal{A}f(x_0) = \int_{\mathbb{R}^d} \mathcal{R}_{x_0} f(x) \mu(dx) = \langle \mu, \mathcal{R}_{x_0} f \rangle, \quad \mathcal{B}f(y_0) = \int_{\mathbb{R}^d} \mathcal{R}_{y_0} f(y) \nu(dy) = \langle \nu, \mathcal{R}_{y_0} f \rangle, \quad (3.5)$$

and for $\mathcal{J}$,

$$\mathcal{J}F(x_0, y_0) = \int_{\mathbb{R}^{2d}} \mathcal{R}_{(x_0, y_0)} F(x, y) \gamma(dx, dy) = \langle \gamma, \mathcal{R}_{(x_0, y_0)} F \rangle. \quad (3.6)$$

The remainder operator tensorizes in the following way:

$$\mathcal{R}_{(x_0, y_0)}[f \otimes 1] = (\mathcal{R}_{x_0} f) \otimes 1, \quad \mathcal{R}_{(x_0, y_0)}[1 \otimes f] = 1 \otimes (\mathcal{R}_{y_0} f). \quad (3.7)$$

→. Suppose $\mathcal{J} \in \Gamma(\mathcal{A}, \mathcal{B})$. This means that for any $f \in C_b^2(\mathbb{R}^d)$ we have

$$\mathcal{J}[f \otimes 1] = \mathcal{A}f \otimes 1, \quad \mathcal{J}[1 \otimes f] = 1 \otimes \mathcal{B}f.$$

Fix $f \in C_b^2(\mathbb{R}^d)$ vanishing on a neighbourhood of 0. Then $f(0) = 0$ and $\nabla f(0) = 0$, hence $\mathcal{R}_0 f = f$ and by (3.7) $\mathcal{R}_{(0,0)}[f \otimes 1] = f \otimes 1$. Combining these with (3.5), (3.6), we find

$$\langle \gamma, f \otimes 1 \rangle = \langle \gamma, \mathcal{R}_{(0,0)}[f \otimes 1] \rangle = \mathcal{J}[f \otimes 1](0, 0) = \mathcal{A}f(0) = \langle \mu, \mathcal{R}_0 f \rangle = \langle \mu, f \rangle.$$

Similarly, we have $\langle \gamma, 1 \otimes f \rangle = \langle \nu, f \rangle$. By Proposition 3.8 (see (3.3)), we conclude that $\gamma \in \Gamma^\Lambda(\mu, \nu)$.

←. Suppose $\gamma \in \Gamma^\Lambda(\mu, \nu)$. Let $f \in C_0^2(\Pi)$ and fix $x_0, y_0 \in \mathbb{R}^d$. By Taylor's theorem, the remainder $\mathcal{R}_{x_0} f$ satisfies $|\mathcal{R}_{x_0} f(x)| \leq C|x|^2$ for some $C \geq 0$ depending on the C^2 -norm of f . Invoking Proposition 3.8 and using (3.5)–(3.7), we obtain

$$\mathcal{J}[f \otimes 1](x_0, y_0) = \langle \gamma, \mathcal{R}_{(x_0, y_0)}(f \otimes 1) \rangle = \langle \mu, \mathcal{R}_{x_0} f \rangle = \mathcal{A}f(x_0).$$

By a similar argument, $\mathcal{J}[1 \otimes f](x_0, y_0) = \mathcal{B}f(y_0)$, and hence by Proposition 2.12 $\mathcal{J} \in \Gamma(\mathcal{A}, \mathcal{B})$. Note we use the fact that $C_0^2(\mathbb{R}^d)$ is dense in $D(\mathcal{A}), D(\mathcal{B})$ respectively. $\square$

We now formulate the optimality condition for Lévy couplings with respect to the squared cost, interpreted in the sense of Lévy transport.

Definition 3.13 (2-Optimal Lévy transport cost and couplings). Let $\mu, \nu \in \Lambda_2(\mathbb{R}^d)$ be two Lévy measures on $\mathbb{R}^d$ with finite second moments. The *2-optimal Lévy transport cost* is defined by

$$\mathcal{C}_2^\Lambda(\mu, \nu) := \inf_{\gamma \in \Gamma^\Lambda(\mu, \nu)} \int_{\mathbb{R}^{2d}} \frac{1}{2} |x - y|^2 d\gamma(x, y). \quad (3.8)$$

A Lévy measure $\gamma_* \in \Lambda(\mathbb{R}^{2d})$ is called a *2-optimal Lévy coupling* of μ and ν if

$$\int_{\mathbb{R}^{2d}} \frac{1}{2} |x - y|^2 d\gamma_*(x, y) = \mathcal{C}_2^\Lambda(\mu, \nu). \quad (3.9)$$

As seen in the definition above, the objective functional (3.8) coincides formally with that of the classical optimal transport problem. However, the key distinction lies in the admissible set of couplings. The classical formulation considers all couplings between probability measures of equal mass, while the Lévy formulation restricts to *Lévy couplings* between possibly unequal Lévy measures. This change in admissibility conditions fundamentally alters the structure and interpretation of the problem.

This formulation will later be shown to coincide with the optimal coupling problem for Lévy jump processes under suitable conditions, a connection that will be explored in the next section. In that context, the Lévy measure associated with the optimal Markovian coupling of two jump processes with Lévy measures μ and ν corresponds precisely to a 2-optimal Lévy coupling.

We remark that several variants of optimal transport, in which the cost functional is optimized under alternative constraints, have also been extensively studied. Notable examples include unbalanced optimal transport [17], partial optimal transport [10, 20], and martingale optimal transport [3, 5], among others.

Convention 3.14. Throughout this section, we will use the terms *classical* and *Lévy* to distinguish between the two parallel frameworks of optimal transport: the classical formulation (Definition 3.1) and the Lévy-based formulation (Definition 3.13). Accordingly, we will refer to notions such as the *classical* optimal transport problem, *classical* 2-optimal coupling, and, in contrast, the *Lévy* optimal transport problem and *Lévy* 2-optimal coupling. We encourage the reader to be mindful of this distinction, as many definitions and results will appear formally similar but belong to distinct theoretical settings.

3.3. Some preliminary results of the Lévy optimal transport cost. Before we proceed with solving the Lévy optimal transport problem, we collect some elementary properties of the transport cost.

Proposition 3.15. *Let $\mu, \nu, \mu', \nu' \in \Lambda(\mathbb{R}^d)$, and let $\mathcal{C}_2^\Lambda$ be the Lévy transport cost defined in (3.8). Then the following properties hold:*

- (i) $\mathcal{C}_2^\Lambda(\mu, \nu) \in [0, \infty]$.
- (ii) $\mathcal{C}_2^\Lambda(\mu, \nu) = \mathcal{C}_2^\Lambda(\nu, \mu)$.
- (iii) $\mathcal{C}_2^\Lambda(\mu, \mu) = 0$.
- (iv) If $\mu, \nu \in \Lambda_2(\mathbb{R}^d)$, then $\mathcal{C}_2^\Lambda(\mu, \nu) < \infty$.
- (v) $\mathcal{C}_2^\Lambda(\mu, 0) = \frac{1}{2} \int_{\mathbb{R}^d} |x|^2 d\mu(x)$.
- (vi) $\mathcal{C}_2^\Lambda(\mu + \mu', \nu + \nu') \leq \mathcal{C}_2^\Lambda(\mu, \nu) + \mathcal{C}_2^\Lambda(\mu', \nu')$.
- (vii) For any $\alpha \geq 0$, $\mathcal{C}_2^\Lambda(\alpha\mu, \alpha\nu) = \alpha \mathcal{C}_2^\Lambda(\mu, \nu)$.

Proof. Recall $c_2(x, y) = \frac{1}{2}|x - y|^2$ and the notation $\langle \cdot, \cdot \rangle$ of natural pairings between Lévy measures and functions.

- (i) The nonnegativity $\mathcal{C}_2^\Lambda(\mu, \nu) \geq 0$ is immediate from the definition.
- (ii) Let $T : (\mathbb{R}^d)^2 \rightarrow (\mathbb{R}^d)^2$ be the coordinate swap map. Then $\gamma \in \Gamma^\Lambda(\mu, \nu)$ if and only if $T_\# \gamma \in \Gamma^\Lambda(\nu, \mu)$, which gives symmetry.
- (iii) Nonnegativity follows from (i). The diagonally supported measure $\gamma_*(E) := \mu(\{x : (x, x) \in E\})$ lies in $\Gamma^\Lambda(\mu, \mu)$ and satisfies $\langle \gamma_*, c_2 \rangle = 0$, hence $\mathcal{C}_2^\Lambda(\mu, \mu) = 0$.
- (iv) Using $c_2(x, y) \leq |x|^2 + |y|^2$, for any $\gamma \in \Gamma^\Lambda(\mu, \nu)$ (non-empty by Remark 3.7),

$$\langle \gamma, c_2 \rangle \leq \int_{\mathbb{R}^{2d}} (|x|^2 + |y|^2) d\gamma(x, y) = \int_{\mathbb{R}^d} |x|^2 d\mu(x) + \int_{\mathbb{R}^d} |y|^2 d\nu(y) < \infty.$$

This follows $\mathcal{C}_2^\Lambda(\mu, \nu) < \infty$.

(v) There is only one Lévy coupling between μ and 0, given by $\gamma = \mu \otimes \delta_0$, supported on $\mathbb{R}^d \times \{0\}$.

(vi) Fix $\epsilon > 0$ and choose $\gamma \in \Gamma^\Lambda(\mu, \nu)$, $\gamma' \in \Gamma^\Lambda(\mu', \nu')$ such that $\langle \gamma, c_2 \rangle \leq \mathcal{C}_2^\Lambda(\mu, \nu) + \epsilon$, $\langle \gamma', c_2 \rangle \leq \mathcal{C}_2^\Lambda(\mu', \nu') + \epsilon$. Then $\gamma + \gamma' \in \Gamma^\Lambda(\mu + \mu', \nu + \nu')$, and by linearity,

$$\mathcal{C}_2^\Lambda(\mu + \mu', \nu + \nu') \leq \langle \gamma + \gamma', c_2 \rangle = \langle \gamma, c_2 \rangle + \langle \gamma', c_2 \rangle \leq \mathcal{C}_2^\Lambda(\mu, \nu) + \mathcal{C}_2^\Lambda(\mu', \nu') + 2\epsilon.$$

Letting $\epsilon \rightarrow 0$ gives the desired inequality.

(vii) The identity trivially holds for $\alpha = 0$. For $\alpha > 0$, note that $\gamma \in \Gamma^\Lambda(\mu, \nu)$ if and only if $\alpha\gamma \in \Gamma^\Lambda(\alpha\mu, \alpha\nu)$, which yields the identity. $\square$

The next result provides a comparison between the classical and Lévy transport costs, under the assumption that the two Lévy measures have equal (finite) total mass.

Proposition 3.16. *Let $\mu, \nu \in \Lambda(\mathbb{R}^d)$ satisfy $\mu(\mathbb{R}^d) = \nu(\mathbb{R}^d) < \infty$. Then*

$$\mathcal{C}_2^\Lambda(\mu, \nu) \leq \mathcal{C}_2(\mu, \nu).$$

Proof. Every classical coupling is also a Lévy coupling, i.e., $\Gamma(\mu, \nu) \subset \Gamma^\Lambda(\mu, \nu)$. The inequality follows immediately from the definitions of $\mathcal{C}_2$ and $\mathcal{C}_2^\Lambda$. $\square$

An immediate question naturally arises: if two Lévy measures have equal total mass, does the Lévy transport cost necessarily coincide with the classical optimal transport cost? Surprisingly, the answer is *no*. The following example illustrates this phenomenon. Later, the fundamental theorem of Lévy optimal transport (Theorem 3.20) will provide a necessary and sufficient condition under which the two costs indeed coincide, see Remark 3.21.

Example 3.17. Consider a probability measure $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ with finite second moment, and assume the first moment is nonzero:

$$m_1(\mu) := \int_{\mathbb{R}^d} x d\mu(x) \neq 0, \quad m_2(\mu) := \int_{\mathbb{R}^d} |x|^2 d\mu(x).$$

Fix $h \in \mathbb{R}^d$ and let μ_h be the h -translation of μ , namely, $\mu_h(E) = \mu(E + h)$ for all measurable E . It is well-known here that the classical optimal transport cost is given by $\mathcal{C}_2(\mu, \mu_h) = \frac{1}{2}|h|^2$. We now show that the Lévy transport cost $\mathcal{C}_2^\Lambda(\mu, \mu_h)$ is strictly less than $\mathcal{C}_2(\mu, \mu_h)$ for some $h \in \mathbb{R}^d$.

Let γ be a classical 2-optimal coupling of μ and μ_h , so that $\langle \gamma, c_2 \rangle = \mathcal{C}_2(\mu, \mu_h) = \frac{1}{2}|h|^2$. For $\theta \in (0, 1]$ we consider the following Lévy coupling of μ and μ_h :

$$\tilde{\gamma} = (1 - \theta)\gamma + \theta(\mu \otimes \delta_0 + \delta_0 \otimes \mu_h).$$

That is, $\tilde{\gamma}$ is a convex combination of a classical, and hence Lévy coupling γ of μ, ν and their trivial Lévy coupling (see Remark 3.7). Hence it is also a Lévy coupling of μ, ν . We then compute the following integral:

$$\begin{aligned} \langle \tilde{\gamma}, c_2 \rangle &= (1 - \theta) \langle \gamma, c_2 \rangle + \theta \langle \mu \otimes \delta_0, c_2 \rangle + \theta \langle \delta_0 \otimes \mu_h, c_2 \rangle \\ &= (1 - \theta) \langle \gamma, c_2 \rangle + \frac{\theta}{2} \int_{\mathbb{R}^d} |x|^2 d\mu(x) + \frac{\theta}{2} \int_{\mathbb{R}^d} |y|^2 d\mu_h(y) \\ &= \frac{1 - \theta}{2} |h|^2 + \theta m_2(\mu) + \theta h \cdot m_1(\mu) + \frac{\theta |h|^2}{2} \\ &= \frac{|h|^2}{2} + \theta(m_2(\mu) + h \cdot m_1(\mu)), \end{aligned}$$

where we use the fact:

$$\int_{\mathbb{R}^d} |y|^2 d\mu_h(y) = \int_{\mathbb{R}^d} |y + h|^2 d\mu(y) = m_2(\mu) + 2h \cdot m_1(\mu) + h^2.$$

Since $m_1(\mu) \neq 0$, we may choose $h \in \mathbb{R}^d$ such that $m_2(\mu) + h \cdot m_1(\mu) < 0$. This provides a Lévy coupling with transport cost $< \frac{|h|^2}{2}$. Thus, $\mathcal{C}_2^\Lambda(\mu, \mu_h) < \mathcal{C}_2(\mu, \mu_h)$.

3.4. Construction of optimal Lévy couplings. The main theorem of this section is the following, which states the existence of a 2-optimal Lévy coupling for any pairs of Lévy measures $\mu, \nu \in \Lambda_2(\mathbb{R}^d)$ with finite second moment.

Theorem 3.18 (Existence of optimal Lévy couplings). *Given $\mu, \nu \in \Lambda_2(\mathbb{R}^d)$, a 2-optimal Lévy coupling γ_* exists. Moreover, $\gamma_* \in \Lambda_2(\mathbb{R}^{2d})$*

3.4.1. Weak convergence and tightness in the space of Lévy measures. The main idea behind the construction is to obtain a minimizer as the weak limit of a subsequence of minimizing Lévy couplings γ_n . A technical challenge arises from the fact that the sequence $\{\gamma_n\}_n$ may lack tightness in the classical sense due to potential singularity near the origin. This issue is well known in the literature and is typically handled by introducing a topology—comprising both convergence and compactness criteria—tailored to the integrability structure of Lévy measures.

Let us now briefly review the notion of weak convergence and tightness that are suitable for the space of Lévy measures. A more detailed treatment will be presented in Appendix C. We say that a sequence of Lévy measures $\{\mu_n\}_n \subset \Lambda(\mathbb{R}^d)$ converges *Lévy-weakly*, or simply *Λ -weakly*, to $\mu \in \Lambda(\mathbb{R}^d)$ if

$$\lim_{n \rightarrow \infty} \langle \mu_n, \varphi \rangle = \lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} \varphi(x) d\mu_n(x) = \int_{\mathbb{R}^d} \varphi(x) d\mu(x) = \langle \mu, \varphi \rangle.$$

for all bounded continuous test functions $\varphi \in C_b(\mathbb{R}^d)$ satisfying a quadratic bound for some constant $C \geq 0$:

$$|\varphi(x)| \leq C \min\{1, |x|^2\}.$$

Next, a family $\mathcal{M} \subset \Lambda(\mathbb{R}^d)$ of Lévy measures is said to be *Lévy-tight* (or simply *Λ -tight*) if it satisfies the following two conditions:

(i) Uniform moment bound:

$$\sup_{\mu \in \mathcal{M}} \int_{\mathbb{R}^d} \min\{1, |x|^2\} d\mu(x) < \infty;$$

(ii) Tightness away from the origin: for every $\varepsilon > 0$, there exists a compact set $K \subset \mathbb{R}^d \setminus \{0\}$ such that

$$\sup_{\mu \in \mathcal{M}} \int_{K^c} \min\{1, |x|^2\} d\mu(x) < \varepsilon.$$

In the construction of minimizers for the Lévy optimal transport problem, the following two foundational results concerning Λ -weak convergence and Λ -tightness are essential:

- *Prokhorov-type compactness:* Every Λ -tight sequence of Lévy measures admits a Λ -weakly convergent subsequence.
- *Portmanteau-type theorem:* Let $\mu_n \rightarrow \mu$ Λ -weakly. Then for every function $\varphi \in C_b(\mathbb{R}^d)$ satisfying $|\varphi(x)| \leq C \min\{1, |x|^2\}$, it holds that

$$\liminf_{n \rightarrow \infty} \langle \mu_n, \varphi \rangle = \liminf_{n \rightarrow \infty} \int \varphi d\mu_n \geq \int \varphi d\mu = \langle \mu, \varphi \rangle,$$

provided φ is lower semicontinuous; and similarly, the lim sup inequality holds for upper semicontinuous functions.

Though these results are standard and often used implicitly in the literature, we include detailed proofs in Appendix C for completeness.

3.4.2. Construction of minimizers. One of the key steps in the construction of a minimizer to the optimal transport problem is to establish the tightness of the family of couplings. Let us prove this in a slightly more general setting, whose proof is similar to that of the classical counterpart.

Lemma 3.19. *Let $\mathcal{M}, \mathcal{N} \subset \Lambda(\mathbb{R}^d)$ be two Λ -tight families, and define*

$$\Gamma^\Lambda(\mathcal{M}, \mathcal{N}) = \{\gamma \in \Lambda(\mathbb{R}^{2d}) : \gamma \in \Gamma^\Lambda(\mu, \nu), \mu \in \mathcal{M}, \nu \in \mathcal{N}\}.$$

That is, $\Gamma^\Lambda(\mathcal{M}, \mathcal{N})$ is the family of all Lévy couplings γ of $\mu \in \mathcal{M}$, $\nu \in \mathcal{N}$. Then $\Gamma(\mathcal{M}, \mathcal{N})$ is Λ -tight in $\Lambda(\mathbb{R}^{2d})$.

Proof. Let us abbreviate $\Gamma = \Gamma^\Lambda(\mathcal{M}, \mathcal{N})$ and introduce the minimum operator $a \wedge b = \min\{a, b\}$. We aim to show: (1) it holds:

$$\sup_{\gamma \in \Gamma} \int_{\mathbb{R}^{2d}} [1 \wedge (|x|^2 + |y|^2)] d\gamma(x, y) < \infty;$$

(2) for every $\varepsilon > 0$, there exists a compact set $K \subset \mathbb{R}^{2d} \setminus \{(0, 0)\}$ such that

$$\sup_{\gamma \in \Gamma} \int_{K^c} [1 \wedge (|x|^2 + |y|^2)] d\gamma(x, y) < \varepsilon.$$

To prove (1), observe the bound $[1 \wedge (|x|^2 + |y|^2)] \leq (1 \wedge |x|^2) + (1 \wedge |y|^2)$. Using the marginal condition, for any $\gamma \in \Gamma^\Lambda(\mu, \nu)$ with $\mu \in \mathcal{M}$, $\nu \in \mathcal{N}$, we have

$$\int_{\mathbb{R}^{2d}} 1 \wedge (|x|^2 + |y|^2) d\gamma(x, y) \leq \int_{\mathbb{R}^d} 1 \wedge |x|^2 d\mu(x) + \int_{\mathbb{R}^d} 1 \wedge |y|^2 d\nu(y).$$

Since $\mathcal{M}$ and $\mathcal{N}$ are Λ -tight, the right-hand side is uniformly bounded. This proves (1).

To prove (2), by Λ -tightness of $\mathcal{M}$ and $\mathcal{N}$, there exists a compact set $\tilde{K} \subset \mathbb{R}^d \setminus \{0\}$ such that

$$\sup_{\mu \in \mathcal{M} \cup \mathcal{N}} \int_{\tilde{K}^c} 1 \wedge |x|^2 d\mu(x) < \frac{\varepsilon}{2}.$$

Let $K = \tilde{K} \times \tilde{K}$, which is a compact subset of $\mathbb{R}^{2d} \setminus \{(0, 0)\}$. Then for any $\gamma \in \Gamma^\Lambda(\mu, \nu)$ with $\mu \in \mathcal{M}$, $\nu \in \mathcal{N}$,

$$\begin{aligned} \int_{K^c} 1 \wedge (|x|^2 + |y|^2) d\gamma(x, y) &\leq \int_{(\tilde{K}^c \times \mathbb{R}^d) \cup (\mathbb{R}^d \times \tilde{K}^c)} [(1 \wedge |x|^2) + (1 \wedge |y|^2)] d\gamma(x, y) \\ &\leq \int_{\tilde{K}^c} 1 \wedge |x|^2 d\mu(x) + \int_{\tilde{K}^c} 1 \wedge |y|^2 d\nu(y) \\ &< \varepsilon. \end{aligned}$$

This proves (2), hence Γ is Λ -tight. □

We may now proceed to the proof of Theorem 3.18

Proof of Theorem 3.18. We first observe that for any $\mu, \nu \in \Lambda_2(\mathbb{R}^d)$, the set of Lévy couplings $\Gamma^\Lambda(\mu, \nu)$ is nonempty, see Remark 3.7. Hence, the Lévy cost $\mathcal{C}_2^\Lambda(\mu, \nu)$ is well-defined and finite.

Let $\{\gamma_n\} \subset \Gamma^\Lambda(\mu, \nu)$ be a minimizing sequence, i.e.,

$$\lim_{n \rightarrow \infty} \langle \gamma_n, c_2 \rangle = \mathcal{C}_2^\Lambda(\mu, \nu).$$

By Lemma 3.19, $\Gamma^\Lambda(\mu, \nu)$ is Λ -tight. Hence by the Prokhorov theorem (see Theorem C.7), there is $\gamma_* \in \Lambda(\mathbb{R}^{2d})$ and a subsequence $\{\gamma_{n_k}\}_k$ such that $\gamma_{n_k} \rightarrow \gamma_*$ Λ -weakly.

Let us first show that $\gamma \in \Lambda_2(\mathbb{R}^d)$. Indeed, it follows by Portmanteau lemma (Lemma C.3):

$$\langle \gamma_*, |x|^2 + |y|^2 \rangle \leq \liminf_{n \rightarrow \infty} \langle \gamma_n, |x|^2 + |y|^2 \rangle = \langle \mu, |x|^2 \rangle + \langle \nu, |y|^2 \rangle < \infty.$$

Hence γ_* has a finite second moment.

Next we show $\gamma_* \in \Gamma^\Lambda(\mu, \nu)$. For any $f \in C_b(\mathbb{R}^d)$ that vanishes on some neighbourhood of 0, clearly $|f(x)| \leq C \min\{1, |x|^2\}$ for some constant $C \geq 0$. Since $\gamma_n \rightarrow \gamma$ Λ -weakly we have

$$\langle \gamma_*, f \otimes 1 \rangle = \lim_{n \rightarrow \infty} \langle \gamma_n, f \otimes 1 \rangle = \langle \mu, f \rangle.$$

A similar computation also leads to $\langle \gamma_*, 1 \otimes f \rangle = \langle \nu, f \rangle$. The above holds for all $f \in C_b(\mathbb{R}^d)$ that vanishes on a neighborhood of 0. By Proposition 3.8 (see (3.3)) $\gamma_* \in \Gamma^\Lambda(\mu, \nu)$.

Finally let us prove γ_* is Lévy 2-optimal. Indeed, by Portmanteau lemma we find

$$\langle \gamma_*, c_2 \rangle \leq \liminf_{n \rightarrow \infty} \langle \gamma_n, c_2 \rangle = \mathcal{C}_2^\Lambda(\mu, \nu).$$

This shows γ_* is 2-optimal. The construction of 2-optimal coupling is completed. $\square$

3.5. The fundamental theorem of Lévy optimal transport. In the Lévy optimal transport problem, a similar connection between optimal couplings, cyclical monotonicity, and duality also holds. We present the corresponding result below, which we refer to as the *fundamental theorem of Lévy optimal transport*.

Theorem 3.20 (The fundamental theorem of Lévy optimal transport). *Let $\mu, \nu \in \Lambda_2(\mathbb{R}^d)$ and $\gamma \in \Lambda_2(\mathbb{R}^{2d})$. The following are equivalent.*

- (i) γ is a 2-optimal Lévy coupling of μ, ν .
- (ii) The set $\text{supp}(\gamma) \cup \{(0, 0)\}$ is 2-cyclically monotone.
- (iii) There exists a pair of lower semicontinuous functions $\varphi, \psi : \mathbb{R}^d \rightarrow [-\infty, \infty)$ such that $\varphi \in L^1(\mu), \psi \in L^1(\nu)$, $\varphi \oplus \psi \leq c_2$, $\varphi(0) = \psi(0) = 0$, and $\varphi(x) + \psi(y) = c_2(x, y)$ for all $x, y \in \text{supp}(\gamma)$.

Remark 3.21. As established in Proposition 3.16, if μ, ν have finite equal mass, the Lévy transport cost is always bounded above by the classical transport cost, that is, $\mathcal{C}_2^\Lambda(\mu, \nu) \leq \mathcal{C}_2(\mu, \nu)$. As a consequence of the preceding theorem, the two costs coincide if and only if the point $(0, 0)$ lies in the support of a classical optimal coupling of μ and ν .

Remark 3.22. The pair of functions (φ, ψ) from (iii) is called a *Kantorovich potential* for the pair μ, ν . Particularly it holds

$$\int_{\mathbb{R}^d} \varphi(x) d\mu(x) + \int_{\mathbb{R}^d} \psi(y) d\nu(y) = \langle \mu, \varphi \rangle + \langle \nu, \psi \rangle = \langle \gamma, \varphi \oplus \psi \rangle = \langle \gamma, c_2 \rangle = \mathcal{C}_2^\Lambda(\mu, \nu).$$

The second equality is due to Proposition 3.11 and that $\varphi \in L^1(\mu), \psi \in L^1(\nu)$.

Proof. (i) $\rightarrow$ (ii). Suppose γ is a 2-optimal Lévy coupling of μ, ν . Consider the decomposition $\gamma = \gamma_0 + \zeta$, where

$$d\gamma_0(x, y) = \min \{1, |x|^2 + |y|^2\} d\gamma(x, y), \quad d\zeta(x, y) = (1 - \min \{1, |x|^2 + |y|^2\}) d\gamma(x, y).$$

Note that γ_0 is a bounded Borel measure on $\mathbb{R}^{2d}$ with $\text{supp}(\gamma_0) = \text{supp}(\gamma)$.

Fix $\alpha > 0$ (e.g., $\alpha = 1$) and consider the bounded measure $\tilde{\gamma}_0 = \gamma_0 + \alpha\delta_{(0,0)}$, for then $\text{supp}(\tilde{\gamma}_0) = \text{supp}(\gamma_0) \cup \{(0, 0)\}$. To establish (ii), we will use Theorem 3.4, (iii) $\rightarrow$ (i). So it reduces to show that $\tilde{\gamma}_0$ is classically 2-optimal with respect to its marginals:

$$\tilde{\mu}_0 = \#_1(\gamma_0) + \alpha\delta_0, \quad \tilde{\nu}_0 = \#_2(\gamma_0) + \alpha\delta_0.$$

(Note: $\tilde{\mu}_0, \tilde{\nu}_0$ have equal mass, that is, $\tilde{\mu}_0(\mathbb{R}^d) = \tilde{\nu}_0(\mathbb{R}^d) = \gamma_0(\mathbb{R}^{2d}) + \alpha$.) Suppose $\tilde{\gamma}_0$ is not classically 2-optimal, that is, there exists $\tilde{\gamma}'_0 \in \Gamma(\tilde{\mu}_0, \tilde{\nu}_0)$ such that

$$\langle \tilde{\gamma}'_0, c_2 \rangle < \langle \tilde{\gamma}_0, c_2 \rangle.$$

Consider the measure $\gamma' = \tilde{\gamma}'_0 + \zeta$. We verify that γ' is a Lévy coupling of μ, ν . For any continuous function $\varphi \in C_b(\mathbb{R}^d)$ vanishing on a neighborhood of 0, we have $\langle \delta_0, \varphi \rangle = 0$. Therefore,

$$\begin{aligned} \langle \gamma', \varphi \otimes 1 \rangle &= \langle \tilde{\gamma}'_0, \varphi \otimes 1 \rangle + \langle \zeta, \varphi \otimes 1 \rangle = \langle \#_1(\gamma_0) + \alpha\delta_0, \varphi \rangle + \langle \zeta, \varphi \otimes 1 \rangle \\ &= \langle \gamma_0, \varphi \otimes 1 \rangle + \langle \zeta, \varphi \otimes 1 \rangle = \langle \gamma, \varphi \otimes 1 \rangle = \langle \mu, \varphi \rangle. \end{aligned}$$

The last second equality is due to that $\gamma_0 + \zeta = \gamma$. A similar argument also shows $\langle \gamma', 1 \otimes \varphi \rangle = \langle \nu, \varphi \rangle$. Hence γ' is a Lévy coupling of μ, ν , i.e., $\gamma' \in \Gamma^\Lambda(\mu, \nu)$. Moreover, we find

$$\langle \gamma', c_2 \rangle = \langle \tilde{\gamma}'_0, c_2 \rangle + \langle \zeta, c_2 \rangle < \langle \tilde{\gamma}_0, c_2 \rangle + \langle \zeta, c_2 \rangle = \langle \gamma_0, c_2 \rangle + \langle \zeta, c_2 \rangle = \langle \gamma, c_2 \rangle.$$

This contradicts with the Lévy 2-optimality of γ . Hence $\tilde{\gamma}_0$ is classically 2-optimal.

(ii) $\rightarrow$ (iii). By the fundamental theorem of (classical) optimal transport (Theorem 3.4(ii)), there is a pair of lower semicontinuous potentials (φ, ψ) such that $\varphi \oplus \psi \leq c_2$ and $\varphi(x) + \psi(y) = c_2(x, y)$ for all $(x, y) \in \text{supp}(\gamma) \cup \{(0, 0)\}$. Particularly we have $\varphi(0) + \psi(0) = 0$. By adding and subtracting a constant if necessary, we may normalize so that $\varphi(0) = \psi(0) = 0$.

It remains to verify that $\varphi \in L^1(\mu), \psi \in L^1(\nu)$. Note that $\varphi^+ \in L^1(\mu), \psi^+ \in L^1(\nu)$ since $0 \leq \varphi^+(x), \psi^+(x) \leq \frac{1}{2}|x|^2$, and both μ and ν have finite second moments. By Lemma 3.11,

$$\langle \mu, \varphi \rangle + \langle \nu, \psi \rangle = \langle \gamma, \varphi \otimes 1 \rangle + \langle \gamma, 1 \otimes \psi \rangle = \langle \gamma, \varphi \oplus \psi \rangle = \langle \gamma, c_2 \rangle.$$

where the integrals are understood in the extended real line $[-\infty, \infty)$. The linearity used above is valid in this context. Since γ also has finite second moment, the integral $\langle \gamma, c_2 \rangle$ is finite nonnegative, implying

$$\langle \mu, \varphi \rangle + \langle \nu, \psi \rangle \in (0, \infty).$$

As $\varphi = \varphi^+ - \varphi^-$ and $\varphi^+ \in L^1(\mu)$, it follows that $\varphi^- \in L^1(\mu)$, and similarly $\psi^- \in L^1(\nu)$. Hence, $\varphi \in L^1(\mu)$ and $\psi \in L^1(\nu)$.

(iii) $\rightarrow$ (i). We have $\varphi(0) = \psi(0) = 0$, $\varphi \oplus \psi \leq c_2$, with the equality holds when $(x, y) \in \text{supp}(\gamma) \cup \{(0, 0)\}$. Hence, for any $\tilde{\gamma} \in \Gamma^\Lambda(\mu, \nu)$

$$\langle \gamma, c_2 \rangle = \langle \gamma, \varphi \oplus \psi \rangle = \langle \mu, \varphi \rangle + \langle \nu, \psi \rangle = \langle \tilde{\gamma}, \varphi \oplus \psi \rangle \leq \langle \tilde{\gamma}, c_2 \rangle.$$

Here, the second and third equalities use Lemma 3.11. Therefore (i) is established. $\square$

3.6. Lévy-Kantorovich duality. A consequence of Theorem 3.20 is a Kantorovich-type duality for the Lévy optimal transport problem. In classical optimal transport theory, the Kantorovich duality states that for any $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$,

$$\mathcal{C}_2(\mu, \nu) = \sup_{\varphi, \psi} \left[\int_{\mathbb{R}^d} \varphi(x) d\mu(x) + \int_{\mathbb{R}^d} \psi(y) d\nu(y) \right],$$

where the supremum is taken over all continuous functions φ, ψ such that $\varphi \oplus \psi \leq c_2$, with $c_2(x, y) := \frac{1}{2}|x - y|^2$. By a density argument, one can show that this identity continues to hold when the supremum is restricted to smaller but dense classes of functions, such as $C_b(\mathbb{R}^d)$, $C_0(\mathbb{R}^d)$, $C_c(\mathbb{R}^d)$, or $C_0^2(\mathbb{R}^d)$.

We now turn to the duality principle for the Lévy optimal transport problem, which we refer to as the *Lévy-Kantorovich duality*.

Theorem 3.23 (Lévy-Kantorovich duality). *Let $\mu, \nu \in \Lambda_2(\mathbb{R}^d)$. Then*

$$\mathcal{C}_2^\Lambda(\mu, \nu) = \sup_{\varphi, \psi} \left[\int_{\mathbb{R}^d} \varphi(x) d\mu(x) + \int_{\mathbb{R}^d} \psi(y) d\nu(y) \right], \quad (3.10)$$

where the supremum is taken over all pairs (φ, ψ) satisfying: $\varphi \oplus \psi \leq c_2$, $\varphi(0) = \psi(0) = 0$, and any of the following regularity conditions:

- (i) φ, ψ are lower semicontinuous;
- (ii) $\varphi, \psi \in C_2^2(\mathbb{R}^d)$.
- (iii) $\varphi, \psi \in C_2^2(\mathbb{R}^d)$, and vanish on a neighborhood of the origin;
- (iv) $\varphi, \psi \in C_b^2(\mathbb{R}^d)$, and vanish on a neighborhood of the origin;

Before proving the theorem, we first record the following consequence.

Corollary 3.24. *Let $\mu, \nu \in \Lambda_2(\mathbb{R}^d)$ and set $\mathcal{A} = \Lambda(0, 0, \mu)$, $\mathcal{B} = \Lambda(0, 0, \nu)$. Then, for the restricted dual transport derivative ω'_2 from Definition 2.25, one has*

$$\mathcal{C}_2^\Lambda(\mu, \nu) = \omega'_2(0, 0; \mathcal{A}, \mathcal{B}).$$

Proof. Take $\varphi, \psi \in C_2^2(\mathbb{R}^d)$ with $\varphi(0) = \psi(0) = 0$ and $\varphi \oplus \psi \leq c_2$. In particular, φ and ψ touch $\frac{1}{2}|x|^2$ at 0 from below, which forces $\nabla\varphi(0) = \nabla\psi(0) = 0$. Thus,

$$\mathcal{A}\varphi(0) = \int_{\mathbb{R}^d} [\varphi(x + x') - \varphi(x) - (x')^\top \nabla\varphi(x)] \Big|_{x=0} d\mu(x') = \int_{\mathbb{R}^d} \varphi(x') d\mu(x'), \quad (3.11)$$

and similarly, $\mathcal{B}\psi(0) = \int_{\mathbb{R}^d} \psi(y') d\nu(y')$. Hence, for all admissible φ, ψ ,

$$\mathcal{A}\varphi(0) + \mathcal{B}\psi(0) = \int_{\mathbb{R}^d} \varphi(x) d\mu(x) + \int_{\mathbb{R}^d} \psi(y) d\nu(y).$$

Taking the supremum over such pairs (φ, ψ) , the definition of ω'_2 together with Theorem 3.23(ii) yields the claim. $\square$

Let us now return to the proof of Theorem 3.23. The formula (3.10) holds with the supremum taken over lower semicontinuous pair is contained in Theorem 3.20(iii). To extend the formula to classes of functions with stronger regularity conditions, we need the following lemmas.

Lemma 3.25. *Let X be a topological space and $F : X \rightarrow \mathbb{R} \cup \{-\infty\}$ be an upper semicontinuous function. Suppose $Y \subset X$ is dense. Then it holds*

$$\sup_{x \in X} F(x) = \sup_{x \in Y} F(x).$$

Lemma 3.26. *Let $\mu, \nu \in \Lambda_2(\mathbb{R}^d)$ be two Lévy measures. For any pairs $(\varphi, \psi) \in L^1(\mu) \oplus L^1(\nu)$ satisfying $\varphi \oplus \psi \leq c_2$, there is a sequence of pairs $\{(\varphi_n, \psi_n)\}_n$ such that*

- $\varphi_n, \psi_n \in C_b^2(\mathbb{R}^d)$ and vanish at a neighborhood of the origin;
- $\varphi_n \oplus \psi_n \leq c_2$;
- $\varphi_n \rightarrow \varphi$ in $L^1(\mu)$, $\psi_n \rightarrow \psi$ in $L^1(\nu)$.

The proof of Lemma 3.25 is elementary and omitted. In contrast, the proof of Lemma 3.26 is more technical, relying on the construction of an approximating sequence of smooth function pairs that converge in L^1 to a given admissible pair (φ, ψ) . To avoid disrupting the main flow of the presentation, we defer the proof to the appendix. Let us now return to the proof of Proposition 3.23.

Proof of Proposition 3.23. (i) Let γ be a 2-Lévy optimal coupling of μ, ν , whose existence is established in Theorem 3.18. Consider any pair (φ, ψ) of lower semicontinuous with $\varphi \oplus \psi \leq c_2$ and $\varphi(0) = \psi(0) = 0$. Particularly we have $\varphi(x), \psi(x) \leq \frac{1}{2}|x|^2$, and so $\varphi^+ \in L^1(\mu), \psi^+ \in L^1(\nu)$. By Proposition 3.11, we then have

$$\begin{aligned} \int_{\mathbb{R}^d} \varphi(x) d\mu(x) + \int_{\mathbb{R}^d} \psi(y) d\nu(y) &= \int_{\mathbb{R}^{2d}} [\varphi(x) + \psi(y)] d\gamma(x, y) \\ &\leq \int_{\mathbb{R}^{2d}} c_2(x, y) d\gamma(x, y) = \mathcal{C}_2^\Lambda(\mu, \nu). \end{aligned}$$

Taking the supremum over all such pairs we find (3.10) holds with “ $\geq$ ” in place of “ $=$.” The equality in fact holds, as the maximizer exists, as proved in Theorem 3.20 (see Remark 3.22).

For the remaining of the proof, denote $\mathcal{F}_\bullet$ for $\bullet = (\text{i}), (\text{ii}), (\text{iii}), (\text{iv})$ the family of all pairs (φ, ψ) that satisfies the condition given by (i)–(iv) of Proposition 3.23 respectively, and $\mathcal{C}_\bullet$ for the supremum of

$$\mathcal{C}_\bullet = \sup_{(\varphi, \psi) \in \mathcal{F}_\bullet} F(\varphi, \psi) := \sup_{(\varphi, \psi) \in \mathcal{F}_\bullet} \left[\int_{\mathbb{R}^d} \varphi(x) d\mu(x) + \int_{\mathbb{R}^d} \psi(y) d\nu(y) \right].$$

Note we have the inclusion $\mathcal{F}_{(i)} \supset \mathcal{F}_{(ii)} \supset \mathcal{F}_{(iii)} \supset \mathcal{F}_{(iv)}$, which then implies

$$\mathcal{C}_{(i)} \geq \mathcal{C}_{(ii)} \geq \mathcal{C}_{(iii)} \geq \mathcal{C}_{(iv)}.$$

The proof is complete if we prove $\mathcal{C}_{(i)} = \mathcal{C}_{(iv)}$. By Lemma 3.26, the space $\mathcal{F}_{(iv)}$ is dense in $\mathcal{F}_{(i)}$ w.r.t the metric topology of the product space $L^1(\mu) \oplus L^1(\nu)$. Lemma 3.25 then implies $\mathcal{C}_{(i)} = \mathcal{C}_{(iv)}$. This completes the proof. $\square$

4. OPTIMAL MARKOVIAN COUPLINGS AND STRONG DUALITY FOR DRIFT, DIFFUSION AND JUMP PARTS

As an intermediate step toward proving the existence of optimal Markovian couplings (Theorem 2.28) and establishing strong duality (Theorem 2.31) for a full Lévy generator, we begin by treating the drift, diffusion, and jump components separately. For each component, we construct an explicit 2-optimal Markovian coupling and verify the corresponding strong duality relation. This decomposition will then serve as the basis for combining the components and addressing the general case in the coming section.

Throughout this and the following sections, we simplify notation by suppressing the dependence on the fixed generators $\mathcal{A}, \mathcal{B}$ whenever the context is clear. In particular, for the transport derivatives $\theta_2, \omega_2^\pm, \omega_2^*, \omega_2'$, etc., we write

$$\theta_2 = \theta_2(\cdot, \cdot; \mathcal{A}, \mathcal{B}), \quad \theta_2(x, y) = \theta_2(x, y; \mathcal{A}, \mathcal{B}).$$

4.1. Affine structure of θ_2, ω_2' . We begin by establishing the affine identity for θ_2 from Theorem 2.28(i), along with the corresponding identity for ω_2' .

4.1.1. Affine structure of θ_2 . Let $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$. Recall the definition of Markovian transport derivative θ_2 from Definition 2.13:

$$\theta_2(x, y; \mathcal{A}, \mathcal{B}) = \inf_{\mathcal{J} \in \Gamma(\mathcal{A}, \mathcal{B})} (\mathcal{J}^- c_2)(x, y), \quad c_2(x, y) = \frac{1}{2}|x - y|^2,$$

where the infimum is taken over all Feller couplings $\mathcal{J}$ of $\mathcal{A}, \mathcal{B}$, and $\mathcal{J}^-$ is the lower generator of $\mathcal{J}$, as given in Definition 2.2. We note the lower generator $\mathcal{J}^-$ has the following additive property: if $\Phi \in C(\mathcal{J}), \Psi \in \bar{D}(\mathcal{J})$, then

$$\mathcal{J}^-(\Phi + \Psi) = \mathcal{J}^-\Phi + \mathcal{J}\Psi. \quad (4.1)$$

To prove Theorem 2.28(i), we first introduce a technical lemma. For $h \in \mathbb{R}^d$, denote by $\sigma_h : C(\mathbb{R}^d) \rightarrow C(\mathbb{R}^d)$ the translation operator defined by

$$(\sigma_h \varphi)(x) := \varphi(x + h).$$

Given a Feller generator $\mathcal{J} \in \mathcal{G}(\mathbb{R}^d)$, we define its h -translated generator by

$$\mathcal{J}^h := \sigma_h^{-1} \mathcal{J} \sigma_h = \sigma_{-h} \mathcal{J} \sigma_h.$$

It is straightforward to check that $\mathcal{J}^h$ is again a Feller generator, and it generates the Feller semigroup $e^{t\mathcal{J}^h} = \sigma_h^{-1} e^{t\mathcal{J}} \sigma_h$, see [21, Chapter II, Section 2.1]. It is also easy to verify that the lower generator of $\mathcal{J}^h$ satisfies the same relation:

$$\mathcal{J}^{h-} := (\mathcal{J}^h)^- = \sigma_h^{-1} \mathcal{J}^- \sigma_h. \quad (4.2)$$

Lemma 4.1. *Let $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$, $m_{\mathcal{A}}, m_{\mathcal{B}} \in \mathbb{R}^d$ be their mean vectors, and $\mathcal{J} \in \Gamma(\mathcal{A}, \mathcal{B})$ be a Feller coupling of two Lévy generators. For any $h = (\tau, \tau') \in \mathbb{R}^{2d}$, it holds*

$$[\mathcal{J}^- \sigma_h c_2](x, y) = \mathcal{J}^- c_2(x, y) + (m_{\mathcal{A}} - m_{\mathcal{B}})^\top (\tau - \tau').$$

Proof. Observe the following basic identity for squared cost:

$$(\sigma_h c_2)(x, y) = c_2(x + \tau, y + \tau') = c_2(x, y) + c_2(\tau, \tau') + (x - y)^\top (\tau - \tau').$$

Write $\Theta(x) = x^\top (\tau - \tau')$. Note that $\Theta \otimes 1, 1 \otimes \Theta \in \bar{D}(\mathcal{J})$, by the definition of coupling generators. Since $c_2(\tau, \tau')$ is a constant (w.r.t. x, y), acting $\mathcal{J}$ to it becomes zero. This follows from (4.1):

$$\begin{aligned} [\mathcal{J}^- \sigma_h c_2](x, y) &= (\mathcal{J}^- c_2)(x, y) + \mathcal{J}(\Theta \otimes 1)(x, y) - \mathcal{J}(1 \otimes \Theta)(x, y) \\ &= (\mathcal{J}^- c_2)(x, y) + m_{\mathcal{A}}^\top (\tau - \tau') - m_{\mathcal{B}}^\top (\tau - \tau') \\ &= (\mathcal{J}^- c_2)(x, y) + (m_{\mathcal{A}} - m_{\mathcal{B}})^\top (\tau - \tau'). \end{aligned} \quad \square$$

Let us first establish Theorem 2.28(i).

Proof of Theorem 2.28(i). Let $x_1, y_1, x_2, y_2 \in \mathbb{R}^d$, and $h = (\tau, \tau') = (x_2 - x_1, y_2 - y_1)$. Fix $\epsilon > 0$ and let $\mathcal{J}_\epsilon \in \Gamma(\mathcal{A}, \mathcal{B})$ be a coupling generator (depending on x_1, y_1) such that

$$\mathcal{J}_\epsilon^-(c_2)(x_1, y_1) \leq \inf_{\mathcal{J} \in \Gamma(\mathcal{A}, \mathcal{B})} \mathcal{J}^-(c_2)(x_1, y_1) + \epsilon = \theta_2(x_1, y_1) + \epsilon, \quad (4.3)$$

where $\mathcal{J}_\epsilon^-$ is the lower generator of $\mathcal{J}_\epsilon$. Consider the h -translated generator $\mathcal{J}_\epsilon^h = \sigma_h^{-1} \mathcal{J}_\epsilon \sigma_h$. Note we have $\mathcal{J}_\epsilon^h \in \Gamma(\mathcal{A}, \mathcal{B})$. To see why, notice for any $\varphi \in \bar{D}(\mathcal{A})$ we have $\sigma_h[\varphi \otimes 1] = (\sigma_\tau \varphi) \otimes 1$. Thus, by the translation invariance of Lévy generators, particularly $\sigma_\tau \mathcal{A} = \mathcal{A} \sigma_\tau$, we have

$$\mathcal{J}_\epsilon^h[\varphi \otimes 1] = (\sigma_h^{-1} \mathcal{J}_\epsilon \sigma_h)[\varphi \otimes 1] = \sigma_h^{-1}[(\mathcal{A} \sigma_\tau \varphi) \otimes 1] = (\sigma_\tau^{-1} \mathcal{A} \sigma_\tau \varphi) \otimes 1 = (\mathcal{A} \varphi) \otimes 1.$$

Similarly we can argue that $\mathcal{J}_\epsilon^h[1 \otimes \psi] = \mathcal{B} \psi$ for any $\psi \in \bar{D}(\mathcal{B})$. Hence $\mathcal{J}_\epsilon^h \in \Gamma(\mathcal{A}, \mathcal{B})$, and we therefore conclude

$$\theta_2(x_2, y_2) = \inf_{\mathcal{J} \in \Gamma(\mathcal{A}, \mathcal{B})} \mathcal{J}^-(c_2)(x_2, y_2) \leq (\mathcal{J}_\epsilon^h c_2)(x_2, y_2). \quad (4.4)$$

We now calculate $(\mathcal{J}_\epsilon^h c_2)(x_2, y_2)$. First by Lemma 4.1, we have

$$(\mathcal{J}_\epsilon^- \sigma_h c_2)(x_2, y_2) = (\mathcal{J}_\epsilon^- c_2)(x_2, y_2) + (m_{\mathcal{A}} - m_{\mathcal{B}})^\top (\tau - \tau').$$

Applying the inverse translation σ_h^{-1} and the identity (4.2), this follows

$$\begin{aligned} (\mathcal{J}_\epsilon^h c_2)(x_2, y_2) &= (\sigma_h^{-1} \mathcal{J}_\epsilon^- \sigma_h c_2)(x_2, y_2) = (\mathcal{J}_\epsilon^- c_2)(x_2 - \tau, y_2 - \tau') + (m_{\mathcal{A}} - m_{\mathcal{B}})^\top (\tau - \tau') \\ &= (\mathcal{J}_\epsilon^- c_2)(x_1, y_1) + (m_{\mathcal{A}} - m_{\mathcal{B}})^\top ((x_2 - x_1) - (y_2 - y_1)). \end{aligned}$$

By (4.3) and (4.4), this follows

$$\theta_2(x_2, y_2) \leq \theta_2(x_1, y_1) - (m_{\mathcal{A}} - m_{\mathcal{B}})^\top ((x_2 - x_1) - (y_2 - y_1)) + \epsilon.$$

Passing $\epsilon \searrow 0$,

$$\theta_2(x_2, y_2) \leq \theta_2(x_1, y_1) + (m_{\mathcal{A}} - m_{\mathcal{B}})^\top ((x_2 - x_1) - (y_2 - y_1)).$$

Switching the role of (x_1, y_1) and (x_2, y_2) , the above holds with $\leq$ replaced by $=$. The proposition follows as the special case $(x_1, y_1) = (0, 0)$. $\square$

4.1.2. *Affine structure of ω'_2 .* We next prove the affine identity for ω'_2 . Let us recall the definition of the *restricted dual transport derivative*. For $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$, it is given by

$$\omega'_2(x_0, y_0; \mathcal{A}, \mathcal{B}) := \sup_{(\varphi, \psi) \in \mathcal{D}'(x_0, y_0)} [\mathcal{A}\varphi(x_0) + \mathcal{B}\psi(y_0)],$$

where $\mathcal{D}'(x_0, y_0)$ denotes the set of admissible pairs $(\varphi, \psi) \in [C_2^2(\mathbb{R}^d)]^2$ satisfying the *touching condition*

$$\varphi(x) + \psi(y) \leq \frac{1}{2}|x - y|^2, \quad \text{with equality at } (x_0, y_0).$$

The first step towards proving the chain of equalities in (5.1) is to show that ω'_2 admits the same affine structure as θ_2 , established in Theorem 2.28(i).

Proposition 4.2. *Let $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$. It holds for all $x_0, y_0 \in \mathbb{R}^d$:*

$$\omega'_2(x_0, y_0; \mathcal{A}, \mathcal{B}) = \omega'_2(0, 0; \mathcal{A}, \mathcal{B}) + (m_{\mathcal{A}} - m_{\mathcal{B}})^\top (x_0 - y_0).$$

Proof. Let us assume that $\mathcal{A}, \mathcal{B}$ admit the global form $\mathcal{A} = \Lambda_{\text{glob}}(\kappa, \alpha, \mu)$, $\mathcal{B} = \Lambda_{\text{glob}}(\zeta, \beta, \nu)$. Suppressing the generators $\mathcal{A}, \mathcal{B}$ in ω'_2 , we are to prove: for all $x_0, y_0 \in \mathbb{R}^d$,

$$\omega'_2(x_0, y_0) = \omega'_2(0, 0) + (\kappa - \zeta)^\top (x_0 - y_0).$$

Fix $(\varphi, \psi) \in \mathcal{D}'(x_0, y_0)$ and consider the second-order Taylor remainder of φ, ψ at x_0, y_0 respectively:

$$\begin{aligned} \tilde{\varphi}(x) &:= \varphi(x + x_0) - \varphi(x_0) - x^\top \nabla \varphi(x_0), \\ \tilde{\psi}(y) &:= \psi(y + y_0) - \psi(y_0) - y^\top \nabla \psi(y_0). \end{aligned}$$

We claim that $(\tilde{\varphi}, \tilde{\psi}) \in \mathcal{D}'(0, 0)$. To see this, first note that since $\varphi \oplus \psi$ touches c_2 from below at (x_0, y_0) , their gradients agree at (x_0, y_0) , i.e.,

$$(\nabla \varphi(x_0), \nabla \psi(y_0)) = (\nabla_{(x, y)}(\varphi \oplus \psi))(x_0, y_0) = (\nabla_{(x, y)} c_2)(x_0, y_0) = (x_0 - y_0, y_0 - x_0).$$

This yields

$$\begin{aligned} \tilde{\varphi}(x) + \tilde{\psi}(y) &= \varphi(x + x_0) + \psi(y + y_0) - (\varphi(x_0) + \psi(y_0)) - x^\top \nabla \varphi(x_0) - y^\top \nabla \psi(y_0) \\ &\leq \frac{1}{2}|(x + x_0) - (y + y_0)|^2 - \frac{1}{2}|x_0 - y_0|^2 - (x - y)^\top (x_0 - y_0) \\ &= \frac{1}{2}|x - y|^2 = c_2(x, y). \end{aligned}$$

Clearly the pair is twice differentiable and satisfies the quadratic bound (2.1). Since $\tilde{\varphi}(0) = \tilde{\psi}(0) = 0$, this shows $(\tilde{\varphi}, \tilde{\psi}) \in \mathcal{D}'(0, 0)$.

We next observe that

$$D^2 \varphi(x_0) = D^2 \tilde{\varphi}(0), \quad D^2 \psi(y_0) = D^2 \tilde{\psi}(0), \quad \nabla \tilde{\varphi}(0) = 0 = \nabla \tilde{\psi}(0).$$

If we decompose the generators $\mathcal{A}, \mathcal{B}$ into their diffusion, drift, and jump parts, we find

$$\mathcal{A}^\Delta \varphi(x_0) = \mathcal{A}^\Delta \tilde{\varphi}(0), \quad \mathcal{A}^\nabla \varphi(x_0) = \kappa^\top (x_0 - y_0),$$

$$\mathcal{B}^\Delta \psi(y_0) = \mathcal{B}^\Delta \tilde{\psi}(0), \quad \mathcal{B}^\nabla \psi(y_0) = -\zeta^\top(x_0 - y_0).$$

Moreover, since $\tilde{\varphi}(0) = 0$ and $\nabla \tilde{\varphi}(0) = 0$, it follows that

$$\mathcal{A}^J \varphi(x_0) = \int_{\mathbb{R}^d} [\varphi(x + x_0) - \varphi(x_0) - x \cdot \nabla \varphi(x_0)] d\mu(x) = \int_{\mathbb{R}^d} \tilde{\varphi}(x) d\mu(x) = \mathcal{A}^J \tilde{\varphi}(0).$$

Likewise, $\mathcal{B}^J \psi(y_0) = \mathcal{B}^J \tilde{\psi}(0)$. Combining all these observations, we have

$$\begin{aligned} \mathcal{A}\varphi(x_0) + \mathcal{B}\psi(y_0) &= \mathcal{A}^\Delta \tilde{\varphi}(0) + \mathcal{B}^\Delta \tilde{\psi}(0) + \mathcal{A}^J \tilde{\varphi}(0) + \mathcal{B}^J \tilde{\psi}(0) + (\kappa - \zeta)^\top(x_0 - y_0) \\ &= \mathcal{A} \tilde{\varphi}(0) + \mathcal{B} \tilde{\psi}(0) + (\kappa - \zeta)^\top(x_0 - y_0) \\ &\leq \omega'_2(0, 0) + (\kappa - \zeta)^\top(x_0 - y_0). \end{aligned}$$

Taking the supremum of the left-hand side over all $(\varphi, \psi) \in \mathcal{D}'(x_0, y_0)$ we find the bound

$$\omega'_2(x_0, y_0; \mathcal{A}, \mathcal{B}) \leq \omega'_2(0, 0; \mathcal{A}, \mathcal{B}) + (\kappa - \zeta)^\top(x_0 - y_0).$$

The reverse inequality follows from the same argument above, together with the following observation: if $(\tilde{\varphi}, \tilde{\psi}) \in \mathcal{D}'(0, 0)$, and

$$\varphi(x) = \tilde{\varphi}(x - x_0) + x^\top(x_0 - y_0), \quad \psi(y) = \tilde{\psi}(y - y_0) + y^\top(y_0 - x_0),$$

then $(\varphi, \psi) \in \mathcal{D}'(x_0, y_0)$. □

4.2. Reduction of global to local minimization. The general c -optimal Markovian coupling problem for a pair of Feller generators $\mathcal{A}, \mathcal{B} \in \mathcal{G}(\Pi)$ is, by default, a *global* optimization problem. That is, the aim is to find a *Feller coupling*—a Feller generator on Π^2 —denoted by $\mathcal{J}_*$, such that the minimum is attained pointwise:

$$(\mathcal{J}_* c)(x, y) = \theta_c(x, y; \mathcal{A}, \mathcal{B}), \quad (x, y) \in \Pi^2.$$

In the case of optimal Markovian couplings between Lévy generators with respect to the quadratic cost $c_2(x, y) = \frac{1}{2}|x - y|^2$, the problem admits a significant simplification. Owing to the translation invariance of Lévy generators—and, in particular, the affine structure of θ_2 and ω'_2 —the global optimization problem reduces to a *local* one. In fact, it suffices to minimize the pointwise expression $(\mathcal{J} c_2)(x, y)$ at a single reference point, such as $(x, y) = (0, 0)$, among all *Lévy* (rather than merely *Feller*) coupling generators.

In what follows, for $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$, we denote by $\Gamma^\Lambda(\mathcal{A}, \mathcal{B}) \subset \Gamma(\mathcal{A}, \mathcal{B})$ the set of all Lévy coupling generators of $\mathcal{A}$ and $\mathcal{B}$.

Proposition 4.3. *Let $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$.*

(i) *Suppose there exists a Lévy generator $\mathcal{J}_* \in \Gamma^\Lambda(\mathcal{A}, \mathcal{B})$ such that*

$$(\mathcal{J}_* c_2)(0, 0) = \theta_2(0, 0; \mathcal{A}, \mathcal{B}).$$

Then $\mathcal{J}_$ is a 2-optimal Markovian coupling of $\mathcal{A}$ and $\mathcal{B}$.*

(ii) *The same conclusion as in (i) holds if θ_2 is replaced by ω'_2 . In this case, the strong duality relation is satisfied:*

$$\omega'_2(x, y; \mathcal{A}, \mathcal{B}) = \theta_2(x, y; \mathcal{A}, \mathcal{B}).$$

Proof. Let us again suppress the generators $\mathcal{A}, \mathcal{B}$ in θ_2, ω'_2 .

(i) Let $\mathcal{J}_*$ be given in the corollary. To show $\mathcal{J}_*$ is optimal, we need to establish the identity $(\mathcal{J}_*c_2)(x_0, y_0) = \theta_2(x_0, y_0)$ for all $(x_0, y_0) \in \mathbb{R}^{2d}$. Let $h = (x_0, y_0)$, for then

$$(\mathcal{J}_*c_2)(x_0, y_0) = (\sigma_h \mathcal{J}_*c_2)(0, 0).$$

Since $\mathcal{J}_*$ is a Lévy generator, it commutes with any translation operator, that is $\sigma_h \mathcal{J}_* = \mathcal{J}_* \sigma_h$. By Lemma 4.1 we find

$$\begin{aligned} (\mathcal{J}_*c_2)(x_0, y_0) &= (\sigma_h \mathcal{J}_*c_2)(0, 0) = (\mathcal{J}_* \sigma_h c_2)(0, 0) = (\mathcal{J}_*c_2)(0, 0) + (m_{\mathcal{A}} - m_{\mathcal{B}})^\top (x_0 - y_0) \\ &= \theta_2(0, 0) + (m_{\mathcal{A}} - m_{\mathcal{B}})^\top (x_0 - y_0) = \theta_2(x_0, y_0). \end{aligned}$$

The proof is complete.

(ii) We note the inequality holds: $\omega'_2(0, 0) \leq \theta_2(0, 0) \leq (\mathcal{J}_*c_2)(0, 0)$. The first is due to Proposition 2.21 and the second is from the definition of θ_2 . If $(\mathcal{J}_*c_2)(0, 0) = \omega'_2(0, 0)$, then it forces the chain of inequality holds. Since $\omega'_2(0, 0) = \theta_2(0, 0)$, the affine structure of θ_2, ω'_2 , particularly Theorem 2.28(i) and Proposition 4.2 imply $\theta_2 = \omega'_2$. $\square$

4.3. Optimal Markovian couplings and strong duality of drift-diffusion parts. Let us begin with considering the optimal Markovian couplings between a pair of drift-diffusion operators. Let $\alpha, \beta \in \mathcal{S}_{\geq 0}(\mathbb{R}^d)$ be two nonnegative definite $(d \times d)$ -matrices, $\kappa, \zeta \in \mathbb{R}^d$ be two drift vectors, and $\mathcal{A} = \Lambda_2(\kappa, \alpha, 0), \mathcal{B} = \Lambda_2(\zeta, \beta, 0)$, namely,

$$(\mathcal{A}\varphi)(x) = \kappa \cdot \nabla \varphi(x) + \frac{1}{2} \text{tr}[\alpha D^2 \varphi](x), \quad (\mathcal{B}\psi)(y) = \zeta \cdot \nabla \psi(y) + \frac{1}{2} \text{tr}[\beta D^2 \psi](y).$$

As discussed, the more general case of drift-diffusion generators with x -dependent coefficients was treated in [16]. We present the construction here for the sake of completeness.

A prototypical coupling of the drift-diffusion operators $\mathcal{A}$ and $\mathcal{B}$ above is itself a diffusion operator $\mathcal{J} = \Lambda_2(\eta, \sigma, 0)$ on $\mathbb{R}^{2d}$, where $\sigma \in \mathcal{S}_{\geq 0}(\mathbb{R}^{2d})$ takes the block form

$$\sigma = \begin{bmatrix} \alpha & K \\ K^\top & \beta \end{bmatrix} \geq 0, \quad (4.5)$$

and $\eta = \kappa \oplus \zeta \in \mathbb{R}^{2d}$. The off-diagonal blocks are transposes of each other to ensure that σ is symmetric. Furthermore, the matrix K is chosen so that σ remains nonnegative semidefinite.

The optimal Markovian coupling problem for two drift-diffusion generators is deeply related to the classical optimal transport problem between two *Gaussian measures*, which has been thoroughly studied in the literature. Given a probability measure $\mu \in \mathcal{P}(\mathbb{R}^d)$, we write

$$\mu \sim \mathcal{N}(m, \sigma), \quad m \in \mathbb{R}^d, \sigma \in \mathcal{S}_{\geq 0}(\mathbb{R}^d),$$

to indicate that μ is a d -dimensional Gaussian measure with mean m and covariance matrix σ . In [23], Givens and Shortt showed that if $\mu \sim \mathcal{N}(m_1, \alpha)$ and $\nu \sim \mathcal{N}(m_2, \beta)$, a (classical) 2-optimal transport plan γ_* between μ and ν is given by a Gaussian measure $\gamma_* \sim \mathcal{N}(m, \sigma)$ on $\mathbb{R}^{2d}$, where

$$m = (m_1, m_2), \quad \sigma = \begin{bmatrix} \alpha & K_* \\ K_*^\top & \beta \end{bmatrix},$$

and K_* is a matrix that solves the following positive semidefinite programming problem:

$$\text{minimize } K \mapsto \frac{1}{2} \text{tr}(\alpha + \beta - 2K), \quad \text{subject to } \begin{bmatrix} \alpha & K \\ K^\top & \beta \end{bmatrix} \geq 0. \quad (4.6)$$

We also note the uniqueness of such coupling holds if the covariance matrices α, β are non-degenerate, i.e., strictly positive definite.

As a result, the 2-optimal transport cost is given by

$$\mathcal{C}_2(\mu, \nu) = \frac{1}{2} \int_{\mathbb{R}^{2d}} |x - y|^2 d\gamma_*(x, y) = \frac{1}{2} |m_1 - m_2|^2 + \mathcal{W}_S(\alpha, \beta)^2,$$

where

$$\mathcal{W}_S(\alpha, \beta)^2 := \frac{1}{2} \text{tr}(\alpha + \beta - 2K_*) = \min_K \frac{1}{2} \text{tr}(\alpha + \beta - 2K), \quad (4.7)$$

with the minimum taken over all K satisfying the constraint in (4.6). Moreover, $\mathcal{W}_S(\alpha, \beta)$ admits the following explicit formula:

$$\mathcal{W}_S(\alpha, \beta)^2 := \frac{1}{2} \text{tr} \left[\alpha + \beta - 2 \left(\alpha^{1/2} \beta \alpha^{1/2} \right)^{1/2} \right].$$

The functional $\mathcal{W}_S$ defines on the space $\mathcal{S}_{\geq 0}(\mathbb{R}^d)$ of nonnegative definite matrices is called the *Bures-Wasserstein distance*, which is a metric on the space.

With these results, we may now address the optimal Markovian coupling problem for drift-diffusion operators.

Proposition 4.4. *Let $\mathcal{A} = \Lambda_{\text{glob}}(\kappa, \alpha, 0)$ and $\mathcal{B} = \Lambda_{\text{glob}}(\zeta, \beta, 0)$ be drift-diffusion generators on $\mathbb{R}^d$. Then the Markovian transport derivative is given by*

$$\theta_2(x, y; \mathcal{A}, \mathcal{B}) = (\kappa - \zeta)^\top (x - y) + \mathcal{W}_S(\alpha, \beta)^2, \quad (4.8)$$

where $\mathcal{W}_S(\alpha, \beta)$ is as defined in (4.7). Moreover, an optimal Markovian coupling of $\mathcal{A}$ and $\mathcal{B}$ is given by the drift diffusion generator $\mathcal{J} = \Lambda_{\text{glob}}(\eta, \sigma, 0)$, where σ is defined as in (4.5) with K being a minimizer of the semidefinite program (4.6), and $\eta = (\kappa, \zeta) \in \mathbb{R}^{2d}$.

Proof. By Theorem 2.28(i), it suffices to prove (4.8) for $x = y = 0$, that is,

$$\theta_2(0, 0; \mathcal{A}, \mathcal{B}) = \mathcal{W}_S(\alpha, \beta)^2. \quad (4.9)$$

Let $\mu_t = \delta_0 e^{t\mathcal{A}}, \nu_t = \delta_0 e^{t\mathcal{B}}$. We have $\mu_t \sim \mathcal{N}(t\kappa, t\alpha), \nu_t \sim \mathcal{N}(t\zeta, t\beta)$. From the discussion earlier, $\gamma_t \sim \mathcal{N}(t\eta, t\sigma_*)$, $\eta = (\kappa, \zeta) \in \mathbb{R}^{2d}$, is a 2-optimal coupling of μ_t, ν_t , where $\sigma_* \in \mathcal{S}_{\geq 0}(\mathbb{R}^{2d})$ is given by (4.5), (4.6). If we let $\mathcal{J}_* = \Lambda_{\text{glob}}(\eta, \sigma_*, 0)$, then $\mathcal{J}_* \in \Gamma(\mathcal{A}, \mathcal{B})$ and we have $\gamma_t = \delta_{(0,0)} e^{t\mathcal{J}_*}$. We specifically point out that the semidefinite programming problem (4.6) has the following linear property:

$$\mathcal{W}_S(t\alpha, t\beta)^2 = t\mathcal{W}_S(\alpha, \beta)^2.$$

Hence we have

$$(e^{t\mathcal{J}_*} c_2)(0, 0) = \int_{\mathbb{R}^{2d}} c_2(x, y) d\gamma_t = \mathcal{C}_2(\delta_0 e^{t\mathcal{A}}, \delta_0 e^{t\mathcal{B}}) = \mathcal{C}_2(\mu_t, \nu_t) = \frac{1}{2} t^2 |\kappa - \zeta|^2 + t\mathcal{W}_S(\alpha, \beta)^2,$$

Subtracting $c_2(0, 0) = 0$ and dividing t then passing $t \searrow 0$, it leads to

$$(\mathcal{J}_* c_2)(0, 0) = \omega_2^+(0, 0; \mathcal{A}, \mathcal{B}) = \limsup_{t \searrow 0} \frac{\mathcal{C}_2(\delta_0 e^{t\mathcal{A}}, \delta_0 e^{t\mathcal{B}})}{t} = \mathcal{W}_S(\alpha, \beta)^2,$$

where we recall the pointwise upper transport derivative ω_2^+ from Definition 2.4. We point out the above also holds if ω_2^+ is replaced by the lower derivative ω_2^- (and hence $\omega_2^+ = \omega_2^-$). Since $\theta_2(0, 0; \mathcal{A}, \mathcal{B})$ is the infimum of $(\mathcal{J} c_2)(0, 0)$ over all $\mathcal{J} \in \Gamma(\mathcal{A}, \mathcal{B})$, it follows

$$\theta_2(0, 0; \mathcal{A}, \mathcal{B}) \leq \omega_2^\pm(0, 0; \mathcal{A}, \mathcal{B}) = \mathcal{W}_S(\alpha, \beta)^2.$$

Recall the bound $\omega_2^\pm \leq \theta_2$ was established in Proposition 2.21. This proves (4.9). Moreover, $\mathcal{J}_*$ is the 2-optimal coupling generator of $\mathcal{A}, \mathcal{B}$, because $(\mathcal{J}_* c_2)(0, 0) = \theta_2(0, 0; \mathcal{A}, \mathcal{B})$. $\square$

We now present a duality formula for the minimization problem (4.8), which will play a key role in the next chapter. This duality arises from the theory of positive semidefinite linear programming. For completeness, we provide a proof using tools from optimal transport.

Lemma 4.5. *Let $\alpha, \beta \in \mathcal{S}_{\geq 0}(\mathbb{R}^d)$, and let $\mathcal{W}_S(\alpha, \beta)^2$ denote the minimum value of the primal problem (4.6).*

$$\mathcal{W}_S(\alpha, \beta)^2 = \sup_{A, B} \operatorname{tr}[\alpha A + \beta B], \quad (4.10)$$

where the supremum is taken over all pairs of positive semidefinite matrices (A, B) satisfying the matrix inequality

$$\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \leq \frac{1}{2} \begin{bmatrix} I & -I \\ -I & I \end{bmatrix}. \quad (4.11)$$

Proof. As discussed above, $\mathcal{W}_S(\alpha, \beta)$ is the 2-Wasserstein distance between two centered Gaussian measures $\mu \sim \mathcal{N}(0, \alpha)$ and $\nu \sim \mathcal{N}(0, \beta)$. By classical Kantorovich duality, we have

$$\mathcal{W}_2^2(\mu, \nu) = \sup_{\varphi \oplus \psi \leq c_2} \left\{ \int_{\mathbb{R}^d} \varphi(x) d\mu(x) + \int_{\mathbb{R}^d} \psi(y) d\nu(y) \right\}.$$

For Gaussian marginals, the optimal Kantorovich potentials (φ, ψ) are known to be generalized quadratic functions. When α and β are positive definite, the optimal potentials take the form

$$\varphi(x) = x^\top A x, \quad \psi(y) = y^\top B y,$$

for some symmetric matrices $A, B \geq 0$ satisfying the constraint

$$\varphi(x) + \psi(y) \leq \frac{1}{2}|x - y|^2 \quad \text{for all } x, y.$$

This pointwise inequality is equivalent to the matrix inequality (4.11).

In the case where α or β is singular, the optimal potentials are limits of such quadratic forms supported on subspaces, and the dual value is still attained as the supremum over such quadratic pairs. Therefore, it suffices to restrict the dual formulation to quadratic potentials

satisfying the cost constraint. In this case, using the property of Gaussian measures, we compute

$$\int_{\mathbb{R}^d} \varphi(x) d\mu(x) + \int_{\mathbb{R}^d} \psi(y) d\nu(y) = \int_{\mathbb{R}^d} x^\top A x d\mu(x) + \int_{\mathbb{R}^d} y^\top B y d\nu(y) = \text{tr}(\alpha A) + \text{tr}(\beta B).$$

This proves the identity (4.10). $\square$

Corollary 4.6. *Assume the setting of Proposition 4.4. The strong duality holds: for all $x, y \in \mathbb{R}^d$,*

$$\omega'_2(x, y; \mathcal{A}, \mathcal{B}) = \theta_2(x, y; \mathcal{A}, \mathcal{B}) = (\kappa - \zeta)^\top (x - y) + \mathcal{W}_S(\alpha, \beta)^2.$$

Proof. The second equality is established in Proposition 4.4. To prove the first, by the affine structure of θ_2, ω'_2 , it suffices to verify for $x = y = 0$, that is,

$$\omega'_2(0, 0) = \theta_2(0, 0) = \mathcal{W}_S(\alpha, \beta)^2.$$

Since $\omega'_2(0, 0) \leq \theta_2(0, 0)$ is established in Proposition 2.21, it suffices to prove the reverse inequality. By Lemma 4.5, for any $\varepsilon > 0$, there exist positive definite matrices $A, B \in \mathcal{S}_{>0}(\mathbb{R}^d)$ satisfying the matrix condition (4.11) such that

$$\mathcal{W}_S(\alpha, \beta)^2 \leq \text{tr}[\alpha A + \beta B] + \varepsilon.$$

Define test functions $\varphi(x) = \frac{1}{2}x^\top A x$ and $\psi(y) = \frac{1}{2}y^\top B y$. Then $\varphi \oplus \psi \leq c_2$ and $\varphi(0) = \psi(0) = 0$, so $(\varphi, \psi) \in \mathcal{D}'(0, 0)$. Hence,

$$\omega'_2(0, 0) \geq \mathcal{A}\varphi(0) + \mathcal{B}\psi(0) = \text{tr}[\alpha A + \beta B] \geq \mathcal{W}_S(\alpha, \beta)^2 - \varepsilon.$$

Taking $\varepsilon \searrow 0$ gives the required inequality. $\square$

4.4. Optimal Markovian coupling and strong duality of jump parts. We now solve the optimal Markovian coupling problem for pure jump generators. As hinted in the previous section, the solution is closely related to the Lévy optimal transport problem between the underlying Lévy measures.

Proposition 4.7. *Let $\mu, \nu \in \Lambda_2(\mathbb{R}^d)$ and set $\mathcal{A} = \Lambda_{\text{glob}}(0, 0, \mu)$, $\mathcal{B} = \Lambda_{\text{glob}}(0, 0, \nu)$. Let γ_* be a Lévy 2-optimal coupling of μ, ν . Then the operator $\mathcal{J}_* := \Lambda_{\text{glob}}(0, 0, \gamma_*)$ is an optimal Markovian coupling of $\mathcal{A}, \mathcal{B}$. Moreover, the strong duality holds:*

$$\omega'_2(x, y; \mathcal{A}, \mathcal{B}) = \theta_2(x, y; \mathcal{A}, \mathcal{B}) = \mathcal{C}_2^\Lambda(\mu, \nu).$$

Proof. To show $\mathcal{J}_*$ is an optimal Markovian coupling and the strong duality (the first equality), by Proposition 4.3(ii) it suffices to show $(\mathcal{J}_* c_2)(0, 0) = \omega'_2(0, 0)$. Indeed,

$$(\mathcal{J}_* c_2)(0, 0) = \int_{\mathbb{R}^{2d}} \left(c_2(x + x', y + y') - c_2(x, y) - (x', y')^\top \nabla_{x, y} c_2(x, y) \right) \Big|_{(x, y) = (0, 0)} d\gamma_*(x', y').$$

Since $c_2(0, 0) = 0$ and $\nabla_{x, y} c_2(0, 0) = (0, 0)$, the integrand reduces to $c_2(x', y') = \frac{1}{2}|x' - y'|^2$. Hence, by the Lévy 2-optimality of γ_* ,

$$(\mathcal{J}_* c_2)(0, 0) = \int_{\mathbb{R}^{2d}} \frac{1}{2}|x' - y'|^2 d\gamma_*(x', y') = \mathcal{C}_2^\Lambda(\mu, \nu).$$

The identity $\mathcal{C}_2^\Lambda(\mu, \nu) = \omega'_2(0, 0)$ is established in Corollary 3.24. This follows $(\mathcal{J}_*c_2)(0, 0) = \omega'_2(0, 0)$.

Now to show the second equality, we note in the global form, any pure jump operators have zero mean, i.e., $m_{\mathcal{A}} = m_{\mathcal{B}} = 0$. We have shown $\theta_2(0, 0) = \mathcal{C}_2^\Lambda(\mu, \nu)$. The affine structure from Theorem 2.28(i) implies the equality also holds for all $x, y \in \mathbb{R}^d$. $\square$

5. STRONG DUALITY AND OPTIMAL MARKOVIAN COUPLINGS FOR LÉVY GENERATORS

In this section, we establish strong duality and the existence of an optimal Markovian coupling for a given pair of full Lévy generators $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$. In the previous section, we verified these results separately for the drift, diffusion, and jump components. Here, we combine those componentwise constructions to handle the full Lévy generator, thereby completing the proof of the general case.

5.1. Strong duality: proofs of Theorems 2.31 and 2.32. Let us start with proving the strong duality Theorem 2.31, that is, establishing the equality

$$\omega'_2(x, y; \mathcal{A}, \mathcal{B}) = \omega_2^*(x, y; \mathcal{A}, \mathcal{B}) = \omega_2^\pm(x, y; \mathcal{A}, \mathcal{B}) = \theta_2(x, y; \mathcal{A}, \mathcal{B}) \quad (5.1)$$

for all Lévy generators $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$ with bounded second moment and $x, y \in \mathbb{R}^d$. We remind the readers that the inequality $\omega'_2 \leq \omega_2^* \leq \omega_2^\pm \leq \theta_2$ holds, see Proposition 2.21 (and its proof in Appendix B) and Remark 2.26.

The main tool in the proof of the strong duality is the following lemma, which establishes the additivity of the functional $\omega'_2(\cdot, \cdot; \mathcal{A}, \mathcal{B})$ with respect to the Lévy–Khintchine decomposition of the operators $\mathcal{A}$ and $\mathcal{B}$. Due to the technical nature of the proof, we will state the lemma now but postpone its proof to the end of this section.

Lemma 5.1. *Let $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$ be in the global form, and $\mathcal{A}^\bullet, \mathcal{B}^\bullet$, $\bullet \in \{\nabla, \Delta, J\}$ be the diffusion, drift, and jump part of $\mathcal{A}, \mathcal{B}$ respectively. It holds for all $x_0, y_0 \in \mathbb{R}^d$:*

$$\omega'_2(x_0, y_0; \mathcal{A}, \mathcal{B}) = \sum_{\bullet \in \{\nabla, \Delta, J\}} \omega'_2(x_0, y_0; \mathcal{A}^\bullet, \mathcal{B}^\bullet).$$

We may now establish the strong duality.

Proof of Theorem 2.31. As pointed out earlier, the chain of inequalities holds:

$$\omega'_2(x, y) \leq \omega_2^*(x, y) \leq \omega_2^-(x, y) \leq \omega_2^+(x, y) \leq \theta_2(x, y), \quad \text{for all } (x, y) \in \mathbb{R}^{2d}.$$

To complete the proof, it suffices to show that $\theta_2(x, y) \leq \omega'_2(x, y)$. By the affine structure of these functions, namely, Lemma 4.2 and Theorem 2.28(i), it is enough to verify the identity at the origin:

$$\theta_2(0, 0) = \omega'_2(0, 0). \quad (5.2)$$

As in Lemma 5.1, for $\bullet \in \{\nabla, \Delta, J\}$ denote $\mathcal{A}^\bullet, \mathcal{B}^\bullet$ the drift, diffusion and jump part of $\mathcal{A}, \mathcal{B}$. Let $\mathcal{J}_*^\bullet \in \Gamma^\Lambda(\mathcal{A}^\bullet, \mathcal{B}^\bullet)$ be an optimal Lévy coupling generator of $\mathcal{A}^\bullet, \mathcal{B}^\bullet$, whose existence is guaranteed by Propositions 4.4 and 4.7. Specifically it holds for $\bullet = \nabla, \Delta, J$:

$$(\mathcal{J}_*^\bullet c_2)(0, 0) = \theta_2(0, 0; \mathcal{A}^\bullet, \mathcal{B}^\bullet).$$

Consider the superposition operator $\mathcal{J}_* = \mathcal{J}_*^\nabla + \mathcal{J}_*^\Delta + \mathcal{J}_*^J$. It is easy to verify that $\mathcal{J}_*$ is a Lévy coupling generator of $\mathcal{A}, \mathcal{B}$, i.e., $\mathcal{J}_* \in \Gamma^\Lambda(\mathcal{A}, \mathcal{B})$. Hence, by the definition of θ_2 , it holds

$$\theta_2(0, 0) \leq (\mathcal{J}_* c_2)(0, 0) = \sum_{\bullet \in \{\nabla, \Delta, J\}} \theta_2(0, 0; \mathcal{A}^\bullet, \mathcal{B}^\bullet).$$

As verify earlier in Corollary 4.6 and Proposition 4.7, the strong duality $\theta_2 = \omega'_2$ holds for each part of generators. It then follows from Lemma 5.1 that

$$\theta_2(0, 0) \leq \sum_{\bullet \in \{\nabla, \Delta, J\}} \theta_2(0, 0; \mathcal{A}^\bullet, \mathcal{B}^\bullet) = \sum_{\bullet \in \{\nabla, \Delta, J\}} \omega'_2(0, 0; \mathcal{A}^\bullet, \mathcal{B}^\bullet) = \omega'_2(0, 0).$$

This is (5.2). The strong duality is established. $\square$

Proof of Theorem 2.32. It follows immediately from Lemma 5.1 and Theorem 2.31, particularly $\omega'_2 = \theta_2$. $\square$

5.2. Optimal Markovian couplings: Proofs of Theorems 2.28(ii, iii). With the strong duality and additivity of θ_2 in place, we may now complete the construction of an optimal Markovian coupling for any given pair $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$, and the proof of Theorem 2.28(ii).

Proof of Theorem 2.28, (ii). As in the theorem, let $\mathcal{A} = \Lambda_{\text{glob}}(\kappa, \alpha, \mu)$, $\mathcal{B} = \Lambda_{\text{glob}}(\zeta, \beta, \nu)$, and $\mathcal{J}_* = \Lambda_{\text{glob}}(\eta_*, \sigma_*, \gamma_*)$, where the triplet $(\eta_*, \sigma_*, \gamma_*)$ is given as in the statement. Clearly, $\mathcal{J}_* \in \Gamma^\Lambda(\mathcal{A}, \mathcal{B})$.

Decompose $\mathcal{J}_*$ into its drift, diffusion, and jump components, $\mathcal{J}_* = \mathcal{J}_*^\nabla + \mathcal{J}_*^\Delta + \mathcal{J}_*^J$. By construction of the triplet, each component $\mathcal{J}_*^\bullet, \bullet = \nabla, \Delta, J$, is an optimal Markovian coupling of $\mathcal{A}^\bullet$ and $\mathcal{B}^\bullet$. In particular, it holds

$$(\mathcal{J}_*^\bullet c_2)(0, 0) = \theta_2(0, 0; \mathcal{A}^\bullet, \mathcal{B}^\bullet).$$

It then follows from Theorem 2.32 that

$$(\mathcal{J}_* c_2)(0, 0) = \sum_{\bullet \in \{\nabla, \Delta, J\}} (\mathcal{J}_*^\bullet c_2)(0, 0) = \sum_{\bullet \in \{\nabla, \Delta, J\}} \theta_2(0, 0; \mathcal{A}^\bullet, \mathcal{B}^\bullet) = \theta_2(0, 0; \mathcal{A}, \mathcal{B}).$$

By Proposition 4.3, $\mathcal{J}_*$ is therefore a 2-optimal Markovian coupling of $\mathcal{A}$ and $\mathcal{B}$, which completes the construction. $\square$

We now present the proof of Theorem 2.28(iii). The next lemma provides a lower bound on the right-hand lower Dini derivative of $t \mapsto (T_t c_2)(x, y)$, which is the only remaining ingredient.

Lemma 5.2. *Let $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$, and let $\omega_2^+ = \omega_2^+(\cdot, \cdot; \mathcal{A}, \mathcal{B})$ denote the upper point-wise transport derivative. Let $\{T_t\}_{t \geq 0}$ be a Markovian coupling semigroup of $\mathcal{A}, \mathcal{B}$. For $(D^- T_t c_2)(x, y)$ denoting the right-hand lower Dini derivative of the map $t \mapsto (T_t c_2)(x, y)$, we have*

$$(D^- T_t c_2)(x, y) \geq (T_t \omega_2^+)(x, y).$$

Proof. We begin with the case $t = 0$. Observe that

$$D^- \Big|_{t=0} T_t c_2(x, y) = \liminf_{h \searrow 0} \frac{T_h c_2(x, y) - c_2(x, y)}{h} \geq \omega_2^+(x, y). \quad (5.3)$$

The proof of (5.3) is identical to the argument establishing $\theta_c \geq \omega_c^+$ in Proposition 2.21; see Appendix B.

For general $t \geq 0$, the semigroup property yields

$$(D^- T_t c_2)(x, y) = \liminf_{h \searrow 0} \frac{T_{t+h} c_2 - T_t c_2}{h}(x, y) = \liminf_{h \searrow 0} T_t \left(\frac{T_h c_2 - c_2}{h} \right)(x, y).$$

By Fatou's lemma applied with respect to the transition kernel of T_t , and using that Proposition 2.27(i) provides a uniform lower bound in h , we obtain

$$(D^- T_t c_2)(x, y) \geq T_t \left(\liminf_{h \searrow 0} \frac{T_h c_2 - c_2}{h} \right)(x, y).$$

Finally, combining this with (5.3) and the order-preserving property of Markov semigroups gives the desired inequality. $\square$

Proof of Theorem 2.28(iii). Fix a Markovian coupling semigroup $\{T_t\}_{t \geq 0}$ of $\mathcal{A}, \mathcal{B}$. By Proposition 2.27(i), the map $t \mapsto (T_t c_2)(x, y)$ is locally Lipschitz, hence absolutely continuous. By the fundamental theorem of calculus for absolutely continuous functions, its derivative exists for almost every $t \geq 0$ and coincides with the lower Dini derivative $(D^- T_t c_2)(x, y)$. Invoking Lemma 5.2 together with strong duality (namely $\omega_2^+ = \theta_2$), we obtain

$$(T_t c_2)(x, y) = c_2(x, y) + \int_0^t (D^- T_s c_2)(x, y) ds \geq c_2(x, y) + \int_0^t (T_s \theta_2)(x, y) ds.$$

By the affine structure of θ_2 established in Theorem 2.28(i), and using the marginal property of $\{T_t\}$, we compute for any $t \geq 0$ and $(x, y) \in \mathbb{R}^{2d}$:

$$(T_s \theta_2)(x, y) = \theta_2(0, 0) + (m_{\mathcal{A}} - m_{\mathcal{B}})^\top [(x - y) + s(m_{\mathcal{A}} - m_{\mathcal{B}})].$$

Substituting this into the integral yields

$$\begin{aligned} (T_t c_2)(x, y) &\geq c_2(x, y) + \int_0^t [\theta_2(0, 0) + (m_{\mathcal{A}} - m_{\mathcal{B}})^\top [(x - y) + s(m_{\mathcal{A}} - m_{\mathcal{B}})]] ds \\ &= c_2(x, y) + t[\theta_2(0, 0) + (m_{\mathcal{A}} - m_{\mathcal{B}})^\top (x - y)] + \frac{t^2}{2} |m_{\mathcal{A}} - m_{\mathcal{B}}|^2. \end{aligned}$$

Moreover, equality holds when $T_t = e^{t\mathcal{J}_*}$ for a 2-optimal coupling $\mathcal{J}_*$, giving precisely the identity (2.23). Consequently, we conclude $T_t c_2 \geq e^{t\mathcal{J}_*} c_2$. The proof is complete. $\square$

5.3. Additivity of ω'_2 (proof of Lemma 5.1). We now return to the proof of Lemma 5.1, which asserts the additivity of ω'_2 with respect to the Lévy–Khintchine decomposition of the generators. The key idea is to observe how each component of the generators interacts with the test functions φ, ψ . When $\mathcal{A}$ and $\mathcal{B}$ are diffusion or drift operators, the expression $\mathcal{A}\varphi(x_0) + \mathcal{B}\psi(y_0)$ depends only on the local behavior of φ and ψ near x_0 and y_0 , respectively. In contrast, when $\mathcal{A}$ and $\mathcal{B}$ are pure jump operators, the same quantity depends on the global

behavior of φ and ψ , due to the nonlocal nature of jumps. This separation of influence—local for drift and diffusion, global for jumps—implies that the contributions from each component can be optimized independently. Consequently, the total value $\omega'_2(x_0, y_0; \mathcal{A}, \mathcal{B})$ decomposes additively across the three types of components.

Although the underlying idea is straightforward, the proof involves some technical subtleties. The strategy is to consider, for each part of the Lévy–Khintchine decomposition, an optimal (or approximately optimal) pair of test functions. The key idea is to combine these pairs in a suitable way to construct an admissible test function pair that approximates a near-maximizer for the full generators $\mathcal{A}, \mathcal{B}$. The main technical difficulty lies in this construction: due to the differing local and nonlocal dependencies of the generator components, several careful modifications are required to ensure that the combined test functions satisfy the admissibility constraints while preserving near-optimality. These modifications form the core of the technical part of the proof.

Given $\varepsilon > 0$, define the truncated cost functions c_2^ε and the remainder δ_2^ε by

$$c_2^\varepsilon(x, y) = \frac{1}{2} \max\{0, |x - y| - \varepsilon\}^2, \quad \delta_2^\varepsilon(x, y) = c_2(x, y) - c_2^\varepsilon(x, y).$$

Specifically, δ_2^ε can be computed explicitly: $\delta_2^\varepsilon(x, y) = q_2^\varepsilon(|x - y|)$, where $q_2^\varepsilon : [0, \infty) \rightarrow [0, \infty)$ is given by

$$q_2^\varepsilon(r) = \frac{1}{2} r^2 \chi_{\{r \leq \varepsilon\}}(r) + \left(\varepsilon r - \frac{1}{2} \varepsilon^2 \right) \chi_{\{r > \varepsilon\}}(r). \quad (5.4)$$

In particular, δ_2^ε is a C^1 function such that $\delta_2^\varepsilon = c_2$ for all $|x - y| \leq \varepsilon$, and is affine (i.e., linear) for $|x - y| > \varepsilon$. Clearly we have the decomposition

$$c_2(x, y) = c_2^\varepsilon(x, y) + \delta_2^\varepsilon(x, y). \quad (5.5)$$

The key idea is to construct a pair of test functions of the form

$$(\varphi_\varepsilon, \psi_\varepsilon) := (\varphi_\varepsilon^\Delta + \varphi_\varepsilon^J, \psi_\varepsilon^\Delta + \psi_\varepsilon^J),$$

where each pair $(\varphi_\varepsilon^\bullet, \psi_\varepsilon^\bullet)$ is a near-optimal pair for $\omega'_2(0, 0; \mathcal{A}^\bullet, \mathcal{B}^\bullet)$, with $\bullet \in \{\Delta, J\}$. These are chosen to satisfy

$$\varphi_\varepsilon^J \oplus \psi_\varepsilon^J \leq c_2^\varepsilon, \quad \varphi_\varepsilon^\Delta \oplus \psi_\varepsilon^\Delta \leq \delta_2^\varepsilon.$$

Summing the two components, we obtain $\varphi_\varepsilon \oplus \psi_\varepsilon \leq c_2$, which ensures the admissibility of the combined pair in the dual formulation.

To proceed, we require two approximation lemmas. The first extends Lemma 3.26 by constructing a sequence of C_2^2 -pairs suitable for the jump part. The second handles the diffusion component by building a localized approximation near the origin.

Lemma 5.3. *Let $\mu, \nu \in \Lambda_2(\mathbb{R}^d)$ be Lévy measures. For any pair $(\varphi, \psi) \in L^1(\mu) \oplus L^1(\nu)$ satisfying $\varphi \oplus \psi \leq c_2$, there exists a family $\{(\varphi_\varepsilon, \psi_\varepsilon)\}_{\varepsilon > 0}$ such that for all sufficiently small $\varepsilon > 0$, the following hold:*

- (i) $\varphi_\varepsilon, \psi_\varepsilon \in C_2^2(\mathbb{R}^d)$, and both vanish in a neighborhood of the origin;
- (ii) $\varphi_\varepsilon \oplus \psi_\varepsilon \leq c_2^\varepsilon$;
- (iii) $\varphi_\varepsilon \rightarrow \varphi$ in $L^1(\mu)$, and $\psi_\varepsilon \rightarrow \psi$ in $L^1(\nu)$ as $\varepsilon \searrow 0$.

Lemma 5.4. *Let $\mu, \nu \in \Lambda_2(\mathbb{R}^d)$, and suppose $\varphi, \psi \in C_2^2(\mathbb{R}^d)$ satisfy the quadratic growth condition $|\varphi(x)|, |\psi(x)| \leq C|x|^2$ for some $C \geq 0$, along with $\varphi \oplus \psi \leq c_2$. Then there exists a family $\{(\varphi_\varepsilon, \psi_\varepsilon)\}_{\varepsilon>0}$ such that for all $\varepsilon > 0$, the following hold:*

- (i) $\varphi_\varepsilon, \psi_\varepsilon \in C_b^2(\mathbb{R}^d)$, and $\varphi_\varepsilon(x) = \varphi(x)$, $\psi_\varepsilon(x) = \psi(x)$ for all $|x| < \frac{\varepsilon}{4}$;
- (ii) $\varphi_\varepsilon \oplus \psi_\varepsilon \leq \delta_2^\varepsilon$;
- (iii) $\varphi_\varepsilon \rightarrow 0$ in $L^1(\mu)$, and $\psi_\varepsilon \rightarrow 0$ in $L^1(\nu)$ as $\varepsilon \searrow 0$.

The proof of Lemma 5.3, similar to that of Lemma 3.26, is technical and thus deferred to the appendix. The proof of Lemma 5.4 is presented below.

Proof of Lemma 5.4. Fix a C^∞ function $\rho : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ such that

$$\rho(r) = r \text{ for } r \in [0, \frac{1}{2}], \quad 0 \leq \rho(r) \leq 1, \quad \rho(\infty) = 1, \quad |\rho'(r)| \leq 1 \text{ for all } r \geq 0.$$

For each $\varepsilon > 0$, define $\rho_\varepsilon : [0, \infty) \rightarrow [0, \infty)$ and the cutoff map $T_\varepsilon : \mathbb{R}^d \rightarrow \mathbb{R}^d$ by

$$\rho_\varepsilon(r) := \varepsilon \rho\left(\frac{r}{\varepsilon}\right), \quad T_\varepsilon(x) := \frac{x}{|x|} \rho_\varepsilon(|x|), \quad \text{with } T_\varepsilon(0) := 0.$$

One can easily verify the following properties:

- (1) T_ε is smooth and 1-Lipschitz on $\mathbb{R}^d$;
- (2) $T_\varepsilon(x) = x$ for all $x \in B_{\varepsilon/2}(0)$;
- (3) $T_\varepsilon(x) \in B_\varepsilon(0)$ for all $x \in \mathbb{R}^d$;
- (4) $T_\varepsilon(x) = 0$ for all sufficiently large $|x|$.

Following this, given φ, ψ from the lemma, let us define for $\varepsilon > 0$:

$$\varphi_\varepsilon(x) := \varphi(T_{\frac{\varepsilon}{2}}(x)), \quad \psi_\varepsilon(x) := \psi(T_{\frac{\varepsilon}{2}}(x)).$$

We now verify Properties (i)–(iii) holds for the family $\{(\varphi_\varepsilon, \psi_\varepsilon)\}_{\varepsilon>0}$. Property (i) follows directly from the construction. To verify (ii), note that $T_{\frac{\varepsilon}{2}}$ is 1-Lipschitz and satisfies $|T_{\frac{\varepsilon}{2}}(x)| \leq \frac{\varepsilon}{2}$. Since $\varphi \oplus \psi \leq c_2$, we estimate:

$$\varphi_\varepsilon(x) + \psi_\varepsilon(y) = \varphi(T_{\frac{\varepsilon}{2}}(x)) + \psi(T_{\frac{\varepsilon}{2}}(y)) \leq \frac{1}{2} |T_{\frac{\varepsilon}{2}}(x) - T_{\frac{\varepsilon}{2}}(y)|^2 \leq \tilde{q}_2^\varepsilon(|x - y|),$$

where $\tilde{q}_2^\varepsilon(r) := \frac{1}{2} \min\{r, \varepsilon\}^2$. Since $\tilde{q}_2^\varepsilon \leq q_2^\varepsilon$ from (5.4), it follows that $\varphi_\varepsilon \oplus \psi_\varepsilon \leq \delta_2^\varepsilon$, as desired.

For (iii), observe that since $T_{\frac{\varepsilon}{2}}$ is 1-Lipschitz and satisfies $|T_{\frac{\varepsilon}{2}}(x)| \leq \frac{\varepsilon}{2}$, we have

$$|\varphi_\varepsilon(x)| = |\varphi(T_{\frac{\varepsilon}{2}}(x))| \leq C |T_{\frac{\varepsilon}{2}}(x)|^2 \leq C \min\left\{|x|, \frac{\varepsilon}{2}\right\}^2 \leq C|x|^2.$$

As $|x|^2 \in L^1(\mu)$, the dominated convergence theorem implies $\varphi_\varepsilon \rightarrow 0$ in $L^1(\mu)$. An identical argument shows $\psi_\varepsilon \rightarrow 0$ in $L^1(\nu)$. $\square$

We may now proceed to prove Lemma 5.1.

Proof of Lemma 5.1. For notational simplicity, we suppress the subscript 2 and the generator arguments $\mathcal{A}, \mathcal{B}$ in $\omega' = \omega'_2$ and write

$$\omega'_\bullet(x, y) := \omega'_2(x, y; \mathcal{A}^\bullet, \mathcal{B}^\bullet), \quad \bullet \in \{\Delta, \nabla, J\}.$$

By Proposition 4.2, it suffices to verify the identity at the origin, i.e., $x = y = 0$. Since $\omega'_\nabla(0,0) = 0$ by Corollary 4.6, the goal reduces to proving

$$\omega'(0,0) = \omega'_\Delta(0,0) + \omega'_J(0,0),$$

where we set $\omega' := \omega'_2(0,0; \mathcal{A}, \mathcal{B})$. The inequality $\omega'(0,0) \leq \omega'_\Delta(0,0) + \omega'_J(0,0)$ follows immediately from the definition and subadditivity of the supremum. We now prove the reverse inequality. Fix $\eta > 0$, and we will show that

$$\omega'(0,0) \geq \omega'_\Delta(0,0) + \omega'_J(0,0) - \eta. \quad (5.6)$$

Let us introduce the shorthand for $\varphi, \psi \in C_2^2(\mathbb{R}^d)$:

$$\mathcal{K}(\varphi, \psi) := \mathcal{A}\varphi(0) + \mathcal{B}\psi(0), \quad \mathcal{K}_\bullet(\varphi, \psi) := \mathcal{A}^\bullet\varphi(0) + \mathcal{B}^\bullet\psi(0) \text{ for } \bullet \in \{\nabla, \Delta, J\}.$$

Clearly, $\mathcal{K} = \mathcal{K}_\nabla + \mathcal{K}_\Delta + \mathcal{K}_J$, and by definition

$$\omega'_\bullet = \sup_{(\varphi, \psi) \in \mathcal{D}'(0,0)} \mathcal{K}_\bullet(\varphi, \psi).$$

We first make a key observation. For any $(\varphi, \psi) \in \mathcal{D}'(0,0)$, we may normalize by adding a constant so that $\varphi(0) = \psi(0) = 0$. By the same computation from (3.11), we have

$$\mathcal{K}_J(\varphi, \psi) = \mathcal{A}^J\varphi(0) + \mathcal{B}^J\psi(0) = \int_{\mathbb{R}^d} \varphi(x) d\mu(x) + \int_{\mathbb{R}^d} \psi(y) d\nu(y). \quad (5.7)$$

Let $(\varphi^\bullet, \psi^\bullet) \in \mathcal{D}'(0,0)$, $\bullet \in \{\Delta, J\}$, be such that

$$\mathcal{K}_\bullet(\varphi^\bullet, \psi^\bullet) \geq \omega'_\bullet(0,0) - \frac{\eta}{2}. \quad (5.8)$$

Let $\{(\varphi_\varepsilon^\bullet, \psi_\varepsilon^\bullet)\}_{\varepsilon>0}$ be the approximating families constructed in Lemmas 5.3 and 5.4, respectively. For the jump part, the pair $(\varphi_\varepsilon^J, \psi_\varepsilon^J) \in \mathcal{D}'(0,0)$ vanishes near the origin and converges to (φ^J, ψ^J) in $L^1(\mu) \oplus L^1(\nu)$. By (5.7), we conclude:

$$\mathcal{K}_\nabla(\varphi^J, \psi^J) = \mathcal{K}_\Delta(\varphi^J, \psi^J) = 0, \quad \text{and} \quad \lim_{\varepsilon \rightarrow 0} \mathcal{K}_J(\varphi_\varepsilon^J, \psi_\varepsilon^J) = \mathcal{K}_J(\varphi^J, \psi^J). \quad (5.9)$$

For the diffusion part, the approximants satisfy

$$(\varphi_\varepsilon^\Delta, \psi_\varepsilon^\Delta) = (\varphi^\Delta, \psi^\Delta) \quad \text{on } B_{\varepsilon/4}(0), \quad \text{and} \quad (\varphi_\varepsilon^\Delta, \psi_\varepsilon^\Delta) \rightarrow (0,0) \text{ in } L^1(\mu) \oplus L^1(\nu).$$

Since $\mathcal{A}^\Delta, \mathcal{B}^\Delta$ are local operators, we have

$$\mathcal{K}_\nabla(\varphi_\varepsilon^\Delta, \psi_\varepsilon^\Delta) = 0, \quad \mathcal{K}_\Delta(\varphi_\varepsilon^\Delta, \psi_\varepsilon^\Delta) = \mathcal{K}_\Delta(\varphi^\Delta, \psi^\Delta), \quad \text{and} \quad \lim_{\varepsilon \rightarrow 0} \mathcal{K}_J(\varphi_\varepsilon^\Delta, \psi_\varepsilon^\Delta) = 0. \quad (5.10)$$

Now define the combined test pair:

$$\varphi_\varepsilon := \varphi_\varepsilon^\Delta + \varphi_\varepsilon^J, \quad \psi_\varepsilon := \psi_\varepsilon^\Delta + \psi_\varepsilon^J.$$

By construction and Lemmas 5.3 and 5.4, we have

$$\varphi_\varepsilon \oplus \psi_\varepsilon \leq \delta_2^\varepsilon + c_2^\varepsilon = c_2,$$

and $\varphi_\varepsilon(0) + \psi_\varepsilon(0) = 0$, so $(\varphi_\varepsilon, \psi_\varepsilon) \in \mathcal{D}'(0,0)$. Using (5.9) and (5.10), we compute

$$\mathcal{K}(\varphi_\varepsilon, \psi_\varepsilon) = \sum_{\bullet \in \{\nabla, \Delta, J\}} \sum_{* \in \{\Delta, J\}} \mathcal{K}_\bullet(\varphi_\varepsilon^*, \psi_\varepsilon^*)$$

$$= \mathcal{K}_\Delta(\varphi^\Delta, \psi^\Delta) + \mathcal{K}_J(\varphi_\varepsilon^J, \psi_\varepsilon^J) + \mathcal{K}_J(\varphi_\varepsilon^\Delta, \psi_\varepsilon^\Delta).$$

Passing to the limit as $\varepsilon \searrow 0$ and using (5.9), (5.10) and (5.8), we obtain

$$\omega'(0, 0) \geq \limsup_{\varepsilon \rightarrow 0} \mathcal{K}(\varphi_\varepsilon, \psi_\varepsilon) \geq \omega'_\Delta(0, 0) + \omega'_J(0, 0) - \eta,$$

which proves (5.6). Since $\eta > 0$ was arbitrary, the result follows. $\square$

6. WASSERSTEIN-TYPE METRIC ON THE SPACE $\mathcal{G}_2^\Lambda(\mathbb{R}^d)$ AND $\Lambda_2(\mathbb{R}^d)$

With the groundwork laid in the preceding sections, we now turn to the metric on the space $\mathcal{G}_2^\Lambda(\mathbb{R}^d)$ of Lévy-type generators with finite second moments, induced by optimal couplings. The metric functional is given in (2.25), namely

$$\mathcal{W}_\mathcal{G}(\mathcal{A}, \mathcal{B})^2 := \frac{1}{2}|m_\mathcal{A} - m_\mathcal{B}|^2 + \theta_2(0, 0; \mathcal{A}, \mathcal{B}), \quad (6.1)$$

where $m_\mathcal{A}, m_\mathcal{B} \in \mathbb{R}^d$ denote the mean vectors associated with the generators $\mathcal{A}$ and $\mathcal{B}$. We will call this metric *Wasserstein generator metric* on $\mathcal{G}_2^\Lambda(\mathbb{R}^d)$. Since Theorem 2.31 established the identity $\omega_2^* = \omega_2 = \theta_2$, the term θ_2 in (6.1) can be equivalently replaced by any of these expressions.

The contribution of $\theta_2(0, 0; \mathcal{A}, \mathcal{B})$ depends only on the *centered parts* of $\mathcal{A}$ and $\mathcal{B}$. Indeed, writing

$$\mathcal{A} = \bar{\mathcal{A}} + m_\mathcal{A} \cdot \nabla, \quad \mathcal{B} = \bar{\mathcal{B}} + m_\mathcal{B} \cdot \nabla,$$

we see that $\bar{\mathcal{A}}$ and $\bar{\mathcal{B}}$ have vanishing mean vectors, and hence are referred to as centered. In particular, if $\mathcal{A}, \mathcal{B}$ are given in the global Lévy–Khintchine form, the centered generator consists precisely of the diffusion and jump parts, i.e. $\bar{\mathcal{A}} = \mathcal{A}^\Delta + \mathcal{A}^J$.

By Theorem 2.32, the term θ_2 decomposes naturally into drift, diffusion, and jump contributions. Moreover, Lemma 4.6 shows that the drift contribution (evaluating at $x = y = 0$) vanishes. The remaining contributions correspond to the squared Bures–Wasserstein distance in the diffusion part, and to the Lévy optimal transport cost $(\mathcal{W}_\Lambda)^2 := \mathcal{C}_2^\Lambda$ (see Definition 3.13) in the jump part. Consequently, we arrive at the explicit expression

$$\mathcal{W}_\mathcal{G}(\mathcal{A}, \mathcal{B})^2 = \frac{1}{2}|\kappa - \zeta|^2 + \mathcal{W}_\mathcal{S}(\alpha, \beta)^2 + \mathcal{W}_\Lambda(\mu, \nu)^2, \quad (6.2)$$

where $\mathcal{A} = \Lambda_{\text{glob}}(\kappa, \alpha, \mu)$ and $\mathcal{B} = \Lambda_{\text{glob}}(\zeta, \beta, \nu)$. The metric $\mathcal{W}_\Lambda$ on $\Lambda_2(\mathbb{R}^d)$ will be called *Lévy–Wasserstein metric*.

In the subsequent subsections, we establish several fundamental properties of the metric $\mathcal{W}_\mathcal{G}$ and $\mathcal{W}_\Lambda$, contained in Theorems 2.33, 2.35, and Corollary 2.34, including:

- that $\mathcal{W}_\mathcal{G}, \mathcal{W}_\Lambda$ define metrics;
- lower semicontinuity and weak compactness;
- completeness;
- separability;
- a maximal-type bound on the path space.

6.1. Wasserstein generator metric. We begin by establishing that the functional $\mathcal{W}_G$ indeed defines a metric on the space $\mathcal{G}_2^\Lambda(\mathbb{R}^d)$. This fact is a direct consequence of the strong duality result in Theorem 2.31.

Proof of Theorem 2.33, $\mathcal{W}_G$ is a metric. We first show that the functional is symmetric in $(\mathcal{A}, \mathcal{B})$. By (6.1) and $\theta_2 = \omega_2^+$, it suffices to verify that $\omega_2^+(0, 0; \mathcal{A}, \mathcal{B}) = \omega_2^+(0, 0; \mathcal{B}, \mathcal{A})$. Indeed, from the definition of the pointwise transport derivative,

$$\omega_2^+(0, 0; \mathcal{A}, \mathcal{B}) = \limsup_{t \searrow 0} \frac{1}{t} \mathcal{C}_2(\delta_0 e^{t\mathcal{A}}, \delta_0 e^{t\mathcal{B}}) = \limsup_{t \searrow 0} \frac{1}{t} \mathcal{C}_2(\delta_0 e^{t\mathcal{B}}, \delta_0 e^{t\mathcal{A}}) = \omega_2^+(0, 0; \mathcal{B}, \mathcal{A}).$$

Hence, $\mathcal{W}_G^\Lambda$ is symmetric and clearly nonnegative.

We now show that $\mathcal{W}_G(\mathcal{A}, \mathcal{B}) = 0$ if and only if $\mathcal{A} = \mathcal{B}$. The reverse direction is immediate, since if $\mathcal{A} = \mathcal{B}$ then $m_{\mathcal{A}} = m_{\mathcal{B}}$, and

$$\omega_2^+(0, 0; \mathcal{A}, \mathcal{A}) = \limsup_{t \searrow 0} \frac{1}{t} \mathcal{C}_2(\delta_0 e^{t\mathcal{A}}, \delta_0 e^{t\mathcal{A}}) = 0.$$

Conversely, suppose $\mathcal{W}_G(\mathcal{A}, \mathcal{B}) = 0$. Then $m_{\mathcal{A}} = m_{\mathcal{B}}$ and $\theta_2(0, 0; \mathcal{A}, \mathcal{B}) = 0$. By Theorem 2.28(i), this implies $\theta_2(x, x; \mathcal{A}, \mathcal{B}) = 0$ for all $x \in \mathbb{R}^d$. From Theorem 2.28(iii), any 2-optimal Lévy coupling generator $\mathcal{J}_* \in \Gamma^\Lambda(\mathcal{A}, \mathcal{B})$ satisfies $e^{t\mathcal{J}_*} c_2(x, x) = c_2(x, x) = 0$. Since $\delta_{(x,x)} e^{t\mathcal{J}_*}$ is a coupling of $\delta_x e^{t\mathcal{A}}$ and $\delta_x e^{t\mathcal{B}}$, we obtain

$$\mathcal{C}_2(\delta_x e^{t\mathcal{A}}, \delta_x e^{t\mathcal{B}}) \leq \langle \delta_{(x,x)} e^{t\mathcal{J}_*}, c_2 \rangle = (e^{t\mathcal{J}_*} c_2)(x, x) = 0.$$

Hence, for all $x \in \mathbb{R}^d$ and $t \geq 0$, we have $\delta_x e^{t\mathcal{A}} = \delta_x e^{t\mathcal{B}}$, which implies $e^{t\mathcal{A}} = e^{t\mathcal{B}}$ and therefore $\mathcal{A} = \mathcal{B}$.

It remains to show that $\mathcal{W}_G$ satisfies the triangle inequality. By (6.1), it suffices to verify that the map $(\mathcal{A}, \mathcal{B}) \mapsto \omega_2^+(0, 0; \mathcal{A}, \mathcal{B})^{1/2}$ satisfies the triangle inequality. Take any $\mathcal{A}, \mathcal{B}, \mathcal{C} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$. Then using the triangle inequality for Wasserstein-2 metric, we have

$$\begin{aligned} \omega_2^+(0, 0; \mathcal{A}, \mathcal{C})^{1/2} &= \limsup_{t \searrow 0} t^{-1/2} \mathcal{W}_2(\delta_0 e^{t\mathcal{A}}, \delta_0 e^{t\mathcal{C}}) \\ &\leq \limsup_{t \searrow 0} t^{-1/2} [\mathcal{W}_2(\delta_0 e^{t\mathcal{A}}, \delta_0 e^{t\mathcal{B}}) + \mathcal{W}_2(\delta_0 e^{t\mathcal{B}}, \delta_0 e^{t\mathcal{C}})] \\ &\leq \omega_2^+(0, 0; \mathcal{A}, \mathcal{B})^{1/2} + \omega_2^+(0, 0; \mathcal{B}, \mathcal{C})^{1/2}. \end{aligned}$$

This completes the proof. $\square$

The following result provides an alternative characterization of the Wasserstein generator metric $\mathcal{W}_G$.

Proposition 6.1. *Let $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$, and $\mathcal{J}_*$ be an optimal Lévy coupling generator of $\mathcal{A}, \mathcal{B}$. One has*

$$\mathcal{W}_G(\mathcal{A}, \mathcal{B}) = e^{\mathcal{J}_*} c_2(0, 0) = \inf_{\mathcal{J} \in \Gamma^\Lambda(\mathcal{A}, \mathcal{B})} e^{\mathcal{J}} c_2(0, 0).$$

Proof. Let $\mathcal{J}_* \in \Lambda(\mathcal{A}, \mathcal{B})$ be a 2-optimal coupling of $\mathcal{A}, \mathcal{B}$. By Theorem 2.28(iii)

$$(e^{\mathcal{J}_*} c_2)(0, 0) = \theta_2(0, 0; \mathcal{A}, \mathcal{B}) + \frac{1}{2} |m_{\mathcal{A}} - m_{\mathcal{B}}|^2 = \mathcal{W}_G(\mathcal{A}, \mathcal{B}).$$

From the same theorem, $e^{t\mathcal{J}^*}c_2 \leq e^{t\mathcal{J}}c_2$ for all $\mathcal{J} \in \Gamma(\mathcal{A}, \mathcal{B})$ and $t \geq 0$. This implies the desired equality. $\square$

Before proceeding, we state the following proposition, which will serve as a useful tool for subsequent results. Its proof follows directly from (2.23) in Theorem 2.28.

Proposition 6.2. *Let $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$. Then it holds for all $x, y \in \mathbb{R}^d$ and $t \geq 0$:*

$$\mathcal{W}_2(\delta_x e^{t\mathcal{A}}, \delta_y e^{t\mathcal{B}})^2 \leq \frac{1}{2}|x - y|^2 + t[\theta(0, 0; \mathcal{A}, \mathcal{B}) + (x - y)^\top (m_{\mathcal{A}} - m_{\mathcal{B}})] + \frac{1}{2}t^2|m_{\mathcal{A}} - m_{\mathcal{B}}|^2.$$

Specifically, when $x = y$,

$$\mathcal{W}_2(\delta_x e^{t\mathcal{A}}, \delta_x e^{t\mathcal{B}})^2 \leq t\theta_2(0, 0; \mathcal{A}, \mathcal{B}) + \frac{1}{2}t^2|m_{\mathcal{A}} - m_{\mathcal{B}}|^2 \leq \max\{t, t^2\} \mathcal{W}_G(\mathcal{A}, \mathcal{B})^2.$$

6.2. Lower semicontinuity, weak compactness and completeness. Since the metric $\mathcal{W}_G$ is naturally induced by the Wasserstein-2 metric, it is natural to investigate compactness properties in analogy with those observed in the Wasserstein-2 setting. In particular, we aim to show that every bounded subset $G \subset (\mathcal{G}_2^\Lambda(\mathbb{R}^d), \mathcal{W}_G)$ is relatively compact with respect to a suitable weaker topology. By boundedness, we mean there is $\mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$, or equivalently $\mathcal{B} = 0$, and a constant $R > 0$, such that

$$\sup_{\mathcal{A} \in G} \mathcal{W}_G(\mathcal{A}, \mathcal{B}) < R.$$

Let us provide the following notion of convergence of probability generators.

Definition 6.3. For $n \geq 1$, let $\mathcal{A}, \mathcal{A}_n \in \mathcal{G}(\Pi)$. We say that $\mathcal{A}_n \rightarrow \mathcal{A}$ *weakly in the sense of transition kernels*, or simply *weakly*, if, for all $x \in \Pi$ and $t \geq 0$,

$$\delta_x e^{t\mathcal{A}_n} \rightarrow \delta_x e^{t\mathcal{A}} \quad \text{weakly in } \mathcal{P}(\Pi).$$

Remark 6.4. If $\mathcal{A}_n, \mathcal{A}$ are Lévy generators, then $\mathcal{A}_n \rightarrow \mathcal{A}$ weakly in the sense of transition kernels if and only if $\delta_0 e^{\mathcal{A}_n} \rightarrow \delta_0 e^{\mathcal{A}}$, weakly. This equivalence follows from the translation invariance and infinite divisibility of Lévy measures. Indeed, weak convergence at time $t = 1$ already guarantees convergence for all $t \geq 0$. To see this, recall that the characteristic function of a Lévy process at time t is of the form $\exp(t\psi(\xi))$, where ψ is the Lévy–Khintchine exponent. If $\delta_0 e^{\mathcal{A}_n} \rightarrow \delta_0 e^{\mathcal{A}}$ at $t = 1$, then by Lévy’s continuity theorem we have pointwise convergence of the characteristic functions $\exp(\psi_n(\xi)) \rightarrow \exp(\psi(\xi))$. This in turn implies $\exp(t\psi_n(\xi)) \rightarrow \exp(t\psi(\xi))$ for every $t \geq 0$, which yields weak convergence of $\delta_0 e^{t\mathcal{A}_n}$ to $\delta_0 e^{t\mathcal{A}}$ for all t .

Let us now establish the weak compactness of for every bounded sequence in $\mathcal{G}_2^\Lambda(\mathbb{R}^d)$.

Proposition 6.5. *Let $\{\mathcal{A}_n\}_n$ be a bounded sequence in $(\mathcal{G}_2^\Lambda(\mathbb{R}^d), \mathcal{W}_G)$. Then there exists a subsequence $\{\mathcal{A}_{n_k}\}_k$ and $\mathcal{A} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$ such that $\mathcal{A}_{n_k} \rightarrow \mathcal{A}$ weakly in the sense of transition kernels.*

Proof. Since the sequence is bounded, we may assume without loss of generality that

$$\{\mathcal{A}_n\}_n \subset B_R(0) := \left\{ \mathcal{A} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d) : \mathcal{W}_G(\mathcal{A}, 0) < R \right\},$$

where $\mathcal{A} = 0$ denotes the trivial generator (with $e^{t\mathcal{A}} = I$).

Let $\rho_t^{(n)} := \delta_0 e^{t\mathcal{A}_n}$ be the law at time t of the process generated by $\mathcal{A}_n$. By Proposition 6.2,

$$\mathcal{W}_2(\rho_t^{(n)}, \delta_0)^2 \leq \max\{t, t^2\} \mathcal{W}_G(\mathcal{A}_n, 0)^2 \leq \max\{t, t^2\} R^2. \quad (6.3)$$

Hence the second moments $\int_{\mathbb{R}^d} |x|^2 d\rho_t^{(n)}(x)$ are uniformly bounded (see Proposition 3.15(v)). Fix $t = 1$. By Prokhorov's theorem, there exists a subsequence $\{\rho_1^{(n_k)}\}_k$ converging weakly to some probability measure $\rho_1 \in \mathcal{P}_2(\mathbb{R}^d)$. Note the weak limit ρ_1 has finite second moment due to Portmanteau lemma and the uniform bound of second moment of the sequence.

It is well known that the class of infinitely divisible measures is closed under weak convergence, and every such measure corresponds to the distribution of a Lévy process. Therefore, there exists $\mathcal{A} \in \mathcal{G}^\Lambda(\mathbb{R}^d)$ such that $\rho_1 = \delta_0 e^{\mathcal{A}}$. Moreover, since $\rho_1 \in \mathcal{P}_2(\mathbb{R}^d)$, the associated generator $\mathcal{A}$ necessarily belongs to $\mathcal{G}_2^\Lambda(\mathbb{R}^d)$. Thus we have $\rho_t^{(n_k)} \rightarrow \rho_t$ weakly for $t = 1$. By Remark 6.4, the weak convergence holds for all $t \geq 0$, that is,

$$\rho_t^{(n_k)} = \delta_0 e^{t\mathcal{A}_{n_k}} \rightarrow \delta_0 e^{t\mathcal{A}} = \rho_t \quad \text{for all } t \geq 0.$$

Finally, by translation invariance of Lévy semigroups, the same convergence holds starting from any $x \in \mathbb{R}^d$. Thus $\mathcal{A}_{n_k} \rightarrow \mathcal{A}$ weakly in the sense of transition kernels. $\square$

The coming proposition states the lower semicontinuity of the metric w.r.t. the weak topology on generators introduced earlier.

Proposition 6.6. *Let $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$ and $\{\mathcal{A}_n, \mathcal{B}_n\}_n$ be bounded in $\mathcal{W}_G$ metric. Suppose $\mathcal{A}_n \rightarrow \mathcal{A}, \mathcal{B}_n \rightarrow \mathcal{B}$ weakly in the sense of transition kernels. It holds*

$$\mathcal{W}_G(\mathcal{A}, \mathcal{B}) \leq \liminf_{n \rightarrow \infty} \mathcal{W}_G(\mathcal{A}_n, \mathcal{B}_n).$$

Proof. Let us denote $\kappa_n = m_{\mathcal{A}_n}$, $\zeta_n = m_{\mathcal{B}_n}$, $\kappa = m_{\mathcal{A}}$, $\zeta = m_{\mathcal{B}}$, and define $\rho_t^{(n)} := \delta_0 e^{t\mathcal{A}_n}$, $\sigma_t^{(n)} := \delta_0 e^{t\mathcal{B}_n}$, $\rho_t := \delta_0 e^{t\mathcal{A}}$, and $\sigma_t := \delta_0 e^{t\mathcal{B}}$.

Since $\{\mathcal{A}_n, \mathcal{B}_n\}_n$ is bounded in $\mathcal{G}_2^\Lambda(\mathbb{R}^d)$, the same computation from (6.3) gives: for some $R \geq 0$,

$$\frac{1}{2} \int_{\mathbb{R}^d} |x|^2 d\rho_t^{(n)}(x) = \mathcal{W}_2(\rho_t^{(n)}, \delta_0)^2 \leq \max\{t, t^2\} R^2.$$

The same bound applies for $\sigma_t^{(n)}$ in place of $\rho_t^{(n)}$. This establishes a pointwise (in time) uniform second moment bound for the families $\{\rho_t^{(n)}\}_n$ and $\{\sigma_t^{(n)}\}_n$. By Jensen's inequality, this also implies that the sequence of first moments $\{\kappa_n, \zeta_n\}_n$ is uniformly bounded in $\mathbb{R}^d$.

Now, again using Proposition 6.2, we have for all $t \geq 0$,

$$\mathcal{C}_2(\rho_t^{(n)}, \sigma_t^{(n)}) \leq t\theta_2(0, 0; \mathcal{A}_n, \mathcal{B}_n) + \frac{1}{2}t^2|\kappa_n - \zeta_n|^2 \leq t\theta_2(0, 0; \mathcal{A}_n, \mathcal{B}_n) + Ct^2.$$

Since $\mathcal{A}_n \rightarrow \mathcal{A}$ and $\mathcal{B}_n \rightarrow \mathcal{B}$ weakly in the sense of transition kernels, we have weak convergence $\rho_t^{(n)} \rightarrow \rho_t$ and $\sigma_t^{(n)} \rightarrow \sigma_t$ weakly as $n \rightarrow \infty$. By the lower semicontinuity of the quadratic transport cost $\mathcal{C}_2$, it follows that

$$\mathcal{C}_2(\rho_t, \sigma_t) \leq \liminf_{n \rightarrow \infty} \mathcal{C}_2(\rho_t^{(n)}, \sigma_t^{(n)}) \leq t \liminf_{n \rightarrow \infty} \theta_2(0, 0; \mathcal{A}_n, \mathcal{B}_n) + Ct^2.$$

Dividing both sides by t and letting $t \searrow 0$, we obtain

$$\omega_2^+(0, 0; \mathcal{A}, \mathcal{B}) = \theta_2(0, 0; \mathcal{A}, \mathcal{B}) \leq \liminf_{n \rightarrow \infty} \theta_2(0, 0; \mathcal{A}_n, \mathcal{B}_n).$$

Next, we claim that $\kappa_n \rightarrow \kappa$ and $\zeta_n \rightarrow \zeta$ in $\mathbb{R}^d$. Indeed, since $\kappa_n = \langle \rho_1^{(n)}, x \rangle$ and $\rho_1^{(n)} \rightarrow \rho_1$ weakly with uniform second moment bounds, it follows that

$$\kappa_n = \int_{\mathbb{R}^d} x d\rho_1^{(n)}(x) \rightarrow \int_{\mathbb{R}^d} x d\rho_1(x) = \kappa.$$

An analogous argument shows $\zeta_n \rightarrow \zeta$.

Combining the convergence $(\kappa_n, \zeta_n) \rightarrow (\kappa, \zeta)$ with the lower semicontinuity result above, we conclude that

$$\begin{aligned} \mathcal{W}_{\mathcal{G}}(\mathcal{A}, \mathcal{B})^2 &= \frac{1}{2}|\kappa - \zeta|^2 + \omega_2(0, 0; \mathcal{A}, \mathcal{B}) \\ &\leq \liminf_{n \rightarrow \infty} \left[\frac{1}{2}|\kappa_n - \zeta_n|^2 + \theta_2(0, 0; \mathcal{A}_n, \mathcal{B}_n) \right] = \liminf_{n \rightarrow \infty} \mathcal{W}_{\mathcal{G}}(\mathcal{A}_n, \mathcal{B}_n)^2. \quad \square \end{aligned}$$

Now the completeness of the Wasserstein generator metric $\mathcal{W}_{\mathcal{G}}$ follows easily from the two results above.

Proof of Theorem 2.33, completeness. Let $\{\mathcal{A}_n\}_n \subset \mathcal{G}_2^{\Lambda}(\mathbb{R}^d)$ be a Cauchy sequence with respect to the metric $\mathcal{W}_{\mathcal{G}}$. Since $\{\mathcal{A}_n\}_n$ is bounded, by the weak compactness established in Proposition 6.5, along a subsequence we have $\mathcal{A}_{n_k} \rightarrow \mathcal{A}$ weakly in the sense of transition kernels, where $\mathcal{A} \in \mathcal{G}_2^{\Lambda}(\mathbb{R}^d)$. It remains to show $\mathcal{A}_n \rightarrow \mathcal{A}$ in the metric $\mathcal{W}_{\mathcal{G}}$.

Let us first prove $\mathcal{A}_n \rightarrow \mathcal{A}$ weakly in the sense of transition kernel in the full sequence. It suffices to prove: for every subsequence $\{\mathcal{A}_n\}_n$ there is a further subsequence that converges to $\mathcal{A}$ weakly. Indeed, for any subsequence of $\{\mathcal{A}_n\}_n$ by the weak compactness in Proposition 6.5 there is a further subsequence, indexed by $\{\mathcal{A}_{m_k}\}_k$ that converges weakly to some $\mathcal{A}' \in \mathcal{G}_2^{\Lambda}(\mathbb{R}^d)$. Since $\mathcal{A}_{n_k} \rightarrow \mathcal{A}$ along the subsequence $\{n_k\}$, by the lower semicontinuity of the metric (Proposition 6.6) we have

$$\mathcal{W}_{\mathcal{G}}(\mathcal{A}, \mathcal{A}') \leq \liminf_{n \rightarrow \infty} \mathcal{W}_{\mathcal{G}}(\mathcal{A}_{m_k}, \mathcal{A}_{n_k}) = 0.$$

The limit is zero because $\{\mathcal{A}_n\}_n$ is Cauchy in the metric $\mathcal{W}_{\mathcal{G}}$. This shows the weak convergence of the full sequence $\{\mathcal{A}_n\}_n$ to $\mathcal{A}$.

Finally we show $\mathcal{W}_{\mathcal{G}}(\mathcal{A}, \mathcal{A}_n) \rightarrow 0$. Since $\mathcal{A}_m \rightarrow \mathcal{A}$ weakly, the lower semicontinuity again implies

$$\mathcal{W}_{\mathcal{G}}(\mathcal{A}, \mathcal{A}_n) \leq \liminf_{m \rightarrow \infty} \mathcal{W}_{\mathcal{G}}(\mathcal{A}_m, \mathcal{A}_n) \leq \sup_{m \geq n} \mathcal{W}_{\mathcal{G}}(\mathcal{A}_m, \mathcal{A}_n).$$

Since $\{\mathcal{A}_n\}_n$ is Cauchy, the right hand side vanishes as $n \rightarrow \infty$. Hence it follows $\mathcal{W}_{\mathcal{G}}(\mathcal{A}, \mathcal{A}_n) \rightarrow 0$. The proof is complete. $\square$

6.3. Lévy-Wasserstein metric and its topological properties. Let us now restrict our attention to the jump part of generators. In which case the generator metric reduces to a metric on the space $\Lambda_2(\mathbb{R}^d)$ of Lévy measures with finite second moment, which will be called *Lévy-Wasserstein metric*, denoted as $\mathcal{W}_\Lambda$. The metric property of $\mathcal{W}_\Lambda$ follows easily from Theorem 2.33.

Proof of Corollary 2.34. Let $\mu, \nu \in \Lambda_2(\mathbb{R}^d)$, and set $\mathcal{A} = \Lambda_{\text{glob}}(0, 0, \mu)$ and $\mathcal{B} = \Lambda_{\text{glob}}(0, 0, \nu)$. By Proposition 4.7, we have

$$\mathcal{W}_\Lambda(\mu, \nu)^2 = \mathcal{C}_2^\Lambda(\mu, \nu) = \theta_2(0, 0; \mathcal{A}, \mathcal{B}) = \mathcal{W}_\mathcal{G}(\mathcal{A}, \mathcal{B})^2.$$

Since $\mathcal{W}_\mathcal{G}$ defines a metric, it follows that $\mathcal{W}_\Lambda$ also defines a metric. $\square$

Remark 6.7. In the classical theory of Wasserstein metrics, the triangle inequality is usually established by constructing a coupling of three probability measures, often referred to as the *gluing lemma*. By contrast, in the present Lévy setting, such a gluing construction is less transparent and technically more involved. Instead, our approach exploits the strong duality established earlier, which allows us to derive the triangle inequality through analytic arguments on generators rather than couplings.

We next establish the following equivalent characterization of convergence in the metric $\mathcal{W}_\Lambda$, analogous to the standard characterization of convergence in the Wasserstein-2 metric. We recall the reader the notion of Λ -weak convergence of Lévy measures from Section 3.4.1, see also Appendix C for the related results.

Theorem 6.8. *For $n \geq 1$, let $\mu, \mu_n \in \Lambda_2(\mathbb{R}^d)$. The following are equivalent:*

- (i) $\mathcal{W}_\Lambda(\mu_n, \mu) \rightarrow 0$;
- (ii) for all Lipschitz functions $\varphi \in C(\mathbb{R}^d)$ satisfying $|\nabla \varphi(x)| \leq C|x|$ Lebesgue-a.e. for some $C \geq 0$, it holds

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} \varphi d\mu_n = \int_{\mathbb{R}^d} \varphi d\mu;$$

- (iii) $\mu_n \rightarrow \mu$ Λ -weakly, and

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} |x|^2 d\mu_n = \int_{\mathbb{R}^d} |x|^2 d\mu.$$

Proof. (i) $\rightarrow$ (ii). First a bounded Lipschitz functions φ as in (ii). Using the fundamental theorem of calculus for absolutely continuous functions and the bound for gradient, we have

$$\begin{aligned} |\varphi(x) - \varphi(y)| &= \left| \int_0^1 (x - y) \cdot \nabla \varphi(sx + (1-s)y) ds \right| \\ &\leq C \int_0^1 |x - y|(s|x| + (1-s)|y|) ds \leq C|x - y|(|x| + |y|). \end{aligned}$$

For $n \geq 1$, let $\gamma_n \in \Gamma^\Lambda(\mu_n, \mu)$ be an optimal coupling realizing $\mathcal{W}_\Lambda(\mu_n, \mu)^2 = \int c_2 d\gamma_n$. Using the Cauchy-Schwartz inequality, $(a + b)^2 \leq 2(a^2 + b^2)$, and the bound above we find

$$\left| \int \varphi d\mu_n - \int \varphi d\mu \right| = \left| \int_{\mathbb{R}^{2d}} [\varphi(x) - \varphi(y)] d\gamma_n(x, y) \right|$$

$$\begin{aligned}
&\leq C \int_{\mathbb{R}^{2d}} |x - y|(|x| + |y|) d\gamma_n(x, y) \\
&\leq C \left[\int_{\mathbb{R}^{2d}} |x - y|^2 d\gamma_n(x, y) \right]^{1/2} \cdot \sqrt{2} \left[\int_{\mathbb{R}^{2d}} (|x|^2 + |y|^2) d\gamma_n(x, y) \right]^{1/2} \\
&= 2C\mathcal{W}_\Lambda(\mu_n, \mu) \left[\int_{\mathbb{R}^d} |x|^2 d\mu_n(y) + \int_{\mathbb{R}^d} |y|^2 d\mu(y) \right]^2.
\end{aligned}$$

Since $\{\mu_n\}_n$ is bounded in $\mathcal{W}_\Lambda$, their second moment is uniformly bounded. Passing $n \rightarrow \infty$, we find $\int \varphi d\mu_n \rightarrow \int \varphi d\mu$ for all such φ .

Next we show the second moment converges: $\int_{\mathbb{R}^d} \frac{1}{2}|x|^2 d\mu_n \rightarrow \int_{\mathbb{R}^d} \frac{1}{2}|x|^2 d\mu$. We observe that for any Lévy measure $\nu \in \Lambda_2(\mathbb{R}^d)$, $\int_{\mathbb{R}^d} \frac{1}{2}|x|^2 d\nu(x) = \mathcal{W}_\Lambda(\nu, 0)^2$. Since $\mu_n \rightarrow \mu$ in $\mathcal{W}_\Lambda$, and $\nu \mapsto \mathcal{W}_\Lambda(\nu, 0)$ is continuous w.r.t. the metric, the convergence of the second moment holds from here.

(ii) $\rightarrow$ (iii). $\mu_n \xrightarrow{\Lambda} \mu$ follows from Proposition C.4. We note $\varphi(x) = |x|^2$ satisfies $|\nabla \varphi(x)| \leq 2|x|$. Hence (ii) implies the convergence of second moment.

(iii) $\rightarrow$ (i). For $n \geq 1$, let γ_n be a 2-Lévy optimal coupling of μ_n, μ . We first verify that the sequence $\{\gamma_n\}_n$ is Λ -tight. Since $\mu_n \xrightarrow{\Lambda} \mu$, the sequence $\{\mu_n\}_n \subset \Lambda_2(\mathbb{R}^d)$ is Λ -tight (see Lemma C.5). By Lemma 3.19, $\{\gamma_n\}_n \subset \Lambda_2(\mathbb{R}^{2d})$ is Λ -tight.

By Prokhorov's theorem (Theorem C.7), there exists a subsequence $\{\gamma_{n_k}\}_k$ and a Lévy measure $\gamma \in \Lambda(\mathbb{R}^{2d})$ such that $\gamma_{n_k} \xrightarrow{\Lambda} \gamma$. We now show that γ is a coupling of μ with itself. Since the marginals of γ_{n_k} are μ_{n_k}, μ , and $\mu_{n_k} \xrightarrow{\Lambda} \mu$, the marginal projections of γ_{n_k} converge in the Λ -sense to μ . By the continuity of pushforward under weak convergence, the marginals of γ are both equal to μ . That is, $\gamma \in \Gamma^\Lambda(\mu, \mu)$.

We now claim that the second moments of γ_{n_k} converge to that of γ . Indeed, we have:

$$\int_{\mathbb{R}^{2d}} (|x|^2 + |y|^2) d\gamma_n(x, y) = \int_{\mathbb{R}^d} |x|^2 d\mu_n(x) + \int_{\mathbb{R}^d} |y|^2 d\mu(y),$$

and the right-hand side converges by assumption. By Proposition C.6, we have

$$\mathcal{W}_\Lambda(\mu_{n_k}, \mu) = \int_{\mathbb{R}^{2d}} \frac{1}{2}|x - y|^2 d\gamma_{n_k}(x, y) \rightarrow \int_{\mathbb{R}^{2d}} \frac{1}{2}|x - y|^2 d\gamma(x, y) = 0.$$

Since every subsequence admits a further subsequence along which convergence to zero holds, the full sequence must also converge: $\mathcal{W}_\Lambda(\mu_n, \mu) \rightarrow 0$. $\square$

6.4. Separability of the Wasserstein generator metric. We now address the separability of the metric $\mathcal{W}_\mathcal{G}$. By the decomposition in (6.2), this metric is the direct sum of three parts: the Euclidean metric on $\mathbb{R}^d$, the Bures–Wasserstein metric on $\mathcal{S}_{\geq 0}(\mathbb{R}^d)$, and the Lévy–Wasserstein metric on $\Lambda_2(\mathbb{R}^d)$. Hence, to establish the separability of $(\mathcal{G}_2^\Lambda(\mathbb{R}^d), \mathcal{W}_\mathcal{G}^\Lambda)$, it suffices to show the separability of each component.

The Euclidean space $\mathbb{R}^d$ is separable, as is $\mathcal{S}_{\geq 0}(\mathbb{R}^d)$ under the Bures–Wasserstein metric (as a closed subset of the space of symmetric matrices with respect to the Frobenius norm). The remaining part is handled by the following result:

Proposition 6.9. *The metric space $(\Lambda_2(\mathbb{R}^d), \mathcal{W}_\Lambda)$ is separable.*

Proof. Fix $r > 0$, and define $\mathcal{M}_r \subset \Lambda_2(\mathbb{R}^d)$ to be the set of Lévy measures with total mass $\mu(\mathbb{R}^d) = r$. We first claim that $(\mathcal{M}_r, \mathcal{W}_\Lambda)$ is separable.

Indeed, since $\mathcal{M}_r$ consists of finite measures with fixed total mass, it is a subset of the space of probability measures (up to scaling) with finite second moment. The classical 2-Wasserstein space over $\mathbb{R}^d$ with fixed total mass r is known to be separable. Moreover, we have $\mathcal{W}_\Lambda \leq \mathcal{W}_2$ from Proposition 3.16, that is, the Lévy-Wasserstein metric is dominated by the standard Wasserstein-2 metric. Therefore, the separability of $(\mathcal{M}_r, \mathcal{W}_\Lambda)$ follows from that of $(\mathcal{M}_r, \mathcal{W}_2)$.

Now define

$$\mathcal{M} := \bigcup_{q \in \mathbb{Q}^+} \mathcal{M}_q.$$

This is a countable union of separable subsets, hence separable.

We now show that $\mathcal{M}$ is dense in $\Lambda_2(\mathbb{R}^d)$. Let $\mu \in \Lambda_2(\mathbb{R}^d)$ and fix $\varepsilon > 0$. Choose small $\delta, \delta' > 0$, and define

$$d\mu_{\delta, \delta'} := (1 - \delta')\chi_{B_\delta(0)^c}(x) d\mu(x),$$

which is a finite Lévy measure. Let $\zeta_{\delta, \delta'} := \mu - \mu_{\delta, \delta'}$, so that $\mu = \mu_{\delta, \delta'} + \zeta_{\delta, \delta'}$. By Proposition 3.15,

$$\begin{aligned} \mathcal{W}_\Lambda(\mu, \mu_{\delta, \delta'})^2 &\leq \mathcal{W}_\Lambda(\mu_{\delta, \delta'}, \mu_{\delta, \delta'})^2 + \mathcal{W}_\Lambda(0, \zeta_{\delta, \delta'})^2 \\ &= \frac{1}{2} \left(\int_{B_\delta(0)} |x|^2 d\mu(x) + \delta' \int_{B_\delta(0)^c} |x|^2 d\mu(x) \right). \end{aligned}$$

Because $\mu \in \Lambda_2(\mathbb{R}^d)$, this expression can be made arbitrarily small by taking δ, δ' sufficiently small. Thus, we may choose δ, δ' such that

$$\mathcal{W}_\Lambda(\mu, \mu_{\delta, \delta'}) < \varepsilon.$$

Note that $\mu_{\delta, \delta'} \in \mathcal{M}_r$ for $r = (1 - \delta')\mu(B_\delta(0)^c)$. By the density of the rational numbers, we may also choose $\delta' > 0$ small enough so that $r \in \mathbb{Q}$, namely $\mu_{\delta, \delta'} \in \mathcal{M}$, while ensuring the error estimate above still holds. This completes the proof. $\square$

We may now complete the proof of Theorem 2.33 .

Proof of Theorem 2.33, separability. The squared generator metric $\mathcal{W}_\mathcal{G}^2$ is the sum of three components: the squared Euclidean distance on $\mathbb{R}^d$, the squared Bures–Wasserstein distance $\mathcal{W}_\mathcal{S}^2$ on the cone of nonnegative definite matrices, and the squared Lévy–Wasserstein distance $\mathcal{W}_\Lambda^2$ on $\Lambda_2(\mathbb{R}^d)$. Each of these metric spaces is known to be separable. Since a finite product of separable metric spaces remains separable, it follows that $(\mathcal{G}_2^\Lambda(\mathbb{R}^d), \mathcal{W}_\mathcal{G})$ is separable. $\square$

6.5. Pathwise estimate for coupled Lévy processes. We shall next establish a maximal-type inequality for the optimal coupling of two Lévy processes in terms of the metric $\mathcal{W}_\mathcal{G}$.

Proof of Theorem 2.35. Let us first prove the case when the two processes has zero mean. In this case, the process $\{Z_t = (X_t, Y_t)\}_{t \geq 0}$ generated by a 2-optimal Lévy coupling $\mathcal{J}_*$ of $\mathcal{A}, \mathcal{B}$ has zero mean. Particularly, it is a vector martingale in $\mathbb{R}^{2d}$.

Define the symmetric matrix

$$U = \begin{bmatrix} I & -I \\ -I & I \end{bmatrix}.$$

Since $U \geq 0$, there exists a symmetric positive semidefinite matrix Q such that $Q^2 = Q^\top Q = U$. Then

$$|X_t - Y_t|^2 = Z_t^\top U Z_t = |Q Z_t|^2.$$

Because $\{Z_t\}$ is a vector martingale, so is $\{Q Z_t\}$. By Doob's L^2 -maximal inequality,

$$\begin{aligned} \mathbb{E} \left[\sup_{t \in [0, T]} |Q Z_t|^2 \right] &\leq 4 \mathbb{E} |Q Z_T|^2 = 2 \mathbb{E} |X_T - Y_T|^2 \\ &= 4 (e^{T\mathcal{J}^* c_2}) (0, 0) = 4T \theta(0, 0; \mathcal{A}, \mathcal{B}) = 4T \mathcal{W}_G(\mathcal{A}, \mathcal{B})^2, \end{aligned}$$

where the final equality uses Theorem 2.31(iii). This proves the estimate in the zero-mean case.

Now consider general Lévy processes $\{X_t\}, \{Y_t\}$ with potentially nonzero means. Write $X_t = \bar{X}_t + tm_{\mathcal{A}}, Y_t = \bar{Y}_t + tm_{\mathcal{B}}$, where $\{\bar{X}_t\}, \{\bar{Y}_t\}$ are the centered Lévy processes with generators $\bar{\mathcal{A}}, \bar{\mathcal{B}}$. Note that

$$\mathcal{W}_G^\Lambda(\mathcal{A}, \mathcal{B}) = \theta(0, 0; \bar{\mathcal{A}}, \bar{\mathcal{B}}) = \theta(0, 0; \mathcal{A}, \mathcal{B}).$$

Using the triangle inequality and the elementary bound $|a + b|^2 \leq 2(|a|^2 + |b|^2)$, we obtain:

$$\begin{aligned} \mathbb{E}^{(0,0)} \left[\sup_{t \in [0, T]} |X_t - Y_t|^2 \right] &\leq 2 \mathbb{E}^{(0,0)} \left[\sup_{t \in [0, T]} |\bar{X}_t - \bar{Y}_t|^2 \right] + 2|Tm_{\mathcal{A}} - Tm_{\mathcal{B}}|^2 \\ &\leq 8T \theta(0, 0; \mathcal{A}, \mathcal{B}) + 2T^2 |m_{\mathcal{A}} - m_{\mathcal{B}}|^2 \\ &\leq 8 \max\{T, T^2\} \mathcal{W}_G^\Lambda(\mathcal{A}, \mathcal{B})^2. \end{aligned} \quad \square$$

APPENDIX A. APPROXIMATION OF ADMISSIBLE PAIRS

In this appendix, we construct smooth approximations of admissible test function pairs, as required in the proofs of Lemmas 3.26 and 5.3. Throughout, we fix two Lévy measures $\mu, \nu \in \Lambda_2(\mathbb{R}^d)$ with finite second moments. By an *admissible pair*, we mean any lower semicontinuous functions $(\varphi, \psi) \in L^1(\mu) \oplus L^1(\nu)$ such that

$$\varphi(x) + \psi(y) \leq c_2(x, y) := \frac{1}{2}|x - y|^2, \quad \text{for all } x, y \in \mathbb{R}^d.$$

Let $\mathcal{F} = \mathcal{F}(\mu, \nu)$ denote the collection of all admissible pairs. To prove Lemma 3.26, we introduce the following subclasses of $\mathcal{F}$:

- $\mathcal{F}_0$: pairs (φ, ψ) vanishing in a neighborhood of the origin and satisfying the bound

$$|\varphi(x)|, |\psi(x)| \leq C \min\{1, |x|^2\}, \quad \text{for some } C \geq 0.$$

Particularly φ, ψ are bounded lower semicontinuous.

- $\mathcal{F}_0^1$: elements of $\mathcal{F}_0$ with φ, ψ bounded Lipschitz;
- $\mathcal{F}_0^2$: elements of $\mathcal{F}_0^1$ with φ, ψ in $C_b^2(\mathbb{R}^d)$.

These classes satisfy the natural inclusions:

$$\mathcal{F}_0^2 \subset \mathcal{F}_0^1 \subset \mathcal{F}_0 \subset \mathcal{F}.$$

Lemma 3.26 follows as a consequence of the density of $\mathcal{F}_0^2$ in $\mathcal{F}$.

A.1. Density of $\mathcal{F}_0 \subset \mathcal{F}$. Given a pair of lower semicontinuous functions $f, g: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{-\infty\}$ with $f \leq g$, and a function $\varphi: \mathbb{R}^d \rightarrow \mathbb{R}$, we define the *cutoff* function:

$$\hat{\varphi} = \chi(\varphi; f, g)(x) := \max\{f(x), \min\{\varphi(x), g(x)\}\} = \begin{cases} g(x) & \text{if } \varphi(x) \geq g(x), \\ \varphi(x) & \text{if } f(x) < \varphi(x) < g(x), \\ f(x) & \text{if } \varphi(x) \leq f(x). \end{cases}$$

Clearly, the cutoff satisfies $f \leq \hat{\varphi} \leq g$.

Lemma A.1. (a) Let $f \leq g$ with $(f, g) \in \mathcal{F}$. If $(\varphi, \psi) \in \mathcal{F}$, then the pair $(\hat{\varphi}, \hat{\psi})$ also belongs to $\mathcal{F}$, where

$$\hat{\varphi} := \chi(\varphi; f, g), \quad \hat{\psi} := \chi(\psi; f, g).$$

(b) If $(\varphi, \psi) \in \mathcal{F}$ and $T: \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a contraction (i.e., 1-Lipchitz). Suppose $\varphi \circ T \in L^1(\mu), \psi \circ T \in L^1(\nu)$. Then $(\varphi \circ T, \psi \circ T) \in \mathcal{F}$.

Proof. (a) Since lower semicontinuity is preserved under pointwise maxima and minima, and since f, g, φ, ψ are all lower semicontinuous, it follows that $\hat{\varphi}$ and $\hat{\psi}$ are lower semicontinuous as well. Moreover, since $f, g \in L^1(\mu)$ and $L^1(\nu)$, and $\hat{\varphi}, \hat{\psi}$ are bounded by f and g , we also have $\hat{\varphi} \in L^1(\mu), \hat{\psi} \in L^1(\nu)$.

It remains to verify that $\hat{\varphi}(x) + \hat{\psi}(y) \leq c_2(x, y)$ for all $x, y \in \mathbb{R}^d$, where $c_2(x, y) := \frac{1}{2}|x - y|^2$. To this end, we distinguish two cases. If either $\hat{\varphi}(x) = f(x)$ or $\hat{\psi}(y) = f(y)$, say the former without loss of generality, then by the bounds $\hat{\psi}(y) \leq g(y)$ and $(f, g) \in \mathcal{F}$, we obtain

$$\hat{\varphi}(x) + \hat{\psi}(y) \leq f(x) + g(y) \leq c_2(x, y).$$

In the complementary case where both $\hat{\varphi}(x) \neq f(x)$ and $\hat{\psi}(y) \neq f(y)$, it follows from the definition of the cutoff that $\hat{\varphi}(x) = \min\{\varphi(x), g(x)\}$ and $\hat{\psi}(y) = \min\{\psi(y), g(y)\}$. In particular, we then have

$$\hat{\varphi}(x) + \hat{\psi}(y) \leq \varphi(x) + \psi(y) \leq c_2(x, y).$$

In either case, the inequality $\hat{\varphi}(x) + \hat{\psi}(y) \leq c_2(x, y)$ holds, as desired.

(b) Since T is continuous, and (φ, ψ) is lower semicontinuous, so is $(\varphi \circ T, \psi \circ T)$. From the assumption, the pair is in $L^1(\mu) \oplus L^1(\nu)$. Moreover it holds for all $x, y \in \mathbb{R}^d$:

$$\varphi(T(x)) + \psi(T(y)) \leq \frac{1}{2}|T(x) - T(y)|^2 \leq \frac{1}{2}|x - y|^2.$$

Hence $(\varphi \circ T, \psi \circ T) \in \mathcal{F}$. □

To complete the proof of the density of $\mathcal{F}_0 \subset \mathcal{F}$, we invoke Lemma A.1 and construct a sequence $\{(f_n, g_n)\}_{n \in \mathbb{N}}$ of bounded continuous pairs, each vanishing on a neighborhood of the origin, such that $g_n \nearrow \frac{1}{2}|x|^2, f_n \searrow \chi_\infty$, where

$$\chi_\infty(x) = \begin{cases} 0, & x = 0, \\ -\infty, & x \neq 0. \end{cases}$$

Lemma A.2. *There exists a sequence $\{(f_n, g_n)\}_{n \in \mathbb{N}}$ of pairs of bounded Lipschitz functions satisfying:*

- $f_n \leq 0 \leq g_n$ and $(f_n, g_n) \in \mathcal{F}$ for all n ;
- f_n and g_n vanish on a neighborhood of 0;
- $g_n(x) \nearrow \frac{1}{2}|x|^2$ and $f_n(x) \searrow \chi_\infty(x)$ pointwise as $n \rightarrow \infty$.

Proof. We use the following standard construction. For $\delta \in (0, 1)$, set

$$f_\delta(x) = \frac{1}{2}(1 - \delta^{-1})|x|^2, \quad g_\delta(y) = \frac{1}{2}(1 - \delta)|y|^2.$$

Then (f_δ, g_δ) is a c_2 -concave dual pair with $f_\delta \leq 0 \leq g_\delta$, in particular satisfying $f_\delta \oplus g_\delta \leq c_2$. For $\varepsilon > 0$, define $\rho_\varepsilon : [0, \infty) \rightarrow [0, \infty)$ by

$$\rho_\varepsilon(r) = \begin{cases} 0, & r \in [0, \varepsilon], \\ r - \varepsilon, & r \in (\varepsilon, \frac{1}{\varepsilon} - \varepsilon), \\ \frac{1}{\varepsilon}, & r \geq \frac{1}{\varepsilon} - \varepsilon. \end{cases}$$

Let $T_\varepsilon(x) = \frac{x}{|x|}\rho_\varepsilon(|x|)$. For each $\varepsilon > 0$, the map T_ε is a contraction and $T_\varepsilon(x) \rightarrow x$ locally uniformly as $\varepsilon \searrow 0$.

Define

$$f_{\delta, \varepsilon}(x) := f_\delta(T_\varepsilon(x)), \quad g_{\delta, \varepsilon}(y) := g_\delta(T_\varepsilon(y)).$$

From the construction we have $f_{\delta, \varepsilon} \leq 0 \leq g_{\delta, \varepsilon}$; each pair $(f_{\delta, \varepsilon}, g_{\delta, \varepsilon})$ is bounded, Lipschitz, and vanishes on $B_\varepsilon(0)$. Also, we have $(f_{\delta, \varepsilon}, g_{\delta, \varepsilon}) \in \mathcal{F}$, by Lemma A.1. Moreover, $(\delta, \varepsilon) \mapsto f_{\delta, \varepsilon}$ decreases pointwise to χ_∞ , while $(\delta, \varepsilon) \mapsto g_{\delta, \varepsilon}$ increases pointwise to $\frac{1}{2}|x|^2$ as $(\delta, \varepsilon) \searrow (0, 0)$. Choosing a sequence $(\delta_n, \varepsilon_n) \searrow (0, 0)$ and setting $f_n := f_{\delta_n, \varepsilon_n}, g_n := g_{\delta_n, \varepsilon_n}$, yields the desired sequence. $\square$

Lemma A.3. $\mathcal{F}_0 \subset \mathcal{F}$ is dense w.r.t. the topology of $L^1(\mu) \oplus L^1(\nu)$.

Proof. Choose $(\varphi, \psi) \in \mathcal{F}$, and our goal is to construct a sequence $\{(\varphi_n, \psi_n)_n\} \subset \mathcal{F}_0$ such that $(\varphi_n, \psi_n) \rightarrow (\varphi, \psi)$ in $L^1(\mu) \oplus L^1(\nu)$. Let $\{(f_n, g_n)\}_n$ be the sequence given in Lemma A.2, and define the cutoff sequences

$$\varphi_n := \chi(\varphi; f_n, g_n), \quad \psi_n := \chi(\psi; f_n, g_n).$$

By Lemma A.1, $(\varphi_n, \psi_n) \in \mathcal{F}$. Since $\chi_\infty(x) \leq \varphi(x), \psi(x) \leq \frac{1}{2}|x|^2$ and $g_n \nearrow \frac{1}{2}|x|^2, f_n \searrow \chi_\infty(x)$, it follows $|\varphi_n| \nearrow |\varphi|, |\psi_n| \nearrow |\psi|$ pointwise as $n \rightarrow \infty$. The dominated convergence theorem implies $\varphi_n \rightarrow \varphi$ in $L^1(\mu), \psi_n \rightarrow \psi$ in $L^1(\nu)$. $\square$

A.2. Density of smooth admissible test functions (proof of Lemma 3.26). We now show that $\mathcal{F}_0^1$, the subclass of Lipschitz pairs, is dense in $\mathcal{F}$. This follows from the fact that any bounded lower semicontinuous function can be approximated by an increasing sequence of Lipschitz functions. Specifically, given a lower semicontinuous function φ bounded below by $\varphi(x) \geq C$, define

$$\varphi_n(x) := \inf_{y \in \mathbb{R}^d} [\varphi(y) + n|x - y|], \quad (\text{A.1})$$

that is, the infimal convolution of φ with $n|x|$ (see (A.2)). Then φ_n is n -Lipschitz and satisfies $\varphi_n \nearrow \varphi$ pointwise as $n \rightarrow \infty$.

Moreover, the following stability property holds: if f is C -Lipschitz and $\varphi \geq f$, then $\varphi_n \geq f$ for all $n \geq C$. In particular, if φ is bounded and vanishes in a neighbourhood of the origin, then for all sufficiently large n the same holds for φ_n . This property will be useful in establishing a uniform bound in n for the sequence $\{\varphi_n\}_n$.

Lemma A.4. *The set $\mathcal{F}_0^1 \subset \mathcal{F}$ is dense with respect to the topology of $L^1(\mu) \oplus L^1(\nu)$.*

Proof. By Lemma A.3, it suffices to prove that $\mathcal{F}_0^1$ is dense in $\mathcal{F}_0$. Fix $(\varphi, \psi) \in \mathcal{F}_0$, where φ and ψ are bounded, lower semicontinuous, and vanish in a neighbourhood of the origin.

Choose a nonnegative Lipschitz function f that also vanishes in a neighbourhood of the origin and satisfies

$$|\varphi(x)|, |\psi(x)| \leq f(x) \quad \text{for all } x,$$

which is clearly possible. Let (φ_n, ψ_n) be defined as in (A.1). We have $\varphi_n \nearrow \varphi$ and $\psi_n \nearrow \psi$, hence $\varphi_n \oplus \psi_n \leq c_2$. Since f is Lipschitz, the stability property above ensures that $\varphi_n, \psi_n \geq -f$ for all sufficiently large n . Thus for large n we have $|\varphi_n|, |\psi_n| \leq f$, so $(\varphi_n, \psi_n) \in \mathcal{F}_0^1$.

Finally, since f is integrable with respect to both μ and ν , the dominated convergence theorem yields $(\varphi_n, \psi_n) \rightarrow (\varphi, \psi)$ in $L^1(\mu) \oplus L^1(\nu)$, completing the proof. $\square$

Let us now establish the density of $\mathcal{F}_0^2$ in $\mathcal{F}$. The following lemma will be the key ingredient.

Lemma A.5. *Let $q : \mathbb{R}^d \rightarrow [0, \infty)$ be given, and let $\varphi, \psi : \mathbb{R}^d \rightarrow \mathbb{R}$ be lower semicontinuous functions satisfying*

$$\varphi(x) + \psi(y) \leq q(x - y), \quad \text{for all } x, y \in \mathbb{R}^d.$$

*Let $\eta \in L^1(\mathbb{R}^d)$ be a nonnegative mollifier, i.e., $\eta \geq 0$, $\int_{\mathbb{R}^d} \eta(x) dx = 1$. Then the mollified pair $(\eta * \varphi, \eta * \psi)$ also satisfies*

$$(\eta * \varphi)(x) + (\eta * \psi)(y) \leq q(x - y), \quad \text{for all } x, y \in \mathbb{R}^d.$$

Proof. For any $x, y \in \mathbb{R}^d$,

$$(\eta * \varphi)(x) + (\eta * \psi)(y) = \int_{\mathbb{R}^d} [\varphi(x - z) + \psi(y - z)] \eta(z) dz \leq q(x - y),$$

since the integrand is bounded above by $q(x - y)$ pointwise. $\square$

We can now deduce the desired density result.

Lemma A.6. *The set $\mathcal{F}_0^2 \subset \mathcal{F}$ is dense with respect to the topology of $L^1(\mu) \oplus L^1(\nu)$.*

Proof. It suffices to prove that $\mathcal{F}_0^2$ is dense in $\mathcal{F}_0^1$. Fix a C^∞ mollifier η supported on $\overline{B_1(0)}$, and define $\eta_\varepsilon(x) = \varepsilon^{-d}\eta(\varepsilon^{-1}x)$. Given $(\varphi, \psi) \in \mathcal{F}_0^1$, set

$$\varphi_\varepsilon := \eta_\varepsilon * \varphi, \quad \psi_\varepsilon := \eta_\varepsilon * \psi.$$

Since (φ, ψ) vanishes on $B_r(0)$ for some $r > 0$, choosing $\varepsilon \in (0, r)$ ensures that $(\varphi_\varepsilon, \psi_\varepsilon)$ also vanishes on some (possibly smaller) neighborhood of the origin.

By Lemma A.5, the smoothed pair satisfies $\varphi_\varepsilon \oplus \psi_\varepsilon \leq c_2$, hence $(\varphi_\varepsilon, \psi_\varepsilon) \in \mathcal{F}_0^2$. Moreover, since (φ, ψ) is bounded Lipschitz, we have uniform convergence $(\varphi_\varepsilon, \psi_\varepsilon) \rightarrow (\varphi, \psi)$. By the bounded convergence theorem, this implies $\varphi_\varepsilon \rightarrow \varphi$ in $L^1(\mu)$ and $\psi_\varepsilon \rightarrow \psi$ in $L^1(\nu)$. Therefore $\mathcal{F}_0^2$ is dense in $\mathcal{F}_0^1$. $\square$

Proof of Lemma 3.26. Denote $\mathcal{F}_{(iv)}$ the family of all pairs given in Lemma 3.26(iv). We note $\mathcal{F}_0^2 \subset \mathcal{F}_{(iv)} \subset \mathcal{F}$. Since $\mathcal{F}_0^2$ is dense in $\mathcal{F}$ by Lemma A.6, so is $\mathcal{F}_{(iv)}$. $\square$

A.3. Proof of Lemma 5.3. The proof of Lemma 5.3 relies on the approximation via the *infimal convolution*. Given two functions $\chi, g : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$, their *infimal convolution* is defined by

$$(\chi \square g)(x) := \inf_{z \in \mathbb{R}^d} \{\chi(z) + g(x - z)\} = \inf_{z+z'=x} \{\chi(z) + g(z')\}. \quad (\text{A.2})$$

The operator $\square$ is *commutative* in general, but not *associative*.

In what follows, given a lower semicontinuous function $q : \mathbb{R}^d \rightarrow [0, \infty]$, we write

$$c_q(x, y) := q(x - y).$$

Also, for a function $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$, we define the reflection

$$f^\sim(x) := f(-x).$$

Lemma A.7. *Let $\chi, q : \mathbb{R}^d \rightarrow [0, \infty]$ and let $\varphi, \psi : \mathbb{R}^d \rightarrow \mathbb{R}$ be lower semicontinuous functions satisfying $\varphi \oplus \psi \leq c_q$. Define*

$$\hat{\varphi} := \chi \square \varphi, \quad \hat{\psi} := \chi \square \psi, \quad \hat{q} := q \square (\chi \square \chi^\sim), \quad \hat{c}_q(x, y) := \hat{q}(x - y).$$

Then the pair $(\hat{\varphi}, \hat{\psi})$ satisfies $\hat{\varphi} \oplus \hat{\psi} \leq \hat{c}_q$.

Proof. By the definition of the infimal convolution,

$$\begin{aligned} \hat{\varphi}(x) + \hat{\psi}(y) &= \inf_z \{\chi(z) + \varphi(x - z)\} + \inf_{z'} \{\chi(z') + \psi(y - z')\} \\ &\leq \inf_{z, z'} \{\chi(z) + \chi(z') + \varphi(x - z) + \psi(y - z')\} \\ &\leq \inf_{z, z'} \{\chi(z) + \chi(z') + q(x - y - (z - z'))\}. \end{aligned}$$

Setting $w := x - y$ and $u := z - z'$, we rewrite the last expression as

$$\begin{aligned} \inf_{u \in \mathbb{R}^d} \left\{ q(w - u) + \inf_{z - z' = u} [\chi(z) + \chi(z')] \right\} &= \inf_{u \in \mathbb{R}^d} \{q(w - u) + (\chi \square \chi^\sim)(u)\} \\ &= q \square (\chi \square \chi^\sim)(w) = \hat{q}(w). \end{aligned}$$

Thus $\hat{\varphi} \oplus \hat{\psi} \leq \hat{c}_q$, which proves the claim. $\square$

In applications, the infimal convolution is often used to produce approximations from below. Let $\{\chi_\varepsilon\}_{\varepsilon>0}$ be a family of nonnegative convex functions such that $\chi_\varepsilon(0) = 0, \chi_\varepsilon \nearrow \chi_0$ as $\varepsilon \searrow 0$, where

$$\chi_0(x) := \begin{cases} 0 & \text{if } x = 0, \\ \infty & \text{otherwise.} \end{cases}$$

Then, for any lower semicontinuous $f: \mathbb{R}^d \rightarrow \mathbb{R}$, we have

$$\chi_\varepsilon \square f \nearrow f \quad \text{pointwise as } \varepsilon \searrow 0.$$

In the proof of Lemma 5.3, we take

$$\chi_\varepsilon(x) := \begin{cases} 0 & \text{if } x \in \overline{B_\varepsilon(0)}, \\ \infty & \text{otherwise.} \end{cases}$$

If we define $q(x) := \frac{1}{2}|x|^2$, then

$$\hat{q}^\varepsilon := q \square (\chi_{\varepsilon/2} \square \chi_{\varepsilon/2}^\sim) = q \square \chi_\varepsilon$$

can be computed explicitly:

$$\hat{q}^\varepsilon(x) = \frac{1}{2} \max(0, |x| - 2\varepsilon)^2.$$

Proof of Lemma 5.3. By Lemma A.4, $\mathcal{F}_0^1$ is dense in $\mathcal{F}$, so it suffices to prove the result for $(\varphi, \psi) \in \mathcal{F}_0^1$. Fix $\varepsilon > 0$, and define

$$\hat{\varphi}_\varepsilon := \varphi \square \chi_{\varepsilon/2}, \quad \hat{\psi}_\varepsilon := \psi \square \chi_{\varepsilon/2}.$$

Since φ and ψ vanish in some $B_r(0)$, for any $\varepsilon \in (0, r)$ the pair $(\hat{\varphi}_\varepsilon, \hat{\psi}_\varepsilon)$ vanishes on $B_{r/2}(0)$. Moreover, infimal convolution with $\chi_{\varepsilon/2}$ preserves Lipschitz continuity, so $(\hat{\varphi}_\varepsilon, \hat{\psi}_\varepsilon) \in \mathcal{F}_0^1$. By Lemma A.7,

$$\hat{\varphi}_\varepsilon \oplus \hat{\psi}_\varepsilon \leq c_2^\varepsilon. \tag{A.3}$$

Let $\{\eta_{\varepsilon'}\}_{\varepsilon'}$ be the smooth mollifiers from Lemma A.6, with $\text{supp}(\eta_{\varepsilon'}) \subset B_{\varepsilon'}(0)$. Choosing $\varepsilon' = \varepsilon/4$ ensures that $(\varphi_\varepsilon, \psi_\varepsilon) := \eta_{\varepsilon'} * (\hat{\varphi}_\varepsilon, \hat{\psi}_\varepsilon)$ still vanishes on $B_{r/4}(0)$. This yields a C^2 pair vanishing near the origin, satisfying condition (i) of Lemma 5.3.

Condition (ii) follows from (A.3) and Lemma A.5. Finally, since $(\hat{\varphi}_\varepsilon, \hat{\psi}_\varepsilon) \rightarrow (\varphi, \psi)$ uniformly as $\varepsilon \searrow 0$, the mollified sequence $(\varphi_\varepsilon, \psi_\varepsilon)$ also converges uniformly. By the bounded convergence theorem, $(\varphi_\varepsilon, \psi_\varepsilon) \rightarrow (\varphi, \psi)$ in $L^1(\mu) \oplus L^1(\nu)$, proving condition (iii). This completes the proof. $\square$

APPENDIX B. PROOFS OF PROPOSITIONS 2.12, 2.21 AND 2.27

B.1. Equivalent characterization of Feller couplings. Let $(\mathcal{A}, D(\mathcal{A})) \in \mathcal{G}(\Pi)$, and let $\{e^{t\mathcal{A}}\}_{t \geq 0}$ denote the probability semigroup generated by $\mathcal{A}$. Since this semigroup admits a

transition kernel $\{\kappa_t(x)\}_{t \geq 0, x \in \Pi}$, it acts as a family of bounded operators $C_0(\Pi) \rightarrow C_0(\Pi)$ in the sense that, for all $\varphi \in C_0(\Pi)$,

$$(e^{t\mathcal{A}}\varphi)(x) = \int_{\Pi} \varphi(y) \kappa_t(x, dy).$$

The operator $e^{t\mathcal{A}}$, $t \geq 0$, naturally extends to the space of bounded continuous functions $C_b(\Pi)$, or more generally to arbitrary continuous or measurable functions, provided the integral is well-defined. The following properties are immediate:

- Order preservation: If $\varphi_1 \leq \varphi_2$, then $e^{t\mathcal{A}}\varphi_1 \leq e^{t\mathcal{A}}\varphi_2$.
- Conservation: $e^{t\mathcal{A}}1 = 1$.
- Monotone convergence: If $0 \leq \varphi_n \nearrow \varphi$, then $e^{t\mathcal{A}}\varphi_n \nearrow e^{t\mathcal{A}}\varphi$.

For $\lambda > 0$, the λ -resolvent operator of $\mathcal{A}$ is the bounded operator $R_\lambda(\mathcal{A}) : C_0(\Pi) \rightarrow C_0(\Pi)$ defined by

$$R_\lambda(\mathcal{A})\varphi(x) = \int_0^\infty e^{-\lambda s} (e^{s\mathcal{A}}\varphi)(x) ds.$$

Equivalently, $R_\lambda(\mathcal{A})$ can be expressed as an integral operator with the λ -potential kernel

$$R_\lambda(\mathcal{A})\varphi(x) = \int_{\Pi} \varphi(y) U_\lambda(x, dy), \quad U_\lambda(x, dy) = \int_0^\infty e^{-\lambda s} \kappa_s(x, dy) ds.$$

Thus, its definition extends in the same way to general measurable functions, including all $\varphi \in C_b(\Pi)$, whenever the integral is well-defined. As with the semigroup, the resolvent is order preserving and satisfies monotone convergence. In particular, the *scaled resolvent operator*

$$I_\lambda(\mathcal{A}) := \lambda R_\lambda(\mathcal{A})$$

is also order preserving and satisfies the monotone convergence and conservation property $I_\lambda(\mathcal{A})1 = 1$. Moreover, it is well known that $I_\lambda(\mathcal{A})$ approximates the identity as $\lambda \rightarrow \infty$, in the sense that

$$I_\lambda(\mathcal{A})\varphi \rightarrow \varphi, \quad \text{for all } \varphi \in C_0(\Pi).$$

It is well known that $R_\lambda(\mathcal{A})$ is the inverse of $\lambda I - \mathcal{A}$. More precisely, for all $\varphi \in C_0(\Pi)$ and $\psi \in D(\mathcal{A})$,

$$(\lambda I - \mathcal{A})R_\lambda(\mathcal{A})\varphi = \varphi, \quad R_\lambda(\mathcal{A})(\lambda I - \mathcal{A})\psi = \psi.$$

These identities extend further to $\varphi \in C_b(\Pi)$ and $\psi \in \bar{D}(\mathcal{A})$. In particular, the second equality—which will be useful later—follows directly from the fundamental theorem of calculus:

$$\begin{aligned} R_\lambda(\mathcal{A})(\lambda I - \mathcal{A})\psi(x) &= \int_0^\infty e^{-\lambda s} e^{s\mathcal{A}}(\lambda I - \mathcal{A})\psi(x) ds \\ &= \int_0^\infty \frac{d}{ds} \left[e^{-\lambda s} e^{s\mathcal{A}}\psi(x) \right] ds = \left(e^{-\lambda s} e^{s\mathcal{A}}\psi(x) \right) \Big|_{s=0}^\infty = \psi(x). \end{aligned}$$

Here, the equality is understood pointwise in $x \in \Pi$. In particular, the fact that the map $s \mapsto e^{-\lambda s} e^{s\mathcal{A}} \psi(x)$ is differentiable for each $x \in \Pi$ is crucial.

In the coming lemma, we present the *Kato–Trotter type* approximation formula for probability semigroups, which expresses $e^{t\mathcal{A}}$ as a limit involving the scaled resolvent operator $I_\lambda(\mathcal{A})$. The formula in fact holds for general C_0 -semigroups (see [21]), and thus for all $\varphi \in C_0(\Pi)$. Since we require an extension to $\varphi \in C_b(\Pi)$, and we are not aware of an explicit reference in the literature, we include a proof here for completeness.

Lemma B.1. *Let $\mathcal{A} \in \mathcal{G}(\Pi)$, and define $I_\lambda(\mathcal{A}) = \lambda R_\lambda(\mathcal{A})$. Then for every $\varphi \in C_b(\Pi)$ and $t > 0$,*

$$\lim_{n \rightarrow \infty} (I_{n/t}(\mathcal{A}))^n \varphi(x) = (e^{t\mathcal{A}} \varphi)(x), \quad \text{for all } x \in \Pi.$$

Proof. Let us write $S_n := I_{n/t}(\mathcal{A})^n$. By the conservation property of both $I_\lambda(\mathcal{A})$ and $e^{t\mathcal{A}}$, we have for any $\alpha, c \in \mathbb{R}$ and $\varphi \in C_b(\Pi)$,

$$S_n(\alpha\varphi + c) = \alpha S_n(\varphi) + c, \quad e^{t\mathcal{A}}(\alpha\varphi + c) = \alpha e^{t\mathcal{A}}\varphi + c.$$

Hence, by rescaling and shifting φ , it is sufficient to prove the claim for functions $\varphi \in C_b(\Pi)$ with $0 \leq \varphi \leq 1$.

Let $\{\chi_m\}_m \subset C_c(\Pi)$ be a sequence of compactly supported functions with $\chi_m \nearrow 1$ pointwise, and set $\chi'_m := 1 - \chi_m$. We decompose

$$e^{t\mathcal{A}}\varphi - S_n\varphi = (e^{t\mathcal{A}}[\varphi\chi_m] - S_n[\varphi\chi_m]) + (e^{t\mathcal{A}}[\varphi\chi'_m] - S_n[\varphi\chi'_m]).$$

Since $\varphi\chi_m \in C_0(\Pi)$, and $S_n \rightarrow e^{t\mathcal{A}}$ strongly on $C_0(\Pi)$ as mentioned before the proof, we have $S_n[\varphi\chi_m] \rightarrow e^{t\mathcal{A}}[\varphi\chi_m]$ uniformly. Passing to the limit $n \rightarrow \infty$ gives, for each $x \in \Pi$,

$$\begin{aligned} \limsup_{n \rightarrow \infty} |(e^{t\mathcal{A}}\varphi - S_n\varphi)(x)| &\leq |e^{t\mathcal{A}}[\varphi\chi'_m](x)| + \limsup_{n \rightarrow \infty} |S_n[\varphi\chi'_m](x)| \\ &\leq e^{t\mathcal{A}}[\chi'_m](x) + \limsup_{n \rightarrow \infty} S_n[\chi'_m](x). \end{aligned}$$

Here we used that both operators are order preserving and $0 \leq \varphi\chi'_m \leq \chi'_m$. Next, since $\chi_m \in C_0(\Pi)$, the strong convergence $S_n \rightarrow e^{t\mathcal{A}}$ implies

$$\lim_{n \rightarrow \infty} S_n[\chi'_m] = \lim_{n \rightarrow \infty} (S_n[1 - \chi_m]) = 1 - \lim_{n \rightarrow \infty} S_n\chi_m = 1 - e^{t\mathcal{A}}\chi_m = e^{t\mathcal{A}}\chi'_m.$$

Hence we obtain

$$\limsup_{n \rightarrow \infty} |(e^{t\mathcal{A}}\varphi - S_n\varphi)(x)| \leq 2(1 - e^{t\mathcal{A}}\chi_m(x)).$$

This estimate holds for all $m \geq 1$. Since $\chi_m \nearrow 1$, the monotone convergence property of $e^{t\mathcal{A}}$ yields $e^{t\mathcal{A}}\chi_m \nearrow 1$. Letting $m \rightarrow \infty$ therefore gives

$$\limsup_{n \rightarrow \infty} |(e^{t\mathcal{A}}\varphi - S_n\varphi)(x)| = 0,$$

which completes the proof. $\square$

The next proposition, which contains Proposition 2.12, collects a range of equivalent characterisations for coupling generators, formulated in terms of semigroup, generator, and resolvent conditions. For a probability semigroup $\mathcal{A} \in \mathcal{G}(\Pi)$ and $T > 0$, recall the continuity domain $C(\mathcal{A})$ of $\mathcal{A}$, as defined in Definition 2.1.

Proposition B.2. *Let $\mathcal{A}, \mathcal{B} \in \mathcal{G}(\Pi)$ and $\mathcal{J} \in \mathcal{G}(\Pi^2)$. The following are equivalent.*

(i) *For all $\varphi, \psi \in C_0(\Pi)$, it holds*

$$e^{t\mathcal{J}}[\varphi \otimes 1] = (e^{t\mathcal{A}}\varphi) \otimes 1, \quad e^{t\mathcal{J}}[1 \otimes \psi] = 1 \otimes (e^{t\mathcal{B}}\psi), \quad t \geq 0. \quad (\text{B.1})$$

(ii) *For all $\varphi, \psi \in C_b(\Pi)$, (B.1) holds.*

(iii) *For all $\varphi \in C(\mathcal{A}), \psi \in C(\mathcal{B})$, (B.1) holds.*

(iv) *For all $\varphi \in \bar{D}(\mathcal{A}), \psi \in \bar{D}(\mathcal{B})$, it holds*

$$\mathcal{J}[\varphi \otimes 1] = (\mathcal{A}\varphi) \otimes 1, \quad \mathcal{J}[1 \otimes \psi] = 1 \otimes (\mathcal{B}\psi). \quad (\text{B.2})$$

(v) *For all $\varphi \in D(\mathcal{A}), \psi \in D(\mathcal{B})$, (B.2) holds.*

(vi) *For all $\lambda > 0$, and $\varphi, \psi \in C_0(\Pi)$ that*

$$R_\lambda(\mathcal{J})[\varphi \otimes 1] = (R_\lambda(\mathcal{A})\varphi) \otimes 1, \quad R_\lambda(\mathcal{J})[1 \otimes \psi] = 1 \otimes (R_\lambda(\mathcal{B})\psi).$$

Proof. By symmetry, it suffices to prove the marginal condition involving $\mathcal{A}$ for each implication.

(i) $\rightarrow$ (ii). It is enough to treat the case $\varphi \geq 0$ with $\varphi \in C_b(\Pi)$. Let $\{\varphi_n\}_n \subset C_0(\Pi)$ be an increasing sequence with $\varphi_n \nearrow \varphi$. Then $\varphi_n \otimes 1 \nearrow \varphi \otimes 1$. Using (i) and the monotone convergence property of $e^{t\mathcal{J}}$ and $e^{t\mathcal{A}}$,

$$e^{t\mathcal{J}}[\varphi \otimes 1] = \lim_{n \rightarrow \infty} e^{t\mathcal{J}}[\varphi_n \otimes 1] = \lim_{n \rightarrow \infty} (e^{t\mathcal{A}}\varphi_n) \otimes 1 = (e^{t\mathcal{A}}\varphi) \otimes 1.$$

(ii) $\rightarrow$ (iii). Let $\varphi \in C(\mathcal{A})$ be nonnegative. Approximate φ monotonically from below by a sequence in $C_b(\Pi)$, and apply (ii) together with the convergence theorem. This yields (B.1) for $\varphi \in C(\mathcal{A})$.

(iii) $\rightarrow$ (iv). Suppose $\varphi \in \bar{D}(\mathcal{A})$, so that $\frac{1}{t}(e^{t\mathcal{A}} - 1)\varphi \rightarrow \mathcal{A}\varphi$ locally uniformly as $t \searrow 0$. By (iii),

$$\frac{e^{t\mathcal{J}} - I}{t}(\varphi \otimes 1) = \left(\frac{e^{t\mathcal{A}} - I}{t}\varphi \right) \otimes 1.$$

Passing to the limit $t \searrow 0$, the right-hand side converges locally uniformly to $(\mathcal{A}\varphi) \otimes 1$, and hence so does the left-hand side. This gives (B.2).

(iv) $\rightarrow$ (v). Immediate, since $D(\mathcal{A}) \subset \bar{D}(\mathcal{A})$.

(v) $\rightarrow$ (vi). Take $\varphi \in C_0(\Pi)$ and set $\varphi' := R_\lambda(\mathcal{A})\varphi \in D(\mathcal{A})$, so that $(\lambda I - \mathcal{A})\varphi' = \varphi$. Then

$$\begin{aligned} (\lambda I - \mathcal{J})[\varphi' \otimes 1] &= \lambda\varphi' \otimes 1 - \mathcal{J}[\varphi' \otimes 1] \\ &= \lambda\varphi' \otimes 1 - (\mathcal{A}\varphi') \otimes 1 = (\lambda I - \mathcal{A})\varphi' \otimes 1 = \varphi \otimes 1. \end{aligned}$$

Applying $R_\lambda(\mathcal{J})$ gives

$$R_\lambda(\mathcal{J})[\varphi \otimes 1] = \varphi' \otimes 1 = (R_\lambda(\mathcal{A})\varphi) \otimes 1.$$

(vi) $\rightarrow$ (i). For $\varphi \in C_0(\Pi)$ we have $\varphi \otimes 1 \in C_b(\Pi)$. By (vi),

$$I_\lambda(\mathcal{J})[\varphi \otimes 1] = (I_\lambda(\mathcal{A})\varphi) \otimes 1.$$

Iterating, the same holds for all powers I_λ^n . Applying Lemma B.1,

$$e^{t\mathcal{J}}[\varphi \otimes 1] = \lim_{n \rightarrow \infty} I_{n/t}(\mathcal{J})^n[\varphi \otimes 1] = \lim_{n \rightarrow \infty} (I_{n/t}(\mathcal{A})^n \varphi) \otimes 1 = (e^{t\mathcal{A}}\varphi) \otimes 1.$$

This proves (i). $\square$

B.2. Proof of Proposition 2.21. We are to prove the inequality:

$$\omega_c^*(x, y; \mathcal{A}, \mathcal{B}) \leq \omega_c^-(x, y; \mathcal{A}, \mathcal{B}) \leq \omega^+(x, y; \mathcal{A}, \mathcal{B}) \leq \theta_c(x, y; \mathcal{A}, \mathcal{B}).$$

As pointed out in Remark 2.5, the second inequality is straightforward from the definition of $\omega_c^\pm$.

Let us proceed with the last inequality. Choose a Feller coupling $\mathcal{J} \in \Gamma(\mathcal{A}, \mathcal{B})$. For any $(x, y) \in \Pi^2$, $\gamma_t = \delta_{(x, y)} e^{t\mathcal{J}} \in \mathcal{P}(\Pi^2)$ is a coupling of the probability measures $\delta_x e^{t\mathcal{A}}, \delta_y e^{t\mathcal{B}}$. This follows

$$\mathcal{C}_c(\delta_x e^{t\mathcal{A}}, \delta_y e^{t\mathcal{B}}) \leq \int_{\Pi^2} c(x, y) d\gamma_t(x, y) = e^{t\mathcal{J}} c(x, y).$$

Subtracting both sides with $c(x, y)$, dividing t , and passing $t \searrow 0$ we find

$$\omega_c^+(x, y; \mathcal{A}, \mathcal{B}) = \limsup_{t \rightarrow 0} \frac{\mathcal{C}_c(\delta_x e^{t\mathcal{A}}, \delta_y e^{t\mathcal{B}}) - c(x, y)}{t} \leq \liminf_{t \searrow 0} \frac{e^{t\mathcal{J}} c(x, y) - c(x, y)}{t} = \mathcal{J}^- c(x, y).$$

This holds for any $\mathcal{J} \in \Gamma(\mathcal{A}, \mathcal{B})$. Taking the infimum over $\mathcal{J} \in \Gamma(\mathcal{A}, \mathcal{B})$ yields the second inequality.

Finally consider the first inequality. Fix $(x_0, y_0) \in \Pi^2$ and any pair $\varphi \in \bar{D}(\mathcal{A}), \psi \in \bar{D}(\mathcal{B})$ so that $\varphi \oplus \psi \leq c$, and $\varphi(x_0) + \psi(y_0) = c(x_0, y_0)$. Fix $t \geq 0$ and denote $\mu_t = \delta_{x_0} e^{t\mathcal{A}}, \nu_t = \delta_{y_0} e^{t\mathcal{B}}$. Let γ_t be a c -optimal coupling of the measures μ_t, ν_t . This follows

$$\begin{aligned} (e^{t\mathcal{A}}\varphi)(x_0) + (e^{t\mathcal{B}}\psi)(y_0) &= \int_{\Pi} \varphi(x) d\mu_t + \int_{\Pi} \psi(x) d\nu_t \\ &= \int_{\Pi^2} [\varphi(x) + \psi(y)] d\gamma_t(x, y) \leq \int_{\Pi^2} c(x, y) d\gamma_t(x, y) \\ &= \mathcal{C}_c(\delta_{x_0} e^{t\mathcal{A}}, \delta_{y_0} e^{t\mathcal{B}}). \end{aligned}$$

Subtracting both side with $\varphi(x_0) + \psi(y_0) = c_2(x_0, y_0)$ yields::

$$(e^{t\mathcal{A}}\varphi)(x_0) - \varphi(x_0) + (e^{t\mathcal{B}}\psi)(y_0) - \psi(y_0) \leq \mathcal{C}_c(\delta_{x_0} e^{t\mathcal{A}}, \delta_{y_0} e^{t\mathcal{B}}) - c(x_0, y_0).$$

Dividing t and taking limit inferior as $t \searrow 0$, it leads us to

$$(\mathcal{A}\varphi)(x_0) + (\mathcal{B}\psi)(y_0) \leq \omega_c^-(x_0, y_0; \mathcal{A}, \mathcal{B}).$$

We note the inequality above holds for any $\varphi \in \bar{D}(\mathcal{A}), \psi \in \bar{D}(\mathcal{B})$ satisfying $\varphi \oplus \psi \leq c$, $\varphi(x_0) + \psi(y_0) = c(x_0, y_0)$. Taking the supremum over all such pairs (φ, ψ) , this leads to the first inequality.

B.3. Proof of Proposition 2.27. (i) Fix $(x_0, y_0) \in \mathbb{R}^{2d}$. We construct two barrier functions $a \leq c_2 \leq b$ that touch c_2 at (x_0, y_0) . The lower barrier a is the supporting plane of c_2 at (x_0, y_0) , while b is a quadratic function touching c_2 from above:

$$\begin{aligned} a(x, y) &= \frac{1}{2}(x_0 - y_0)^\top (x - y) = \frac{1}{2}(x_0 - y_0)^\top x - \frac{1}{2}(x_0 - y_0)^\top y =: \varphi(x) + \psi(y), \\ b(x, y) &= |x - x_0|^2 + |y - y_0|^2 + c_2(x_0, y_0) + (x - x_0)^\top (x_0 - y_0) - (y - y_0)^\top (x_0 - y_0) \\ &=: \Phi(x) + \Psi(y) + c_2(x_0, y_0). \end{aligned}$$

Clearly $a \leq c_2 \leq b$, with equality at (x_0, y_0) .

Let $\{T_t\}_{t \geq 0}$ be a Markovian coupling semigroup of $\mathcal{A}, \mathcal{B}$. By order preservation,

$$T_h a \leq T_h c_2 \leq T_h b, \quad h \geq 0. \quad (\text{B.3})$$

Since both a and b are of direct sum form, the marginality property yields

$$\begin{aligned} (T_h a)(x_0, y_0) &= (e^{h\mathcal{A}}\varphi)(x_0) + (e^{h\mathcal{B}}\psi)(y_0) \\ &= c_2(x_0, y_0) + \frac{h}{2}(m_{\mathcal{A}} - m_{\mathcal{B}})^\top (x_0 - y_0), \\ (T_h b)(x_0, y_0) &= (e^{h\mathcal{A}}\Phi)(x_0) + (e^{h\mathcal{B}}\Psi)(y_0) + c_2(x_0, y_0) \\ &= c_2(x_0, y_0) + h^2(|m_{\mathcal{A}}|^2 + |m_{\mathcal{B}}|^2) + h \left[\text{tr}(Q_{\mathcal{A}} + Q_{\mathcal{B}}) + (m_{\mathcal{A}} - m_{\mathcal{B}})^\top (x_0 - y_0) \right]. \end{aligned}$$

Here we used that for quadratic-linear test functions, the semigroup action can be computed explicitly from the mean vector and covariance matrix of the underlying Lévy generator:

$$\begin{aligned} (e^{t\mathcal{A}}\varphi)(x_0) &= \frac{1}{2}(x_0 + tm_{\mathcal{A}})^\top (x_0 - y_0), \\ (e^{t\mathcal{B}}\psi)(y_0) &= -\frac{1}{2}(y_0 + tm_{\mathcal{B}})^\top (x_0 - y_0), \\ (e^{t\mathcal{A}}\Phi)(x_0) &= t^2|m_{\mathcal{A}}|^2 + t\text{tr}(Q_{\mathcal{A}}) + tm_{\mathcal{A}}^\top (x_0 - y_0), \\ (e^{t\mathcal{B}}\Psi)(y_0) &= t^2|m_{\mathcal{B}}|^2 + t\text{tr}(Q_{\mathcal{B}}) - tm_{\mathcal{B}}^\top (x_0 - y_0). \end{aligned}$$

Recalling $U(t, x, y) = (x - y + t(m_{\mathcal{A}} - m_{\mathcal{B}}))^\top (m_{\mathcal{A}} - m_{\mathcal{B}})$, the identities above for $T_h a, T_h b$, combined with (B.3), give

$$\frac{h}{2} U(0, x_0, y_0) \leq (T_h c_2 - c_2)(x_0, y_0) \leq h^2(|m_{\mathcal{A}}|^2 + |m_{\mathcal{B}}|^2) + h[\text{tr}(Q_{\mathcal{A}} + Q_{\mathcal{B}}) + U(0, x_0, y_0)].$$

Finally, by order preservation again, applying T_t to this inequality yields for all $t, h \geq 0$:

$$\frac{h}{2} U(t, x_0, y_0) \leq (T_{t+h} c_2 - T_t c_2)(x_0, y_0) \leq h^2(|m_{\mathcal{A}}|^2 + |m_{\mathcal{B}}|^2) + h[\text{tr}(Q_{\mathcal{A}} + Q_{\mathcal{B}}) + U(t, x_0, y_0)],$$

which is the desired bound. We note here the identity $T_t U(0, \cdot, \cdot) = U(t, \cdot, \cdot)$ is used.

(ii) Let $\mathcal{J} \in \Gamma(\mathcal{A}, \mathcal{B})$. Since $\{e^{t\mathcal{J}}\}_{t \geq 0}$ is itself a Markovian coupling semigroup, the estimate from (a) applies to $e^{t\mathcal{J}} c_2$, which shows that $t \mapsto e^{t\mathcal{J}} c_2$ is locally uniformly continuous. To conclude $c_2 \in C(\mathcal{J})$, it remains to check that for each $t \geq 0$, one has $e^{t\mathcal{J}} c_2 \in C(\mathbb{R}^d)$. Let $\{\kappa_t(x, y)\}$ denote the transition kernel of $\mathcal{J}$. By the Feller property, the map $(x, y) \mapsto \kappa_t(x, y)$ is weakly continuous. Consider $Q(x, y) = q(x) + q(y)$ with $q(z) = |z|^2$. By the marginal property of the coupling semigroup,

$$\langle \kappa_t(x, y), Q \rangle = e^{t\mathcal{J}} Q(x, y) = (e^{t\mathcal{A}} q)(x) + (e^{t\mathcal{B}} q)(y).$$

Since $q \in \bar{D}(\mathcal{A}) \cap \bar{D}(\mathcal{B})$ (as $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$), both terms on the right-hand side are continuous, and hence $(x, y) \mapsto \kappa_t(x, y)$ is continuous in the 2-Wasserstein topology. Consequently, for any continuous function with quadratic growth, such as c_2 , we obtain

$$(e^{t\mathcal{J}}c_2)(x, y) = \langle \kappa_t(x, y), c_2 \rangle,$$

which is continuous in (x, y) . Thus $e^{t\mathcal{J}}c_2 \in C(\mathbb{R}^d)$, and so $c_2 \in C(\mathcal{J})$.

(iii) If $\mathcal{A}, \mathcal{B} \in \mathcal{G}_2^\Lambda(\mathbb{R}^d)$, and $\mathcal{J} \in \Gamma(\mathcal{A}, \mathcal{B})$ is itself a Lévy generator, then clearly $\mathcal{J} \in \mathcal{G}_2^\Lambda(\mathbb{R}^{2d})$, i.e., it admits finite second moments. This is because $e^{t\mathcal{J}}[|x|^2 + |y|^2] = e^{t\mathcal{A}}[|x|^2] + e^{t\mathcal{B}}[|y|^2] < \infty$. In this case $\bar{D}(\mathcal{J})$ contains all functions in $C_2^2(\mathbb{R}^{2d})$, which includes c_2 . Hence $c_2 \in \bar{D}(\mathcal{J})$.

APPENDIX C. Λ -WEAK TOPOLOGY IN THE SPACE $\Lambda(\mathbb{R}^d)$ OF LÉVY MEASURES

This appendix reviews basic results on the weak topology for Lévy measures. While related notions appear in standard references such as Sato's [42] or Kallenberg's [26] monograph, they are not typically stated in a form tailored to our convergence analysis. The results presented here parallel the classical weak theory for probability measures—such as the Portmanteau theorem and Prokhorov's compactness criterion—but adapted to the Lévy class $\Lambda(\mathbb{R}^d)$, where singularity at the origin.

Let us briefly recall the notions of *weak convergence*, *vague convergence* and *tightness* for bounded measures in the classical setting. Throughout the discussion, let Π be a locally compact, Hausdorff, second countable space, and denote by $\mathcal{M}(\Pi)$ the space of bounded Radon (Borel) measures on Π . A sequence $\{\mu_n\}_n \subset \mathcal{M}(\Pi)$ is said to *converge weakly* to $\mu \in \mathcal{M}(\Pi)$ if

$$\lim_{n \rightarrow \infty} \int_{\Pi} \varphi(x) d\mu_n(x) = \int_{\Pi} \varphi(x) d\mu(x)$$

for all bounded continuous functions $\varphi \in C_b(\Pi)$. The sequence *converges vaguely* to μ if the same holds for all compactly supported test functions $\varphi \in C_c(\Pi)$. A family $\mathcal{N} \subset \mathcal{M}(\Pi)$ is said to be *tight* if:

- (1) $\sup_{\mu \in \mathcal{N}} \mu(\Pi) < \infty$;
- (2) for every $\varepsilon > 0$, there exists a compact set $K \subset \Pi$ such that $\sup_{\mu \in \mathcal{N}} \mu(K^c) < \varepsilon$.

It is well known that:

- If $\mu_n \rightarrow \mu$ weakly, then $\mu_n \rightarrow \mu$ vaguely.
- Conversely, if $\mu_n \rightarrow \mu$ vaguely, the family $\{\mu_n\}$ is tight, then $\mu_n \rightarrow \mu$ weakly.
- In particular, for probability measures (where $\mu_n(\Pi) = 1$), vague convergence together with tightness is equivalent to weak convergence.

In our work, we utilize the following notions of convergence and tightness of Lévy measures in the sense of Lévy.

Definition C.1 (Λ -weak convergence and Λ -tightness). (i) For $n \geq 1$, let $\mu, \mu_n \in \Lambda(\mathbb{R}^d)$. We say that μ_n converges Lévy weakly, or Λ -weakly, to μ , written $\mu_n \xrightarrow{\Lambda} \mu$, if

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} \varphi(x) d\mu_n(x) = \int_{\mathbb{R}^d} \varphi(x) d\mu(x) \quad (\text{C.1})$$

for all bounded continuous test functions $\varphi \in C_b(\mathbb{R}^d)$ satisfying $|\varphi(x)| \leq C \min\{1, |x|^2\}$ for some constant $C \geq 0$.

(ii) A family $\mathcal{M} \subset \Lambda(\mathbb{R}^d)$ is said to be Λ -tight if the following conditions holds: (1) it holds

$$\sup_{\mu \in \mathcal{M}} \int_{\mathbb{R}^d} \min\{1, |x|^2\} d\mu(x) < \infty;$$

(2) for every $\varepsilon > 0$, there exist a compact set $K \subset \mathbb{R}^d \setminus \{0\}$ such that

$$\sup_{\mu \in \mathcal{M}} \int_{K^c} \min\{1, |x|^2\} d\mu(x) < \varepsilon.$$

We now relate these classical notions to their counterparts in the context of Lévy measures. Given a Lévy measure $\mu \in \Lambda(\mathbb{R}^d)$, define its associated re-scaled (bounded) measure $\tilde{\mu} \in \mathcal{M}(\mathbb{R}^d \setminus \{0\})$ by

$$d\tilde{\mu}(x) := \min\{1, |x|^2\} d\mu(x),$$

where $\mathbb{R}^d \setminus \{0\}$ is a Polish space. Likewise, given a measurable function $\tilde{\varphi} : \mathbb{R}^d \setminus \{0\} \rightarrow \mathbb{R}$, define its re-scaled version by

$$\varphi(x) := \min\{1, |x|^2\} \tilde{\varphi}(x), \quad \varphi(0) := 0. \quad (\text{C.2})$$

Then the identity holds:

$$\int_{\mathbb{R}^d} \varphi d\mu = \int_{\mathbb{R}^d \setminus \{0\}} \tilde{\varphi} d\tilde{\mu} \quad (\text{C.3})$$

Moreover, $\tilde{\varphi}$ is continuous (resp. lower semicontinuous, upper semicontinuous) on $\mathbb{R}^d \setminus \{0\}$ iff φ is continuous (resp. lower semicontinuous, upper semicontinuous) on $\mathbb{R}^d$. Similarly, $\tilde{\varphi}$ is bounded above (resp. below) iff φ satisfies $\varphi(x) \leq C \min\{1, |x|^2\}$ (resp. $\varphi(x) \geq -C \min\{1, |x|^2\}$) for some $C \geq 0$.

This rescaling relation allows us to transfer classical weak convergence results for $\tilde{\mu}$ to the setting of Lévy measures μ , as will be seen in the following characterization.

Proposition C.2.

(i) Let $\mu, \mu_n \in \Lambda(\mathbb{R}^d)$ for $n \geq 1$. The following are equivalent:

- $\mu_n \xrightarrow{\Lambda} \mu$ (Λ -weak convergence);
- $\tilde{\mu}_n \rightarrow \tilde{\mu}$ weakly on $\mathbb{R}^d \setminus \{0\}$.

(ii) Let $\mathcal{N} \subset \Lambda(\mathbb{R}^d)$ be a family of Lévy measures. The following are equivalent:

- $\mathcal{N}$ is Λ -tight;
- $\tilde{\mathcal{N}} := \{\tilde{\mu} : \mu \in \mathcal{N}\}$ is tight on $\mathbb{R}^d \setminus \{0\}$.

Proof. (i) By (C.2) and (C.3), testing against φ in the definition of Λ -weak convergence is equivalent to testing against $\tilde{\varphi} \in C_b(\mathbb{R}^d \setminus \{0\})$. Hence $\mu_n \xrightarrow{\Lambda} \mu$ iff $\tilde{\mu}_n \rightarrow \tilde{\mu}$ weakly.

(ii) Likewise, by (C.3), Λ -tightness of $\{\mu_n\}$ is equivalent to tightness of $\{\tilde{\mu}_n\}$ on $\mathbb{R}^d \setminus \{0\}$. $\square$

C.1. Portmanteau Lemma for Λ -weak convergence. Our first result is a Portmanteau-type theorem, giving an equivalent characterization of Λ -weak convergence in terms of lower and upper semicontinuous test functions.

Lemma C.3 (Portmanteau-type lemma for Λ -weak convergence). *Let $\{\mu_n\}_n \subset \Lambda(\mathbb{R}^d)$ be a sequence of Lévy measures, and let $\mu \in \Lambda(\mathbb{R}^d)$. The following are equivalent:*

- (i) $\mu_n \xrightarrow{\Lambda} \mu$.
- (ii) For every lower semicontinuous function $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ satisfying $\varphi(x) \geq -C \min\{1, |x|^2\}$ for some $C \geq 0$, it holds that

$$\int_{\mathbb{R}^d} \varphi(x) d\mu(x) \leq \liminf_{n \rightarrow \infty} \int_{\mathbb{R}^d} \varphi(x) d\mu_n(x).$$

- (iii) For every upper semicontinuous function $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ satisfying $\varphi(x) \leq C \min\{1, |x|^2\}$ for some $C \geq 0$, it holds that

$$\int_{\mathbb{R}^d} \varphi(x) d\mu(x) \geq \limsup_{n \rightarrow \infty} \int_{\mathbb{R}^d} \varphi(x) d\mu_n(x).$$

Proof. By Lemma C.2, $\mu_n \xrightarrow{\Lambda} \mu$ iff the rescaled measures $\tilde{\mu}_n \rightarrow \tilde{\mu}$ weakly on $\mathbb{R}^d \setminus \{0\}$, with the correspondence (C.3), where $\tilde{\varphi}$ is the rescaling of φ defined in (C.2). The conclusion then follows directly from the classical Portmanteau lemma applied to $\tilde{\mu}_n \rightarrow \tilde{\mu}$. $\square$

Proposition C.4. *We have $\mu_n \xrightarrow{\Lambda} \mu$ if and only if (C.1) holds for all bounded Lipschitz test functions $\varphi \in C_b(\mathbb{R}^d)$ satisfying*

$$|\varphi(x)| \leq C \min\{1, |x|^2\}, \quad |\nabla \varphi(x)| \leq C|x| \quad (\text{Lebesgue a.e.})$$

for some constant $C \geq 0$.

Proof. The implication “ $\rightarrow$ ” is immediate from the definition of Λ -weak convergence. Let us prove the converse.

By assumption, (C.1) holds for every bounded Lipschitz function φ satisfying the above growth bounds. In particular, it applies to functions φ that vanish inside some ball $B_r(0)$, since in that case we can find constants $C, C' \geq 0$ with

$$|\nabla \varphi(x)| \leq C' \chi_{B_r(0)^c}(x) \leq C|x|.$$

For each $r > 0$, the family $\{\chi_{B_r(0)^c} \mu_n\}_n$ is tight (as the sequence of truncated measures converges weakly). By density of Lipschitz functions in $C_b(\mathbb{R}^d)$, it follows that (C.1) holds for every $\varphi \in C_b(\mathbb{R}^d)$ that vanishes on some $B_r(0)$.

Next, let us extend (C.1) to all $\varphi \in C_b(\mathbb{R}^d)$ such that $\varphi(x) \leq C \min\{1, |x|^2\}$ for some $C \geq 0$. Fix a cutoff $\chi \in C_b(\mathbb{R}^d)$ that is radially nondecreasing, with $\chi(x) = 0$ for $|x| \leq 1$ and $\chi(x) = 1$ for $|x| \geq 2$. For $\varepsilon > 0$, set $\chi_\varepsilon(x) := \chi(\frac{x}{\varepsilon})$, $\varphi_\varepsilon := \varphi \chi_\varepsilon$. Then

$$|\varphi_\varepsilon(x)| \leq C \min\{1, |x|^2\}, \quad |\varphi(x) - \varphi_\varepsilon(x)| \leq C \chi_{B_{2\varepsilon}(0)}(x) \min\{1, |x|^2\} \leq C \min\{4\varepsilon^2, |x|^2\}.$$

Hence we have

$$\left| \int_{\mathbb{R}^d} \varphi d(\mu_n - \mu) \right| \leq \left| \int_{\mathbb{R}^d} \varphi_\varepsilon d(\mu_n - \mu) \right| + \int_{\mathbb{R}^d} |\varphi - \varphi_\varepsilon| d(\mu_n + \mu).$$

As $n \rightarrow \infty$, the first term vanishes by the previous paragraph (since φ_ε vanishes on $B_\varepsilon(0)$). Therefore

$$\begin{aligned} \limsup_{n \rightarrow \infty} \left| \int_{\mathbb{R}^d} \varphi d(\mu_n - \mu) \right| &\leq \limsup_{n \rightarrow \infty} \int_{\mathbb{R}^d} |\varphi - \varphi_\varepsilon| d(\mu_n + \mu) \\ &\leq \limsup_{n \rightarrow \infty} C \int_{\mathbb{R}^d} \min\{|x|^2, 4\varepsilon^2\} d(\mu_n + \mu) \\ &= 2C \int_{\mathbb{R}^d} \min\{|x|^2, 4\varepsilon^2\} d\mu. \end{aligned}$$

Here we used that $\psi(x) := C \min\{4\varepsilon^2, |x|^2\}$ satisfies $|\nabla \psi(x)| \leq C|x|$, so (C.1) also applies to ψ . Finally, since $\varepsilon > 0$ is arbitrary, sending $\varepsilon \searrow 0$ yields that (C.1) holds for all admissible φ . Thus $\mu_n \xrightarrow{\Lambda} \mu$. $\square$

Proposition C.5. *If $\mu_n \xrightarrow{\Lambda} \mu$ for some $\mu \in \Lambda(\mathbb{R}^d)$, then the sequence $\{\mu_n\}_n$ is Λ -tight.*

Proof. By the characterization given by Lemma C.2, $\mu_n \xrightarrow{\Lambda} \mu$ if and only if $\tilde{\mu}_n \rightarrow \tilde{\mu}$ weakly on $\mathbb{R}^d \setminus \{0\}$. From the classical result, $\{\tilde{\mu}_n\}$ is tight on $\mathbb{R}^d \setminus \{0\}$ in the usual sense. Applying Lemma C.2 again $\{\mu_n\}_n$ is Λ -tight. $\square$

Next, we characterize Λ -convergence for Lévy measures with bounded second moments. This parallels the case of probability measures with finite second moments (cf. [46, Definition 6.8]).

Lemma C.6 (Characterization of convergence in $\Lambda_2(\mathbb{R}^d)$). *Let $\mu_n, \mu \in \Lambda_2(\mathbb{R}^d)$. Suppose $\mu_n \xrightarrow{\Lambda} \mu$ and*

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} |x|^2 d\mu_n(x) = \int_{\mathbb{R}^d} |x|^2 d\mu(x). \quad (\text{C.4})$$

Then for every $\varepsilon > 0$, there exists $R \geq 1$ such that

$$\sup_n \int_{|x| \geq R} |x|^2 d\mu_n(x) \leq \varepsilon.$$

Moreover, for any continuous φ with quadratic growth, $|\varphi(x)| \leq C|x|^2$ for some $C \geq 0$, one has

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} \varphi(x) d\mu_n(x) = \int_{\mathbb{R}^d} \varphi(x) d\mu(x).$$

Proof. Fix $\varepsilon > 0$. Since $|x|^2 \in L^1(\mu)$, there exists $R > 0$ such that

$$\int_{|x| \geq R} |x|^2 d\mu(x) < \frac{\varepsilon}{2}.$$

Consider the truncated function $|x|^2 \chi_{\{|x| < R\}}$, which is lower semicontinuous and bounded below. By the Portmanteau theorem (Lemma C.3),

$$\liminf_{n \rightarrow \infty} \int_{|x| < R} |x|^2 d\mu_n(x) \geq \int_{|x| < R} |x|^2 d\mu(x).$$

Using the convergence of the second moment (C.4), we obtain

$$\limsup_{n \rightarrow \infty} \int_{|x| \geq R} |x|^2 d\mu_n(x) \leq \int_{|x| \geq R} |x|^2 d\mu(x) < \frac{\varepsilon}{2}.$$

Hence, there exists N such that for all $n \geq N$,

$$\int_{|x| \geq R} |x|^2 d\mu_n(x) < \varepsilon.$$

For the finitely many indices $n < N$, finiteness of the second moment allows enlarging R so that the same bound holds. Thus the first claim follows.

For the second claim, let $\varphi \in C(\mathbb{R}^d)$ with $|\varphi(x)| \leq C|x|^2$. For $R > 0$, define the cutoff

$$\varphi_R(x) = \max\{-R, \min\{\varphi(x), R\}\}, \quad \delta_R = \varphi - \varphi_R.$$

Then $\varphi_R \in C_b(\mathbb{R}^d)$ and $|\delta_R(x)| \leq C|x|^2 \chi_{\{|x| \geq R\}}$. Hence,

$$\left| \int \varphi d\mu_n - \int \varphi d\mu \right| \leq \left| \int \varphi_R d\mu_n - \int \varphi_R d\mu \right| + \int |\delta_R(x)| d(\mu_n + \mu).$$

As $n \rightarrow \infty$, the first term vanishes since $\mu_n \xrightarrow{\Lambda} \mu$. For the second term,

$$\limsup_{n \rightarrow \infty} \left| \int \varphi d\mu_n - \int \varphi d\mu \right| \leq 2C \sup_n \int_{|x| \geq R} |x|^2 d\mu_n(x).$$

By the first part, the right-hand side can be made arbitrarily small by choosing R large. This proves the result. $\square$

C.2. Prokhorov-type Compactness Theorem. We conclude with a compactness criterion for the space $\Lambda(\mathbb{R}^d)$ of Lévy measures, in analogy with Prokhorov's theorem.

Theorem C.7 (Prokhorov-type compactness). *Let $\{\mu_n\}_n \subset \Lambda(\mathbb{R}^d)$ be a Λ -tight sequence of Lévy measures. Then there exist a subsequence (still denoted μ_n) and a Lévy measure $\mu \in \Lambda(\mathbb{R}^d)$ such that $\mu_n \xrightarrow{\Lambda} \mu$.*

Proof. Let $\{\tilde{\mu}_n\}_n \subset \mathcal{M}(\mathbb{R}^d \setminus \{0\})$ be the rescaled measures as in (C.2). By Proposition C.2, the Λ -tightness of $\{\mu_n\}_n$ implies that $\{\tilde{\mu}_n\}_n$ is tight in $\mathcal{M}(\mathbb{R}^d \setminus \{0\})$. By Prokhorov's theorem, there exists a subsequence (still denoted $\tilde{\mu}_n$) converging weakly to some $\tilde{\mu} \in \mathcal{M}(\mathbb{R}^d \setminus \{0\})$. Define $\mu \in \Lambda(\mathbb{R}^d)$ by $d\mu(x) = (\min\{1, |x|^2\})^{-1} d\tilde{\mu}(x)$. Then Proposition C.2 yields $\mu_n \xrightarrow{\Lambda} \mu$ along the subsequence. $\square$

REFERENCES

- [1] A. Alfonsi, J. Corbetta, and B. Jourdain. Evolution of the Wasserstein distance between the marginals of two Markov processes. *Bernoulli*, 24(4A):2461–2498, 2018.
- [2] L. Ambrosio and N. Gigli. A user’s guide to optimal transport. In B. Piccoli and M. Rascle, editors, *Modelling and Optimisation of Flows on Networks*, volume 2062 of *Lecture Notes in Mathematics*, pages 1–155. Springer, Berlin, Heidelberg, 2013.
- [3] J. Backhoff-Veraguas and G. Pammer. Stability of martingale optimal transport and weak optimal transport. *Annals of Applied Probability*, 32(1):721–752, 2022.
- [4] A. D. Barbour, L. Holst, and S. Janson. *Poisson Approximation*, volume 2 of *Oxford Studies in Probability*. Oxford University Press, Oxford, 1992.
- [5] M. Beiglböck, P. Henry-Labordère, and F. Penkner. Model-independent bounds for option prices: A mass transport approach. *Finance and Stochastics*, 17(3):477–501, 2013.
- [6] B. Böttcher, R. Schilling, and J. Wang. *Lévy Matters III: Lévy-Type Processes: Construction, Approximation and Sample Path Properties*, volume 2099 of *Lecture Notes in Mathematics*. Springer, Cham, Switzerland, 2013.
- [7] B. Böttcher, R. L. Schilling, and J. Wang. Constructions of coupling processes for lévy processes. *Stochastic Process. Appl.*, 121(6):1201–1216, 2011.
- [8] K. Burdzy and W. S. Kendall. Efficient markovian couplings: Examples and counterexamples. *Annals of Applied Probability*, 10(2):362–409, 2000.
- [9] B. Böttcher. Markovian maximal coupling of Markov processes. *Journal of Theoretical Probability*, 33(3):1220–1260, 2020.
- [10] L. A. Caffarelli and R. J. McCann. Free boundaries in optimal transport and monge–ampère obstacle problems. *Annals of Mathematics*, 171(2):673–730, 2010.
- [11] L.-P. Chaintron and A. Diez. Propagation of chaos: A review of models, methods and applications. I. Models and methods. *Kinetic and Related Models*, 15(6):895–1015, 2022.
- [12] L.-P. Chaintron and A. Diez. Propagation of chaos: A review of models, methods and applications. II. Applications. *Kinetic and Related Models*, 15(6):1017–1173, 2022.
- [13] L. H. Y. Chen. Poisson approximation for dependent trials. *The Annals of Probability*, 3(3):534–545, 1975.
- [14] M.-F. Chen. Optimal markovian couplings and applications. *Acta Mathematica Sinica (English Series)*, 10(3):260–275, 1994.
- [15] M.-F. Chen. Optimal couplings and application to riemannian geometry. In edited by contributors, editor, *Probability Theory and Mathematical Statistics*, pages 121–142. Walter de Gruyter GmbH, 2020.
- [16] M.-F. Chen and S.-F. Li. Coupling methods for multidimensional diffusion processes. *Annals of Probability*, 17(1):151–177, 1989.
- [17] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. Unbalanced optimal transport: Dynamic and Kantorovich formulations. *Journal of Functional Analysis*, 274(11):3090–3123, 2018.
- [18] S. B. Connor. Optimal co-adapted coupling for a random walk on the hyper-complete graph. *arXiv preprint*, 2009.
- [19] W. Doeblin. Exposé de la théorie des chaînes simples constantes de markov à un nombre fini d’états. *Rev. Math. Union Interbalcanique*, 1:77–105, 1937.
- [20] G. Dávila and Y.-H. Kim. Dynamics of optimal partial transport. *Calculus of Variations and Partial Differential Equations*, 55(3):116, 2016.
- [21] K.-J. Engel and R. Nagel. *One-Parameter Semigroups for Linear Evolution Equations*, volume 194 of *Graduate Texts in Mathematics*. Springer New York, 2000.
- [22] S. N. Ethier and T. G. Kurtz. *Markov processes: characterization and convergence*. John Wiley & Sons, 2009.
- [23] C. R. Givens and R. M. Shortt. A class of Wasserstein metrics for probability distributions. *Michigan Mathematical Journal*, 31(2):231–240, 1984.

- [24] S. Goldstein. Maximal coupling. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 45(3):269–274, 1978.
- [25] D. Griffeath. A maximal coupling for markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 31(2):95–106, 1975.
- [26] O. Kallenberg. *Foundations of Modern Probability*. Probability and Its Applications. Springer, New York, 1st edition, 1997.
- [27] W. S. Kendall. Nonnegative Ricci curvature and the Brownian coupling property. *Stochastic Processes and their Applications*, 20(2):245–258, 1986.
- [28] W. S. Kendall. Coupling, local times, immersions. *Bernoulli*, 21(2):1014–1046, 2015.
- [29] W. S. Kendall, M. B. Majka, and A. Mijatović. Optimal markovian coupling for finite-activity lévy processes. *Bernoulli*, 30(4):2821–2845, 2024.
- [30] V. N. Kolokoltsov. The lévy–khintchine type operators with variable lipschitz continuous coefficients generate linear or nonlinear markov processes and semigroups. *Probability Theory and Related Fields*, 151(3):667–696, 2011. Originally submitted as arXiv:0911.5688.
- [31] F. Kühn. *Lévy Matters VI: Lévy-Type Processes: Moments, Construction and Heat Kernel Estimates*, volume 2187 of *Lecture Notes in Mathematics*. Springer, Cham, Switzerland, 2017.
- [32] K. Kuwada. Duality on gradient estimates and wasserstein controls. *Journal of Functional Analysis*, 258(11):3758–3774, 2010.
- [33] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2nd edition, 2017.
- [34] M. Liang, R. L. Schilling, and J. Wang. A unified approach to coupling sdes driven by lévy noise and some applications. *Bernoulli*, 26(1):664–693, 2020.
- [35] T. M. Liggett. *Continuous Time Markov Processes*, volume 113 of *Graduate Studies in Mathematics*. American Mathematical Society, 2010.
- [36] T. S. Lim and C. D. Teoh. Abstract formulation of mean-field models and propagation of chaos. *arXiv preprint*, arXiv:2508.02224, 2025.
- [37] T. Lindvall and L. C. G. Rogers. Coupling of multidimensional diffusions by reflection. *The Annals of Probability*, 14(3):860–872, 1986.
- [38] H. P. McKean. Propagation of chaos for a class of non-linear parabolic equations. In A. K. Aziz, editor, *Lecture Series in Differential Equations, Volume 2*, number 19 in Van Nostrand Mathematical Studies, pages 177–194. Van Nostrand Reinhold Company, 1969.
- [39] N. Nüsken and G. A. Pavliotis. Constructing sampling schemes via coupling: Markov semigroups and optimal transport. *SIAM/ASA Journal on Uncertainty Quantification*, 7(1):324–382, 2019.
- [40] J. G. Propp and D. B. Wilson. Exact sampling with coupled markov chains and applications to statistical mechanics. In *Random Structures & Algorithms*, volume 9, pages 223–252, 1996.
- [41] J. S. Rosenthal. Faithful couplings of markov chains: Now equals forever. *Advances in Applied Mathematics*, 18(3):372–381, 1997. Cited by over 200 works; foundational in Markov chain convergence theory.
- [42] K.-i. Sato. *Lévy Processes and Infinitely Divisible Distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1999.
- [43] R. L. Schilling and J. Wang. On the coupling property of lévy processes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47(4):1147–1159, 2011.
- [44] D. W. Stroock and S. R. S. Varadhan. *Multidimensional Diffusion Processes*. Classics in Mathematics. Springer, Berlin, 2006.
- [45] A.-S. Sznitman. Topics in propagation of chaos. In P.-L. Hennequin, editor, *École d’Eté de Probabilités de Saint-Flour XIX - 1989*, volume 1464 of *Lecture Notes in Mathematics*, pages 165–251. Springer Berlin, Heidelberg, 1991.
- [46] C. Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, Heidelberg, 2009.

- [47] M.-K. von Renesse and K.-T. Sturm. Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Communications on Pure and Applied Mathematics*, 58(7):923–940, 2005.
- [48] F.-Y. Wang. *Harnack Inequalities and Applications for Stochastic Partial Differential Equations*. Springer Briefs in Mathematics. Springer, New York, 2013.
- [49] S. Zhang. Existence and application of optimal markovian coupling with respect to non-negative lower semi-continuous functions. *Acta Mathematica Sinica, English Series*, 16(2):261–270, 2000.

DEPARTMENT OF MATHEMATICS, XIAMEN UNIVERSITY MALAYSIA

Email address: `taushean.lim@xmu.edu.my`, `wyang1217@gmail.com`