ACADÉMIE
DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus
Mathématique

Bohdan Bulanyi

On the equivalence of static and dynamic weak optimal transport

Volume 364 (2026), p. 205-236

Online since: 17 March 2026

https://doi.org/10.5802/crmath.814

CENTRE MERSENNE

The Comptes Rendus. Mathématique *are a member of the*
Mersenne Center for open scientific publishing
www.centre-mersenne.org — e-ISSN : 1778-3569

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Mathématique
2026, Vol. 364, p. 205-236
https://doi.org/10.5802/crmath.814

Research article
Partial differential equations

On the equivalence of static and dynamic weak optimal transport

Bohdan Bulanyi [a]

[a] Université catholique de Louvain, Institut de Recherche en Mathématique et Physique, Chemin du Cyclotron 2 bte L7.01.01, 1348 Louvain-la-Neuve, Belgium

E-mail: bohdan.bulanyi@uclouvain.be

Abstract. We show that there is a PDE formulation in terms of Fokker–Planck equations for weak optimal transport problems. The main novelty is that we introduce a minimization problem involving Fokker–Planck equations in the extended sense of measure-valued solutions and prove that it is equal to the associated weak transport problem.

Keywords. Fokker–Planck equation, weak transport problem, duality, Laplace operator, Green function, subharmonic function, viscosity solution, Hamilton–Jacobi–Bellman equation.

2020 Mathematics Subject Classification. 49Q22, 28A33, 49J45, 49N15, 49Q20.

Funding. This work was partially supported by the Projet de Recherche T.0229.21 "Singular Harmonic Maps and Asymptotics of Ginzburg–Landau Relaxations" of the Fonds de la Recherche Scientifique–FNRS.

Manuscript received 12 June 2024, revised 26 September 2025, accepted 29 December 2025, online since 17 March 2026.

1. Introduction

1.1. *Statement of the problem*

Given probability measures μ, ν on $\mathbb{R}^d$, we denote by $\Pi(\mu,\nu)$ the set of all probability measures (called transport plans) on $\mathbb{R}^d\times\mathbb{R}^d$ whose first and second marginals are μ and ν, respectively. Disintegrating a transport plan $\gamma\in\Pi(\mu,\nu)$ with respect to its first marginal μ, we obtain a μ-almost everywhere uniquely determined family of probability measures $\{\gamma^x\}_{x\in\mathbb{R}^d}\subset\mathscr{P}(\mathbb{R}^d)$ such that $\gamma=\gamma^x\otimes\mu$. Let $G\colon\mathbb{R}^d\times\mathscr{P}(\mathbb{R}^d)\to[0,+\infty]$ be lower semicontinuous in an appropriate sense and $G(x,\cdot)$ be convex on $\mathscr{P}(\mathbb{R}^d)$ for each $x\in\mathbb{R}^d$. The weak transport problem is then defined by

$$H(\mu,\nu)=\inf\left\{\int_{\mathbb{R}^d}G(x,\gamma^x)\,\mathrm{d}\mu(x)\;\middle|\;\gamma\in\Pi(\mu,\nu)\right\}.$$

This problem was first introduced by Gozlan, Roberto, Samson and Tetali [17], and shortly thereafter by Alibert, Bouchitté and Champion [2]. The weak transport problem has been extensively studied in the literature: the results of existence and duality are established in [2,6,17]; the concept of C-monotonicity, which is analogous to cyclical monotonicity, was developed in [5,6,16] in order to provide a characterization of optimizers; in [1,2,5,8] the weak transport viewpoint is used to investigate martingale optimal transport problems; for applications of weak transport theory, the reader may consult [7].

In [19], Huesmann and Trevisan introduced the following minimization problem:

$$f_{\mathrm{FPE}}(\mu,\nu)=\inf\left\{\int_0^1\int_{\mathbb{R}^d} f\big(a_t(x)\big)\,\mathrm{d}\varrho_t(x)\,\mathrm{d}t\;\middle|\;\partial_t\varrho=\mathrm{tr}\Big(\frac{1}{2}\nabla^2 a\varrho\Big),\ \varrho_0=\mu,\ \varrho_1=\nu\right\},$$

where μ, ν are probabilities on $\mathbb{R}^d$ with finite second moment, the Fokker–Planck equation $\partial_t\varrho=\mathrm{tr}\big(\frac{1}{2}\nabla^2 a\varrho\big)$ in $(0,1)\times\mathbb{R}^d$ holds in the weak sense and f is q-admissible for some $q>1$ (namely, $f(a)$ behaves like $f(a)\sim|a|^q$ for a nonnegative symmetric real $d\times d$ matrix a, which means that there exist constants $c,C>0$, independent of a, such that $c|a|^q\le f(a)\le C|a|^q$; see [19, Sections 2 and 3]). They proved that

$$f_{\mathrm{FPE}}(\mu,\nu)=\inf\left\{\int_0^1\mathbb{E}\big[f\big(\langle\dot X\rangle_t\big)\big]\,\mathrm{d}t\right\},\tag{1}$$

where the infimum is taken over all martingales connecting μ and ν whose quadratic variation process $\langle X\rangle$ is absolutely continuous with respect to the Lebesgue measure (see [19, Theorem 3.3]). It is worth noting the following. Let $G(x,p)=f_{\mathrm{FPE}}(\delta_x,p)$ for each $x\in\mathbb{R}^d$ and for each probability measure p on $\mathbb{R}^d$ with finite second moment. Then, in view of [19, Theorem 4.3], the functional G is lower semicontinuous in an appropriate sense and $G(x,\cdot)$ is convex for each $x\in\mathbb{R}^d$. Taking into account (1) and using a standard measurable selection argument (see, for instance, [9, Section 7.7]), we have

$$f_{\mathrm{FPE}}(\mu,\nu)=\inf\left\{\int_{\mathbb{R}^d} f_{\mathrm{FPE}}(\delta_x,\gamma^x)\,\mathrm{d}\mu(x)\;\middle|\;\gamma\in\Pi(\mu,\nu)\right\}=H(\mu,\nu).$$

This provides an equivalent PDE formulation in terms of Fokker–Planck equations for the weak transport problem $H(\mu,\nu)$, where $G(x,p)=f_{\mathrm{FPE}}(\delta_x,p)$ and f is q-admissible for some $q>1$ (the reader may also consult the proof of [7, Theorem 8.2], where an equivalent SDE formulation for the weak transport problem is established using a measurable selection argument). We emphasize that the q-admissibility property of the function f was crucial in the proof of (1) in [19]. Namely, the equality (1) was proved in [19] only for strictly convex functions f behaving like $f(a)\sim|a|^q$ for some $q>1$ and for each nonnegative symmetric real $d\times d$ matrix a. The purpose of this paper is to provide an equivalent PDE formulation for $H(\mu,\nu)$ in terms of Fokker–Planck equations when the cost function f is not strictly convex and behaves like $f(a)\sim|a|$ (i.e., for example, for the trace or spectral radius of a nonnegative symmetric real $d\times d$ matrix), thereby narrowing the “gap” between the classical static formulation for $H(\mu,\nu)$ and the dynamic formulation involving a PDE.

In particular, we introduce the minimization problem

$$F(\mu,\nu):=\inf\left\{\int_0^1\int_{\mathbb{R}^d} f\Big(\frac{\mathrm{d}\lambda_t}{\mathrm{d}|\lambda_t|}\Big)\,\mathrm{d}|\lambda_t|\,\mathrm{d}t\;\middle|\;\partial_t\varrho=\mathrm{tr}\Big(\frac{1}{2}\nabla^2\lambda\Big),\ \varrho_0=\mu,\ \varrho_1=\nu\right\},$$

where f is a nonnegative convex positively 1-homogeneous function and the Fokker–Planck equation $\partial_t\varrho=\mathrm{tr}\big(\frac{1}{2}\nabla^2\lambda\big)$ in $(0,1)\times\mathbb{R}^d$ holds in the weak extended sense of measure-valued solutions (see Section 2.2), namely, λ does not have to be absolutely continuous with respect to ϱ, in contrast to the Fokker–Planck equations $\partial_t\varrho=\mathrm{tr}\big(\frac{1}{2}\nabla^2 a\varrho\big)$ in $(0,1)\times\mathbb{R}^d$ considered by Huesmann and Trevisan in [19], where the analysis is carried out for q-admissible costs. By defining $G(x,p)=F(\delta_x,p)$, our goal is to prove the equality $H(\mu,\nu)=F(\mu,\nu)$. Let us highlight that the *dynamic* problem $F(\mu,\nu)$ is in fact essentially *static* (see Proposition 9) and, unlike the case of q-admissible costs [19], we do not have an equivalent Benamou–Brenier type formulation (1) for $F(\mu,\nu)$ (see Remark 12), and hence we cannot use a standard measurable selection argument to prove the equality between $F(\mu,\nu)$ and $H(\mu,\nu)$. Therefore, we implement a new analytical approach.

1.2. *Main results*

In Proposition 9, we obtain a formulation that states that $F(\mu,\nu)$ is in some sense a *static* problem, namely

$$F(\mu,\nu)=\inf\left\{\int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\lambda}{\mathrm{d}|\lambda|}\right)\mathrm{d}|\lambda| \;\middle|\; \mathrm{tr}\left(\frac{1}{2}\nabla^2\lambda\right)=\nu-\mu\right\}.$$

In Theorem 13, we prove the existence of a solution to $F(\mu,\nu)$ whenever $F(\mu,\nu)$ is finite and derive a dual formulation for $F(\mu,\nu)$. This dual formulation is further refined in Theorem 34 in terms of invariant functions under G-transform $\varphi\mapsto\varphi^G$ (see (24)) in line with [2,17]. In Proposition 19, we prove the existence of a solution to $H(\mu,\nu)$ whenever $H(\mu,\nu)$ is finite and deduce a dual formulation for $H(\mu,\nu)$ relying on the G-transform, which is then refined in Theorems 26 and 33, where the supremum is taken over potentials whose opposite is fixed by the G-transform. Since the dual competitors for $H(\mu,\nu)$ are less regular than the dual competitors for $F(\mu,\nu)$, and the supremum is taken for the same functional for both $F(\mu,\nu)$ and $H(\mu,\nu)$ (see Remark 35), we first prove the equality $F(\mu,\nu)=H(\mu,\nu)$ when G-invariant functions can be approximated by smooth G-invariant functions in a suitable sense (see Theorems 37 and 39). Next, under the coercivity assumption on f, we characterize bounded continuous G-invariant functions as viscosity solutions of the Hamilton–Jacobi–Bellman equation (see Proposition 47). This, under the coercivity and growth assumptions on f, yields the dual formulation for $H(\mu,\nu)$, where the competitors are viscosity solutions of the Hamilton–Jacobi–Bellman equation (see Corollary 49). Then we perform a smoothing procedure and obtain the equality $F(\mu,\nu)=H(\mu,\nu)$ when the function f behaves like $f(a)\sim|a|$ (see Theorem 50). It is worth noting that the established equality between $F(\mu,\nu)$ and $H(\mu,\nu)$ allows us to recover the result of Ghoussoub, Kim and Lin on the existence of a Brownian martingale (see Remark 38), as well as the Strassen theorem (see Remark 40).

1.3. *Structure of the paper*

In Section 2, we introduce our main notation and recall some definitions. In Section 3, we announce our basic assumptions on the cost function f and introduce the minimization problem (5) involving Fokker–Planck equations in the extended sense of measure-valued solutions. We prove Proposition 9 and Theorem 13. We deduce that the functional F is convex, subadditive and lower semicontinuous in an appropriate sense (see Proposition 16). Next, we introduce the weak transport problem (23) (static formulation) associated with the problem (5) (PDE formulation) and prove Proposition 19. We develop the theory of subadditive cost functionals that appeared in [2, Section 6], where the role of $\mathbb{R}^d$ is replaced by the closure of a bounded open convex subset of $\mathbb{R}^d$. Unlike [2], this paper considers the set of probability measures on $\mathbb{R}^d$ that is not compact, which leads to additional difficulties. We obtain new results either under the coercivity assumption on f (see Propositions 24 and 25) or under the growth assumption on f (see Propositions 30 and 31). We prove Theorems 26, 33 and 34. In Section 4, the equality between $F(\mu,\nu)$ and $H(\mu,\nu)$ is proved: under the approximation assumptions on G-invariant functions (see Theorems 37 and 39); under the coercivity and growth assumptions on f (see Theorem 50). We consider in detail four examples (A)–(D) for which we compute $F(\mu,\nu)$ in terms of μ and ν.

2. Preliminaries

2.1. *Conventions and notation*

Conventions. In this paper, we say that a value is positive if it is strictly greater than zero, and a value is nonnegative if it is greater than or equal to zero. Euclidean spaces are endowed with the

Euclidean inner product $a \cdot b = a^{\mathrm{T}} b$ and the induced norm $|a| = \sqrt{a^{\mathrm{T}} a}$. By d we denote a positive integer.

Notation. For $r > 0$, $B_r(x)$, $\overline{B}_r(x)$, and $\partial B_r(x)$ denote, respectively, the open ball, the closed ball, and the $(d-1)$-sphere with center x and radius r. Let S_d be the space of real $d \times d$ symmetric matrices endowed with the Hilbert–Schmidt (or Frobenius) inner product $A : B = \operatorname{tr}(A^{\mathrm{T}} B)$ and the induced norm $\|A\| = \sqrt{A : A}$. We write I_d for the $d \times d$ identity matrix and S_d^+ (resp. S_d^{++}) for the set of symmetric nonnegative (resp. positive) definite real $d \times d$ matrices. We write $\mathscr{P}_2\big((0,1) \times \mathbb{R}^d\big)$ (resp. $\mathscr{P}_2(\mathbb{R}^d)$) for the set of probability measures μ on $(0,1) \times \mathbb{R}^d$ (resp. $\mathbb{R}^d$) such that $\int_{(0,1)\times\mathbb{R}^d} |x|^2 \,\mathrm{d}\mu(t,x)$ (resp. $\int_{\mathbb{R}^d} |x|^2 \,\mathrm{d}\mu(x)$) is finite. We endow $\mathscr{P}_2(\mathbb{R}^d)$ with the topology generated by the 2-Wasserstein distance (see [27, Definition 6.8 and Theorem 6.9]). We write $\mathscr{M}\big((0,1) \times \mathbb{R}^d, S_d^+\big)$ (resp. $\mathscr{M}(\mathbb{R}^d, S_d^+)$) for the set of measures λ on $(0,1) \times \mathbb{R}^d$ (resp. $\mathbb{R}^d$) with values in S_d such that for each Borel set $E \subset (0,1) \times \mathbb{R}^d$ (resp. $E \subset \mathbb{R}^d$), $\lambda(E) \in S_d^+$. For a measure λ, we denote by $|\lambda|$ its total variation. By $C_{\mathrm{b}}(\mathbb{R}^d)$ we denote the space of all real-valued bounded continuous functions on $\mathbb{R}^d$. We write $\mathscr{L}_{\mathrm{b}}(\mathbb{R}^d)$ (resp. $\mathscr{U}_{\mathrm{b}}(\mathbb{R}^d)$) for the set of all bounded lower (resp. upper) semicontinuous functions on $\mathbb{R}^d$. If $U \subset \mathbb{R}^d$ is Lebesgue measurable, then $L^1(U)$ will denote the space consisting of all real measurable functions on U that are integrable on U. By $L^1_{\mathrm{loc}}(U)$ we denote the space of functions u such that $u \in L^1(V)$ for all $V \Subset U$. We use the standard notation for Sobolev spaces. For an open set $U \subset \mathbb{R}^d$, denote by $W_0^{1,p}(U)$ the closure of $C_{\mathrm{c}}^\infty(U)$ in the Sobolev space $W^{1,p}(U)$, where $C_{\mathrm{c}}^\infty(U)$ is the space of functions in $C^\infty(U)$ with compact support in U. By $\mathscr{L}^r$ and $\mathscr{H}^r$ we denote the r-dimensional Lebesgue and Hausdorff measure, respectively. For each $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, $[\mu]$ and $\operatorname{var}(\mu)$ will denote the barycenter and the variance of μ, respectively. Namely, $[\mu] = \int_{\mathbb{R}^d} x \,\mathrm{d}\mu(x)$ and $\operatorname{var}(\mu) = \int_{\mathbb{R}^d} |x|^2 \,\mathrm{d}\mu(x) - \big|[\mu]\big|^2$.

2.2. *Fokker–Planck equation for general measures*

Let $C_{\mathrm{b}}^{1,2}\big((0,1) \times \mathbb{R}^d\big)$ be the space of functions continuously differentiable once in t and twice in x in $(0,1) \times \mathbb{R}^d$ with uniformly bounded $\partial_t \varphi$ and $\nabla_x^2 \varphi$.

Given a pair of measures $(\varrho, \lambda) \in \mathscr{P}_2\big((0,1) \times \mathbb{R}^d\big) \times \mathscr{M}\big((0,1) \times \mathbb{R}^d, S_d^+\big)$ such that one can disintegrate $\varrho = \varrho_t \otimes \mathscr{L}^1 \llcorner (0,1)$, $\lambda = \lambda_t \otimes \mathscr{L}^1 \llcorner (0,1)$, we say that the Fokker–Planck equation in the weak extended sense of measure-valued solutions

$$\partial_t \varrho = \operatorname{tr}\left(\frac{1}{2} \nabla^2 \lambda\right) \quad \text{in } (0,1) \times \mathbb{R}^d \tag{GFPE}$$

holds if

$$\int_0^1 |\lambda_t|(\mathbb{R}^d) \,\mathrm{d}t < +\infty$$

and for each $\varphi \in C_{\mathrm{b}}^{1,2}\big((0,1) \times \mathbb{R}^d\big)$ with (closed) support in $(0,1) \times \mathbb{R}^d$,

$$\int_0^1 \int_{\mathbb{R}^d} \partial_t \varphi(t,x) \,\mathrm{d}\varrho_t(x) \,\mathrm{d}t = -\int_0^1 \int_{\mathbb{R}^d} \frac{1}{2} \nabla_x^2 \varphi(t,x) : \mathrm{d}\lambda_t(x) \,\mathrm{d}t. \tag{2}$$

If $(\varrho, \lambda) \in \mathscr{P}_2\big((0,1) \times \mathbb{R}^d\big) \times \mathscr{M}\big((0,1) \times \mathbb{R}^d, S_d^+\big)$ is a solution to (GFPE), then there exists a narrowly continuous curve $(\widetilde{\varrho}_t)_{t \in [0,1]} \subset \mathscr{P}_2(\mathbb{R}^d)$ such that $\widetilde{\varrho}_t = \varrho_t$ for $\mathscr{L}^1$-a.e. $t \in (0,1)$. Moreover, if $\psi \in C_{\mathrm{b}}^2(\mathbb{R}^d)$ and $0 \le t_1 \le t_2 \le 1$, then

$$\int_{\mathbb{R}^d} \psi(x) \,\mathrm{d}\widetilde{\varrho}_{t_2}(x) - \int_{\mathbb{R}^d} \psi(x) \,\mathrm{d}\widetilde{\varrho}_{t_1}(x) = \int_{t_1}^{t_2} \int_{\mathbb{R}^d} \frac{1}{2} \nabla^2 \psi(x) : \mathrm{d}\lambda_t(x) \,\mathrm{d}t \tag{3}$$

(the reader may consult [4, Lemma 8.1.2] and [26, Remark 2.3]). Thus, for a solution (ϱ, λ) to (GFPE), without loss of generality, we shall assume that $(\varrho_t)_{t \in (0,1)}$ is narrowly continuous.

Let $C_{\mathrm{b}}^2(\mathbb{R}^d)$ be the space of functions twice continuously differentiable on $\mathbb{R}^d$ whose Hessian is uniformly bounded.

Proposition 1. *Let* (ϱ,λ) *solve* (GFPE), $\mu = w*-\lim_{t\searrow 0}\varrho_t$ *and* $\nu = w*-\lim_{t\nearrow 1}\varrho_t$. *If* $\psi\in C^2_{\mathrm{b}}(\mathbb{R}^d)$ *is convex, the function* $t\in[0,1]\mapsto\int_{\mathbb{R}^d}\psi\,\mathrm{d}\varrho_t$ *is nondecreasing. In particular,* $\int_{\mathbb{R}^d}\psi\,\mathrm{d}\mu\le\int_{\mathbb{R}^d}\psi\,\mathrm{d}\nu$.

Proof. For every $0\le t_1\le t_2\le 1$, using (3) and the fact that ψ is convex and $\lambda_t\in\mathscr{M}(\mathbb{R}^d,S^+_d)$, we have

$$\int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}\varrho_{t_2}(x)=\int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}\varrho_{t_1}(x)+\int_{t_1}^{t_2}\int_{\mathbb{R}^d}\frac{1}{2}\nabla^2\psi(x):\mathrm{d}\lambda_t(x)\,\mathrm{d}t\ge\int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}\varrho_{t_1}(x),$$

which completes our proof of Proposition 1. □

For convenience, we recall the next definition (see [24]).

Definition 2. *We say that two measures* $\mu,\nu\in\mathscr{P}_2(\mathbb{R}^d)$ *are in convex order, and we write* $\mu\le_{\mathrm{c}}\nu$, *if for each convex function* $\psi\colon\mathbb{R}^d\to\mathbb{R}$ *it holds*

$$\int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}\mu(x)\le\int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}\nu(x).$$

Remark 3. Notice that $\mu\le_{\mathrm{c}}\nu$ if and only if $\langle\nu-\mu,\varphi\rangle\ge 0$ for each convex function $\varphi\in C^2_{\mathrm{b}}(\mathbb{R}^d)$. Indeed, assume that $\langle\nu-\mu,\varphi\rangle\ge 0$ for each convex function $\varphi\in C^2_{\mathrm{b}}(\mathbb{R}^d)$. Then $\int_{\mathbb{R}^d}\psi\,\mathrm{d}\mu=\int_{\mathbb{R}^d}\psi\,\mathrm{d}\nu$ for each affine map $\psi\colon\mathbb{R}^d\to\mathbb{R}$. Since every convex function is nonnegative up to the addition of some affine map, the relation $\mu\le_{\mathrm{c}}\nu$ will be justified as soon as we show that $\langle\nu-\mu,\varphi\rangle\ge 0$ for each convex function $\varphi\colon\mathbb{R}^d\to[0,+\infty)$. For such a function φ, there exists a sequence $(\varphi_k)_{k\in\mathbb{N}}$ of convex functions $\varphi_k\colon\mathbb{R}^d\to[0,+\infty)$ such that $\varphi_k(\cdot)=\inf\{\varphi(y)+k|\cdot-y|\mid y\in\mathbb{R}^d\}$ is k-Lipschitz and $\varphi_k\nearrow\varphi$ as $k\to+\infty$. For each $k\in\mathbb{N}$, using convolution with $\eta_\varepsilon(\cdot)=\varepsilon^{-d}\eta(\cdot/\varepsilon)$, where $\varepsilon>0$ and η is a standard mollifier (as in Lemma 46), φ_k can be uniformly approximated on $\mathbb{R}^d$ by a sequence of nonnegative convex smooth k-Lipschitz functions. Again, using the above mollification procedure, each convex smooth Lipschitz function on $\mathbb{R}^d$ can be uniformly approximated by a sequence of convex functions lying in $C^2_{\mathrm{b}}(\mathbb{R}^d)$. After all, taking into account the above monotone and uniform convergences, for each convex function $\varphi\colon\mathbb{R}^d\to[0,+\infty)$ we can find a sequence $(\varphi_k)_{k\in\mathbb{N}}\subset C^2_{\mathrm{b}}(\mathbb{R}^d)$ of nonnegative convex functions such that $\int_{\mathbb{R}^d}\varphi_k\,\mathrm{d}\mu\to\int_{\mathbb{R}^d}\varphi\,\mathrm{d}\mu$ and $\int_{\mathbb{R}^d}\varphi_k\,\mathrm{d}\nu\to\int_{\mathbb{R}^d}\varphi\,\mathrm{d}\nu$ as $k\to+\infty$. This actually justifies the above criterion.

2.3. *Subharmonic functions*

Definition 4. *A function* $u\colon\mathbb{R}^d\to\mathbb{R}\cup\{-\infty\}$ *is said to be subharmonic if it satisfies the following conditions:*

(i) *u is not identically equal to* $-\infty$;
(ii) *u is upper semicontinuous;*
(iii) *for each* $x\in\mathbb{R}^d$ *and* $r>0$,

$$u(x)\le\fint_{\partial B_r(x)}u(y)\,\mathrm{d}\mathscr{H}^{d-1}(y),$$

where $\fint_{\partial B_r(x)}u(y)\,\mathrm{d}\mathscr{H}^{d-1}(y)=\frac{1}{\mathscr{H}^{d-1}(\partial B_r(x))}\int_{\partial B_r(x)}u(y)\,\mathrm{d}\mathscr{H}^{d-1}(y)$.

A function $v\colon\mathbb{R}^d\to\mathbb{R}\cup\{+\infty\}$ is said to be superharmonic if $u=-v$ is subharmonic. A function is harmonic if and only if it is both subharmonic and superharmonic. If $u\in C^2(\mathbb{R}^d)$, then u is subharmonic if and only if $\Delta u\ge 0$ in $\mathbb{R}^d$ (see, for instance, [22, Section 3.2]). In one dimension, a function is subharmonic if and only if it is convex; however, in dimension $d\ge 2$ the notions of subharmonicity and convexity are not equivalent (for more details, the reader may consult, for instance, [22]). We shall denote by $\mathscr{SH}(\mathbb{R}^d)$ the cone of all subharmonic functions on $\mathbb{R}^d$.

Next, we recall the notion of the so-called *subharmonic order*, which is stronger than the convex order in dimension $d\ge 2$ and is equivalent to the convex order in one dimension. It is worth mentioning that the convex order between a pair of probability measures μ and ν is a necessary and sufficient condition for the existence of a martingale coupling between μ and ν, as was proved by Strassen in [24]. The subharmonic order between μ and ν is a necessary and sufficient condition for the existence of a Brownian martingale coupling between μ and ν (namely, a transport plan $\gamma \in \Pi(\mu,\nu)$ which is a joint distribution of (B_0, B_τ), where $(B_t)_t$ is the Brownian motion with initial law μ, the law of B_τ is ν and τ is a possibly randomized stopping time for the Brownian filtration), as was proved by Ghoussoub, Kim, and Lin in [14].

Definition 5. *We say that two measures $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$ are in subharmonic order, and we write $\mu \le_{\mathrm{sh}} \nu$, if for each $u \in C^2_{\mathrm{b}}(\mathbb{R}^d) \cap \mathscr{SH}(\mathbb{R}^d)$ the following holds*

$$\int_{\mathbb{R}^d} u \, \mathrm{d}\mu \le \int_{\mathbb{R}^d} u \, \mathrm{d}\nu.$$

If $\mu \le_{\mathrm{sh}} \nu$, using a standard mollification procedure, we observe that $\int_{\mathbb{R}^d} u \, \mathrm{d}\mu \le \int_{\mathbb{R}^d} u \, \mathrm{d}\nu$ for each bounded or Lipschitz $u \in \mathscr{SH}(\mathbb{R}^d)$.

2.4. *Fenchel conjugate*

Let X be a normed vector space and X' be the topological dual space of X.

Definition 6. *Let $f\colon X \to \mathbb{R}\cup\{+\infty\}$ and $\mathrm{dom}(f) = \{x \in X \mid f(x) < +\infty\} \neq \emptyset$. The Fenchel conjugate $f^*\colon X' \to \mathbb{R}\cup\{+\infty\}$ of f at $x' \in X'$ is defined by*

$$f^*(x') = \sup\{x'(x) - f(x) \mid x \in X\}.$$

If $X = S_d$, then the following result holds. Recall that a function $f\colon S_d \to [0,+\infty]$ is said to be positively 1-homogeneous if

$$f(tA) = t f(A) \quad \text{for all } A, B \in S_d \text{ and } t \ge 0.$$

We denote by $C_{\mathrm{b}}(\mathbb{R}^d, S_d)$ the space of all bounded continuous functions on $\mathbb{R}^d$ with values into S_d.

Lemma 7. *Let $f\colon S_d \to [0,+\infty]$ be convex, lower semicontinuous and positively 1-homogeneous with $\mathrm{dom}(f) \subset S_d^+$ and $\mathrm{dom}(f) \neq \emptyset$. Let $\Psi\colon \big(C_{\mathrm{b}}(\mathbb{R}^d, S_d)\big)' \to [0,+\infty]$ be defined by*

$$\Psi(\sigma) = \begin{cases} \displaystyle\int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\sigma}{\mathrm{d}|\sigma|}\right) \mathrm{d}|\sigma| & \text{if } \sigma \in \mathscr{M}(\mathbb{R}^d, S_d^+),\ |\sigma|(\mathbb{R}^d) < +\infty, \\ +\infty & \text{otherwise.} \end{cases}$$

Then Ψ is convex and positively 1-homogeneous. For each $\xi \in C_{\mathrm{b}}(\mathbb{R}^d, S_d)$,

$$\Psi^*(\xi) = \begin{cases} 0 & \text{if } \xi(\mathbb{R}^d) \subset \mathrm{dom}(f^*), \\ +\infty & \text{otherwise.} \end{cases} \tag{4}$$

Furthermore, for each $\sigma \in \mathscr{M}(\mathbb{R}^d, S_d^+)$ such that $|\sigma|(\mathbb{R}^d) < +\infty$, Ψ is (weakly) lower semicontinuous at σ and $\Psi^{**}(\sigma) = \Psi(\sigma)$.

Proof. Given the definition of Ψ, to prove the convexity of Ψ, it suffices to show that if $t \in (0,1)$, $\sigma_1, \sigma_2 \in \mathscr{M}(\mathbb{R}^d, S_d^+)$ and $|\sigma_i|(\mathbb{R}^d) < +\infty$ for each $i \in \{1,2\}$, then $\Psi\big(t\sigma_1 + (1-t)\sigma_2\big) \le t\Psi(\sigma_1) + (1-$

$t)\Psi(\sigma_2)$. This estimate comes by using the facts that f is positively 1-homogeneous and convex. Indeed, we have

$$\begin{aligned}
\Psi\big(t\sigma_1+(1-t)\sigma_2\big) &= \int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}(t\sigma_1+(1-t)\sigma_2)}{\mathrm{d}\big|t\sigma_1+(1-t)\sigma_2\big|}\right)\mathrm{d}\big|t\sigma_1+(1-t)\sigma_2\big| \\
&= \int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}(t\sigma_1+(1-t)\sigma_2)}{\mathrm{d}\big(t|\sigma_1|+(1-t)|\sigma_2|\big)}\right)\mathrm{d}\big(t|\sigma_1|+(1-t)|\sigma_2|\big) \\
&\le t\int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\sigma_1}{\mathrm{d}\big(t|\sigma_1|+(1-t)|\sigma_2|\big)}\right)\mathrm{d}\big(t|\sigma_1|+(1-t)|\sigma_2|\big) \\
&\qquad + (1-t)\int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\sigma_2}{\mathrm{d}\big(t|\sigma_1|+(1-t)|\sigma_2|\big)}\right)\mathrm{d}\big(t|\sigma_1|+(1-t)|\sigma_2|\big) \\
&= t\int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\sigma_1}{\mathrm{d}|\sigma_1|}\right)\mathrm{d}|\sigma_1| + (1-t)\int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\sigma_2}{\mathrm{d}|\sigma_2|}\right)\mathrm{d}|\sigma_2| \\
&= t\Psi(\sigma_1)+(1-t)\Psi(\sigma_2)
\end{aligned}$$

(see Remark 10). The positive 1-homogeneity of Ψ comes from its definition and the positive 1-homogeneity of f.

Next, fix an arbitrary $\xi \in C_{\mathrm{b}}(\mathbb{R}^d, S_d)$. If $\xi(x_0) \notin \mathrm{dom}(f^*)$, then there exists $M \in \mathrm{dom}(f)$ such that $\xi(x_0) : M > f(M)$ (see (6)). Since ξ is continuous, there exists $r > 0$ such that $\xi(x) : M > f(M)$ for each $x \in B_r(x_0)$. Defining $\sigma_n = nM\mathscr{L}^d \llcorner B_r(x_0)$ and using the positive 1-homogeneity of f, we have

$$\Psi^*(\xi) \ge \int_{\mathbb{R}^d} \xi(x) : \mathrm{d}\sigma_n(x) - \int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\sigma_n}{\mathrm{d}|\sigma_n|}\right)\mathrm{d}|\sigma_n| = n\int_{B_r(x_0)} \big(\xi(x) : M - f(M)\big)\,\mathrm{d}x > 0.$$

Letting n tend to $+\infty$, we deduce that $\Psi^*(\xi) = +\infty$. On the other hand, if $\xi(\mathbb{R}^d) \subset \mathrm{dom}(f^*)$, then for each $\sigma \in \mathscr{M}(\mathbb{R}^d, S_d^+)$ such that $|\sigma|(\mathbb{R}^d) < +\infty$, we have

$$\int_{\mathbb{R}^d} \xi(x) : \mathrm{d}\sigma(x) - \int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\sigma}{\mathrm{d}|\sigma|}\right)\mathrm{d}|\sigma| \le 0$$

(see (6)) and hence $\Psi^*(\xi) = 0$. Thus, we have proved (4).

Inasmuch as Ψ is convex and for each $\sigma \in \mathscr{M}(\mathbb{R}^d, S_d^+)$ such that $|\sigma|(\mathbb{R}^d) < +\infty$, Ψ is (weakly) lower semicontinuous at σ (which is a consequence of [3, Theorem 2.34] and Remark 10), we have $\Psi^{**}(\sigma) = \Psi(\sigma)$ (see [10, Theorem 2.1(i)]). This completes our proof of Lemma 7. □

3. A PDE constrained optimization problem

For each $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$, we consider the following minimization problem

$$F(\mu,\nu) = \inf\left\{\int_0^1 \int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\lambda_t}{\mathrm{d}|\lambda_t|}\right)\mathrm{d}|\lambda_t|\,\mathrm{d}t \;\middle|\; \partial_t\varrho = \mathrm{tr}\left(\frac{1}{2}\nabla^2\lambda\right),\ \varrho_0 = \mu,\ \varrho_1 = \nu\right\}, \tag{5}$$

where the function $f\colon S_d \to [0,+\infty]$ with $\mathrm{dom}(f) := \{M \in S_d \mid f(M) < +\infty\} \subset S_d^+$ satisfies the following assumptions:

- (A0) $\mathrm{dom}(f) \cap S_d^{++}$ is dense in $\mathrm{dom}(f) \neq \emptyset$;
- (A1) f is convex and positively 1-homogeneous;
- (A2) f is lower semicontinuous.

It is worth noting that f is convex and positively 1-homogeneous if and only if f is subadditive and positively 1-homogeneous. Among the interesting examples of such functions, we emphasize the following

Examples.

- (A) $f(A) = t$ if $A = tI_d$ for some $t \ge 0$ and $f(A) = +\infty$ otherwise.
- (B) For some $B \in S_d^+$, $f(A) = A : B$ if $A \in S_d^+$ and $f(A) = +\infty$ otherwise.

(C) $f(A) = \mathrm{tr}(A)$ if $A \in S_d^+$ and $f(A) = +\infty$ otherwise.
(D) $f(A)$ is the largest eigenvalue of A if $A \in S_d^+$ and $f(A) = +\infty$ otherwise.

Notice that in the example (A) $\mathrm{dom}(f)$ is a proper convex subset of S_d^+, in the example (B) the function f is not coercive, and the example (C) is the particular case of the example (B) with $B = I_d$.

Using the 1-homogeneity of f, we compute its Fenchel conjugate. For each $A \in S_d$,

$$f^*(A) = \sup\{A : M - f(M) \mid M \in S_d\} = \begin{cases} 0 & \text{if } A : M \le f(M) \ \ \forall\, M \in \mathrm{dom}(f), \\ +\infty & \text{otherwise.} \end{cases} \tag{6}$$

In particular, $A \in \mathrm{dom}(f^*)$ for each $A \in -S_d^+$.

Remark 8. If $F(\mu,\nu) < +\infty$, then $\mu \le_{\mathrm{c}} \nu$ (see Proposition 1 and Remark 3), which implies that $[\mu] = [\nu]$.

Using the 1-homogeneity of f, we eliminate the time variable in (5) via the equation

$$\mathrm{tr}\left(\frac{1}{2}\nabla^2\lambda\right) = \nu - \mu \quad \text{in } \mathbb{R}^d, \tag{7}$$

which means that $\lambda \in \mathscr{M}(\mathbb{R}^d, S_d^+)$, $|\lambda|(\mathbb{R}^d) < +\infty$ and

$$\int_{\mathbb{R}^d} \frac{1}{2}\nabla^2\varphi(x) : \mathrm{d}\lambda(x) = \int_{\mathbb{R}^d} \varphi(x)\,\mathrm{d}\nu(x) - \int_{\mathbb{R}^d} \varphi(x)\,\mathrm{d}\mu(x) \quad \forall\, \varphi \in C_{\mathrm{b}}^2(\mathbb{R}^d). \tag{8}$$

We denote by $C_{\mathrm{b}}\big((0,1)\times\mathbb{R}^d, S_d\big)$ the space of all bounded continuous functions on $(0,1)\times\mathbb{R}^d$ with values into S_d.

Proposition 9. *For each $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$ the following holds*

$$F(\mu,\nu) = \inf\left\{\int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\lambda}{\mathrm{d}|\lambda|}\right)\mathrm{d}|\lambda| \;\middle|\; \mathrm{tr}\left(\frac{1}{2}\nabla^2\lambda\right) = \nu - \mu\right\}. \tag{9}$$

Remark 10. Since f is positively 1-homogeneous, for each positive finite measure m on $\mathbb{R}^d$ such that $|\lambda| \ll m$, it holds

$$\int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\lambda}{\mathrm{d}m}\right)\mathrm{d}m = \int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\lambda}{\mathrm{d}|\lambda|}\right)\mathrm{d}|\lambda|.$$

Proof of Proposition 9. Let $\lambda \in \mathscr{M}(\mathbb{R}^d, S_d^+)$ be a solution to (7). For each $t \in [0,1]$, define

$$\varrho_t = (1-t)\mu + t\nu \in \mathscr{P}_2(\mathbb{R}^d). \tag{10}$$

Fix an arbitrary $\varphi \in C_{\mathrm{b}}^{1,2}\big((0,1)\times\mathbb{R}^d\big)$ with (closed) support in $(0,1)\times\mathbb{R}^d$. Then, using (10), Fubini's theorem and integrating by parts, we have

$$\int_0^1\int_{\mathbb{R}^d} \partial_t\varphi(t,x)\,\mathrm{d}\varrho_t(x)\,\mathrm{d}t = \int_{\mathbb{R}^d}\int_0^1 \varphi(t,x)\,\mathrm{d}t\,\mathrm{d}\mu(x) - \int_{\mathbb{R}^d}\int_0^1 \varphi(t,x)\,\mathrm{d}t\,\mathrm{d}\nu(x). \tag{11}$$

Since $\mathrm{tr}\big(\frac{1}{2}\nabla^2\lambda\big) = \nu - \mu$ and $\int_0^1 \varphi(t,\cdot)\,\mathrm{d}t \in C_{\mathrm{b}}^2(\mathbb{R}^d)$, using (8) and Fubini's theorem, we deduce that

$$\int_0^1\int_{\mathbb{R}^d} \frac{1}{2}\nabla_x^2\varphi(t,x) : \mathrm{d}\lambda(x)\,\mathrm{d}t = \int_{\mathbb{R}^d}\int_0^1 \varphi(t,x)\,\mathrm{d}t\,\mathrm{d}\nu(x) - \int_{\mathbb{R}^d}\int_0^1 \varphi(t,x)\,\mathrm{d}t\,\mathrm{d}\mu(x). \tag{12}$$

Combining (11) and (12), we get

$$\int_0^1\int_{\mathbb{R}^d} \partial_t\varphi(t,x)\,\mathrm{d}\varrho_t(x)\,\mathrm{d}t = -\int_0^1\int_{\mathbb{R}^d} \frac{1}{2}\nabla_x^2\varphi(t,x) : \mathrm{d}\lambda(x)\,\mathrm{d}t,$$

and hence the pair $\big(\varrho_t \otimes \mathscr{L}^1 \llcorner (0,1),\ \lambda \otimes \mathscr{L}^1 \llcorner (0,1)\big)$ is a solution to (GFPE). Thus,

$$F(\mu,\nu) \le \inf\left\{\int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\lambda}{\mathrm{d}|\lambda|}\right)\mathrm{d}|\lambda| \;\middle|\; \mathrm{tr}\left(\frac{1}{2}\nabla^2\lambda\right) = \nu - \mu\right\}. \tag{13}$$

Now assume that $F(\mu,\nu) < +\infty$ and let $\big(\varrho_t \otimes \mathscr{L}^1 \llcorner (0,1),\, \lambda_t \otimes \mathscr{L}^1 \llcorner (0,1)\big)$ be an admissible pair for (5). Using (3) with $t_1 = 0$ and $t_2 = 1$, for each $\varphi \in C_{\mathrm{b}}^2(\mathbb{R}^d)$, we obtain

$$\int_0^1 \int_{\mathbb{R}^d} \frac{1}{2}\nabla^2\varphi(x) : \mathrm{d}\lambda_t(x)\,\mathrm{d}t = \int_{\mathbb{R}^d} \varphi(x)\,\mathrm{d}\nu(x) - \int_{\mathbb{R}^d} \varphi(x)\,\mathrm{d}\mu(x). \tag{14}$$

Define

$$\lambda = \pi^x_{\#}\big(\lambda_t \otimes \mathscr{L}^1 \llcorner (0,1)\big) \in \mathscr{M}(\mathbb{R}^d, S_d^+),$$

where $\pi^x(t,x) = x$ for each $(t,x) \in [0,1] \times \mathbb{R}^d$. Then $|\lambda|(\mathbb{R}^d) < +\infty$ and for each $\xi \in C_{\mathrm{b}}(\mathbb{R}^d, S_d)$,

$$\int_{\mathbb{R}^d} \xi(x) : \mathrm{d}\lambda(x) = \int_0^1 \int_{\mathbb{R}^d} \xi\big(\pi^x(t,x)\big) : \mathrm{d}\lambda_t(x)\,\mathrm{d}t. \tag{15}$$

Gathering together (14) and (15), we get

$$\int_{\mathbb{R}^d} \frac{1}{2}\nabla^2\varphi(x) : \mathrm{d}\lambda(x) = \int_{\mathbb{R}^d} \varphi(x)\,\mathrm{d}\nu(x) - \int_{\mathbb{R}^d} \varphi(x)\,\mathrm{d}\mu(x) \quad \forall\, \varphi \in C_{\mathrm{b}}^2(\mathbb{R}^d).$$

Thus, λ is a solution to (7). Let Ψ be the convex function of Lemma 7. According to Lemma 7,

$$\begin{aligned}
\int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\lambda}{\mathrm{d}|\lambda|}\right)\mathrm{d}|\lambda| &= \sup\left\{\int_{\mathbb{R}^d} \xi(x) : \mathrm{d}\lambda(x) \,\middle|\, \xi \in C_{\mathrm{b}}(\mathbb{R}^d, S_d),\ \xi(\mathbb{R}^d) \subset \operatorname{dom}(f^*)\right\}\\
&= \sup\left\{\int_0^1 \int_{\mathbb{R}^d} \xi\big(\pi^x(t,x)\big) : \mathrm{d}\lambda_t(x)\,\mathrm{d}t \,\middle|\, \xi \in C_{\mathrm{b}}(\mathbb{R}^d, S_d),\ \xi(\mathbb{R}^d) \subset \operatorname{dom}(f^*)\right\}\\
&\le \sup\left\{\int_0^1 \int_{\mathbb{R}^d} \xi(t,x) : \mathrm{d}\lambda_t(x)\,\mathrm{d}t \,\middle|\, \xi \in C_{\mathrm{b}}\big((0,1)\times\mathbb{R}^d, S_d\big),\ \xi\big((0,1)\times\mathbb{R}^d\big) \subset \operatorname{dom}(f^*)\right\}\\
&\le \int_0^1 \int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\lambda_t}{\mathrm{d}|\lambda_t|}\right)\mathrm{d}|\lambda_t|\,\mathrm{d}t.
\end{aligned}$$

Using this together with the fact that λ is a solution to (7) and taking into account (13), we deduce (9), which completes our proof of Proposition 9. □

As a byproduct of the proof of Proposition 9, we obtain the next result.

Corollary 11. *If the infimum in* (5) *is achieved on* $\big(\varrho_t \otimes \mathscr{L}^1 \llcorner (0,1),\, \lambda_t \otimes \mathscr{L}^1 \llcorner (0,1)\big)$, *then the infimum in* (9) *is achieved on* $\lambda = \pi^x_{\#}\big(\lambda_t \otimes \mathscr{L}^1 \llcorner (0,1)\big)$. *Conversely, if the infimum in* (9) *is achieved on* λ, *then the infimum in* (5) *is achieved on* $\big(\varrho_t \otimes \mathscr{L}^1 \llcorner (0,1),\, \lambda_t \otimes \mathscr{L}^1 \llcorner (0,1)\big)$, *where* $\varrho_t = (1-t)\mu + t\nu$ *and* $\lambda_t = \lambda$.

Remark 12. In general, one cannot find a martingale $(X_t)_{t\in[0,1]}$ (for the definition, see, for instance, [19, Section 2]) with continuous paths whose marginals are $\varrho_t = (1-t)\mu + t\nu$ for $t \in [0,1]$. Indeed, if $\mu = \delta_{\frac{x+y}{2}}$ and $\nu = \frac{1}{2}(\delta_x + \delta_y)$, the optimal curve $\varrho_t = (1-t)\mu + t\nu$ is absolutely continuous in the 1-Wasserstein distance, but it is not absolutely continuous in the q-Wasserstein distance with $q > 1$, which from the particle point of view means that the particles must jump from $\frac{x+y}{2}$ to x or y at a certain rate. This phenomenon occurs because, in contrast to [19], in our case the cost f is not q-admissible, which correlates with the fact that for a solution (ϱ, λ) of the problem (5), λ does not have to be absolutely continuous with respect to ϱ (the reader may also consult [11, Remark 3.3]).

3.1. *Existence and duality*

We introduce a dual formulation to (5) and prove the existence of a minimizer when $F(\mu,\nu) < +\infty$.

Theorem 13. *For each* $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ *the following equality holds:*

$$F(\mu,\nu) = \sup\left\{\langle \nu - \mu, \psi\rangle \,\middle|\, \psi \in C_{\mathrm{b}}^2(\mathbb{R}^d),\ -f^*\left(\frac{1}{2}\nabla^2\psi\right) = 0 \text{ in } \mathbb{R}^d\right\}. \tag{16}$$

Moreover, if $F(\mu,\nu) < +\infty$, *then the infimum in* (5) *(and in* (9)*) is actually a minimum.*

Remark 14. The function f is not strictly convex and, generally speaking, is not coercive (see, for instance, the example (B)). Notice that in [19, Theorem 4.3] the strict convexity and q-coercivity (for some $q > 1$; see [19, Section 2] for the definition) of the cost function are used to prove the existence of a minimizer of the primal problem. In particular, the fact that the Fenchel conjugate of the cost function is finite on $\{tI_d \mid t \geq 0\}$ (since the cost function in [19] is q-coercive) is used to prove that a minimizer in the Fenchel-Rockafellar duality theorem (see [19, (4.5) in Theorem 4.3]) solves the Fokker–Planck equation, where the diffusion term is weighted accordingly with the mass. In our setting, if $\operatorname{dom}(f^*)$ contained $\{tI_d \mid t \geq 0\}$, we would have $f = +\infty$ identically on S_d^{++}. It is also worth noting that our dual formulation (16) cannot be derived from the results of [25], established using stochastic control theory. Indeed, in view of Remark 12, the problem (5) and the semimartingale transportation problems studied in [25] are different. Furthermore, the existence and duality results in [25] are established under the crucial coercivity assumption (see [25, Assumption 3.3]), which guarantees tightness of any minimizing sequences of the (primal) problems in [25] and does not hold for the cost functions behaving like $c(a) \sim |a|$ for $a \in S_d^+$, in particular, for our cost f (notice also that [7, Assumption 8.1(3)] does not hold for f). Thirdly, the coercivity of f is not required in Theorem 13.

Proof of Theorem 13. Let Ψ be the convex function of Lemma 7. We write the dual pairing as $\sigma(\xi)$ for $\sigma \in \left(C_{\mathrm{b}}(\mathbb{R}^d, S_d)\right)'$ and $\xi \in C_{\mathrm{b}}(\mathbb{R}^d, S_d)$. Then the following implication holds: if $\Psi^{**}(\sigma) < +\infty$, then $\sigma(\xi) \geq 0$ for each $\xi \in C_{\mathrm{b}}(\mathbb{R}^d, S_d^+)$. Assume by contradiction that $\Psi^{**}(\sigma) < +\infty$ and $\sigma(\xi) > 0$ for some $\xi \in C_{\mathrm{b}}(\mathbb{R}^d, S_d)$ such that $-\xi \in C_{\mathrm{b}}(\mathbb{R}^d, S_d^+)$. Since $\Psi^*(t\xi) = 0$ for each $t \geq 0$, $\Psi^{**}(\sigma) \geq t\sigma(\xi) > 0$. Letting t tend to $+\infty$, we obtain $\Psi^{**}(\sigma) = +\infty$, which leads to a contradiction and proves the above implication.

Next, following [19], we say that $\xi \in C_{\mathrm{b}}(\mathbb{R}^d, S_d)$ is represented by $\varphi \in C_{\mathrm{b}}^2(\mathbb{R}^d)$ if $\xi = -\frac{1}{2}\nabla^2\varphi$. We define $\Theta\colon C_{\mathrm{b}}(\mathbb{R}^d, S_d) \to \mathbb{R} \cup \{+\infty\}$ by

$$\Theta(\xi) = \begin{cases} \langle \mu - \nu, \varphi \rangle & \text{if } \xi \text{ is represented by } \varphi, \\ +\infty & \text{otherwise.} \end{cases}$$

If φ_1 and φ_2 represent ξ, then $\frac{1}{2}\nabla^2(\varphi_1 - \varphi_2) = 0$, and hence $\varphi_1(x) = a + y \cdot x + \varphi_2(x)$ for some fixed $a \in \mathbb{R}$ and $y \in \mathbb{R}^d$. Without loss of generality, we can assume that $\int_{\mathbb{R}^d} (a + y \cdot x)\,\mathrm{d}\mu(x) = \int_{\mathbb{R}^d} (a + y \cdot x)\,\mathrm{d}\nu(x)$, because otherwise $F(\mu, \nu)$ and the supremum in (16) are equal to $+\infty$: $F(\mu, \nu) = +\infty$ because μ and ν would not be in convex order; the supremum in (16) is equal to $+\infty$ by letting $\psi(x) = ty \cdot x$, which satisfies $-f^*\left(\frac{1}{2}\nabla^2\psi\right) = 0$ in $\mathbb{R}^d$, and letting t tend to $\pm\infty$ depending on the sign of the difference. This implies that $\langle \mu - \nu, \varphi_1 - \varphi_2 \rangle = 0$. Thus, Θ does not depend on the choice of φ. It is also worth noting that the set of represented mappings $\xi \in C_{\mathrm{b}}(\mathbb{R}^d, S_d)$ is a linear subspace on which Θ is linear. In particular, Θ is positively 1-homogeneous with the Fenchel conjugate

$$\Theta^*(\sigma) = \sup\left\{\sigma(\xi) + \langle \nu - \mu, \varphi \rangle \,\middle|\, \xi \text{ is represented}\right\}$$

taking values in $\{0, +\infty\}$ and $\Theta^*(\sigma) = 0$ if and only if

$$\int_{\mathbb{R}^d} \frac{1}{2}\nabla^2\varphi : \mathrm{d}\sigma = \int_{\mathbb{R}^d} \varphi\,\mathrm{d}\nu - \int_{\mathbb{R}^d} \varphi\,\mathrm{d}\mu \quad \forall\, \varphi \in C_{\mathrm{b}}^2(\mathbb{R}^d), \tag{17}$$

where we interpret the integral on the left-hand side of the above equality as a duality pairing.

We have proved that if $\Psi^{**}(\sigma) < +\infty$, then σ is a nonnegative bounded linear functional. We claim that σ is tight and hence induced by a measure. Let $g\colon \mathbb{R} \to [0, 1]$ be a smooth nondecreasing function such that $g(t) = 0$ for $t \in (-\infty, 1/2]$, $g(t) = 1$ for $t \in [1, +\infty)$ and $|g'(t)|, |g''(t)| \leq 4$ for $t \in \mathbb{R}$. Define the function $h\colon \mathbb{R} \to [0, +\infty)$ by $h(t) = \int_{-\infty}^t g(s)\,\mathrm{d}s$. Then h is a smooth function, $h'(t) = g(t)$ and $h''(t) = g'(t) \geq 0$ for $t \in \mathbb{R}$. For $L > 0$ and $x \in \mathbb{R}^d$, we set $\varphi^L(x) = h\left(|x|^2 - L^2\right)$. Notice that $\xi^L(x) = -\left(g\left(|x|^2 - L^2\right)I_d + 2g'\left(|x|^2 - L^2\right)x \otimes x\right)$ is represented by φ^L and $-\xi^L \geq I_d$ on

$\mathbb{R}^d \setminus B_{\sqrt{L^2+1}}(0)$. Since $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ and $h(|x|^2 - L^2) \le |x|^2 - L^2 \le |x|^2$ on $\mathbb{R}^d \setminus B_L(0)$, for each fairly small $\varepsilon > 0$, there exists $L > 0$ large enough such that

$$\int_{\mathbb{R}^d} h(|x|^2 - L^2)\, \mathrm{d}\nu(x) < \varepsilon. \tag{18}$$

For each $\xi \in C_{\mathrm{b}}(\mathbb{R}^d, S_d)$ such that $|\xi| \le 1$ and $\mathrm{supp}(\xi) \subset \mathbb{R}^d \setminus B_{\sqrt{L^2+1}}(0)$, it holds $\xi \le -\xi^L$ and $-\xi \le -\xi^L$. Using this, the facts that σ is a nonnegative linear functional and $\xi^L = -\frac{1}{2}\nabla^2\varphi^L$, (17), (18) and the fact that μ is a nonnegative measure, we deduce the following:

$$\big|\varphi(\cdot)\big| \le \sigma(-\xi^L) = \int_{\mathbb{R}^d} h(|x|^2 - L^2)\, \mathrm{d}\big(\nu(x) - \mu(x)\big) < \varepsilon,$$

which proves that σ is tight and hence induced by a measure.

The mapping I_d is represented by $\varphi(x) = -|x|^2$, Ψ^* is continuous at $-I_d$ and $\Theta(I_d) < +\infty$. After all, applying the formula for the conjugate of the sum $\Psi^*(-\cdot) + \Theta(\cdot)$ at $0 \in \big(C_{\mathrm{b}}(\mathbb{R}^d, S_d)\big)'$ (see, for instance, [10, Proposition 2.3(i)] or [12, Theorem 1.12]), we obtain

$$\inf\big\{\Psi^{**}(\sigma) + \Theta^*(\sigma) \;\big|\; \sigma \in \big(C_{\mathrm{b}}(\mathbb{R}^d, S_d)\big)'\big\} = \sup\big\{-\Psi^*(-\xi) - \Theta(\xi) \;\big|\; \xi \in C_{\mathrm{b}}(\mathbb{R}^d, S_d)\big\}, \tag{19}$$

where the infimum is actually a minimum if the supremum, coinciding with the supremum in (16), is finite. The latter holds if and only if $F(\mu, \nu) < +\infty$. Indeed, if the supremum in (19) is finite, then according to [10, Proposition 2.3(ii)] (or [12, Theorem 1.12]), the infimum in (19) is actually a minimum and if σ is a minimizer, then we have proved that $\sigma \in \mathscr{M}(\mathbb{R}^d, S_d^+)$, $|\sigma|(\mathbb{R}^d) < +\infty$, σ solves (7) for μ and ν (see (17)) and $\Psi(\sigma) = \Psi^{**}(\sigma)$. This, together with Proposition 9, implies that the left-hand side in (19) coincides with $F(\mu, \nu)$, and the infimum in (5) (and in (9)) is actually a minimum. On the other hand, if $F(\mu, \nu) < +\infty$, then there exists $\sigma \in \mathscr{M}(\mathbb{R}^d, S_d^+)$ solving (7) for μ and ν. For each $\psi \in C_{\mathrm{b}}^2(\mathbb{R}^d)$ such that $-f^*\big(\frac{1}{2}\nabla^2\psi\big) = 0$ in $\mathbb{R}^d$, we have $\frac{1}{2}\nabla^2\psi : \frac{\mathrm{d}\sigma}{\mathrm{d}|\sigma|} \le f\big(\frac{\mathrm{d}\sigma}{\mathrm{d}|\sigma|}\big)$ $|\sigma|$-a.e. on $\mathbb{R}^d$. Hence

$$\langle \nu - \mu, \psi \rangle = \int_{\mathbb{R}^d} \frac{1}{2}\nabla^2\psi : \mathrm{d}\sigma \le \int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\sigma}{\mathrm{d}|\sigma|}\right) \mathrm{d}|\sigma|,$$

which implies that the supremum in (19) is less than or equal to $\int_{\mathbb{R}^d} f\big(\frac{\mathrm{d}\sigma}{\mathrm{d}|\sigma|}\big)\mathrm{d}|\sigma| < +\infty$. This completes our proof of Theorem 13. □

Corollary 15. *For each $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ the following estimate holds:*

$$F(\mu, \nu) \ge f\left(\int_{\mathbb{R}^d} x \otimes x\big(\mathrm{d}\nu(x) - \mathrm{d}\mu(x)\big)\right). \tag{20}$$

Proof of Corollary 15. Let $A \in \mathrm{dom}(f^*)$ be arbitrary and $\psi(x) = Ax \cdot x$ for each $x \in \mathbb{R}^d$. Then we have $-f^*\big(\frac{1}{2}\nabla^2\psi\big) = -f^*(A) = 0$ in $\mathbb{R}^d$, which, in view of Theorem 13, yields

$$F(\mu, \nu) \ge \int_{\mathbb{R}^d} Ax \cdot x\, \mathrm{d}\nu(x) - \int_{\mathbb{R}^d} Ax \cdot x\, \mathrm{d}\mu(x) = A : \left(\int_{\mathbb{R}^d} x \otimes x\, \mathrm{d}\nu(x) - \int_{\mathbb{R}^d} x \otimes x\, \mathrm{d}\mu(x)\right). \tag{21}$$

Since $A \in \mathrm{dom}(f^*)$ was arbitrarily chosen, $f^* = 0$ on $\mathrm{dom}(f^*)$ and $f = f^{**}$ (this comes from (A1), (A2) and [10, Theorem 2.1(i)]), (21) implies (20), which completes our proof of Corollary 15. □

3.2. *Lower semicontinuity, convexity and subadditivity*

Hereinafter, we use the notation $\Phi_2(\mathbb{R}^d)$ for the space of all functions $\varphi \in C(\mathbb{R}^d)$ such that there exists a constant $C > 0$ such that $\big|\varphi(x)\big| \le C\big(1 + |x|^2\big)$ for each $x \in \mathbb{R}^d$.

Proposition 16. *The following assertions hold:*

(i) *F is convex on $\mathscr{P}_2(\mathbb{R}^d) \times \mathscr{P}_2(\mathbb{R}^d)$ and lower semicontinuous with respect to the weak topology on $\mathscr{P}_2(\mathbb{R}^d) \times \mathscr{P}_2(\mathbb{R}^d)$ in duality with $\Phi_2(\mathbb{R}^d) \times \Phi_2(\mathbb{R}^d)$;*

(ii) *for each choice of* $\mu_1, \mu_2, \mu_3 \in \mathscr{P}_2(\mathbb{R}^d)$,

$$F(\mu_1, \mu_3) \le F(\mu_1, \mu_2) + F(\mu_2, \mu_3).$$

Proof. It is a direct consequence of Theorem 13 that F is convex and lower semicontinuous with respect to the specified product topology, since it is represented as the supremum of the family consisting of linear functionals that are continuous with respect to this topology. This proves (i).

Let us prove (ii). For each $\psi \in C^2_{\mathrm{b}}(\mathbb{R}^d)$ such that $-f^*\left(\frac{1}{2}\nabla^2\psi\right) = 0$ in $\mathbb{R}^d$, using Theorem 13, we have

$$\langle \mu_3 - \mu_1, \psi\rangle = \langle \mu_2 - \mu_1, \psi\rangle + \langle \mu_3 - \mu_2, \psi\rangle \le F(\mu_1, \mu_2) + F(\mu_2, \mu_3),$$

which implies (ii) and completes our proof of Proposition 16. □

3.3. *Associated weak transport problem*

We define the cost function $G\colon \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to [0, +\infty]$ by

$$G(x, p) = F(\delta_x, p). \tag{22}$$

By Proposition 16, G is lower semicontinuous in (x, p) and convex in p. For each $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$, we consider the following weak transport problem

$$\inf\left\{ \int_{\mathbb{R}^d} G(x, \gamma^x)\, \mathrm{d}\mu(x) \;\middle|\; \gamma \in \Pi(\mu, \nu)\right\}$$

and define the functional $H\colon \mathscr{P}_2(\mathbb{R}^d) \times \mathscr{P}_2(\mathbb{R}^d) \to [0, +\infty]$ by

$$H(\mu, \nu) = \inf\left\{ \int_{\mathbb{R}^d} G(x, \gamma^x)\, \mathrm{d}\mu(x) \;\middle|\; \gamma \in \Pi(\mu, \nu)\right\}. \tag{23}$$

Since $G(x, \delta_x) = 0$ for each $x \in \mathbb{R}^d$ and $\gamma^x \otimes \mu \in \Pi(\mu, \mu)$ when $\gamma^x = \delta_x$ for μ-a.e. $x \in \mathbb{R}^d$,

$$H(\mu, \mu) = F(\mu, \mu) = 0 \quad \forall\, \mu \in \mathscr{P}_2(\mathbb{R}^d),$$

which implies that the functional H is proper (i.e., $\mathrm{dom}(H) \neq \emptyset$).

We shall prove the equality $F = H$. First, we show that $H \ge F$, which is a consequence of Theorem 13. To establish the converse inequality, which is a delicate matter, we develop the dual result of [6] and the theory for subadditive costs that appeared in [2, Section 6], where the role of $\mathbb{R}^d$ is replaced by the closure of a bounded open convex subset of $\mathbb{R}^d$. The main difficulty is that, unlike [2], we work with the set of probability measures on $\mathbb{R}^d$, which is not compact with respect to the weak topology.

Proposition 17. *For each* $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$, $H(\mu, \nu) \ge F(\mu, \nu)$.

Proof. Fix an arbitrary $\gamma = \gamma^x \otimes \mu \in \Pi(\mu, \nu)$. According to the definition of G and Theorem 13, for μ-a.e. $x \in \mathbb{R}^d$ and for each $\psi \in C^2_{\mathrm{b}}(\mathbb{R}^d)$ such that $-f^*\left(\frac{1}{2}\nabla^2\psi\right) = 0$ in $\mathbb{R}^d$, we have

$$G(x, \gamma^x) \ge \int_{\mathbb{R}^d} \psi(y)\, \mathrm{d}\gamma^x(y) - \psi(x).$$

Integrating both sides of the above inequality over $\mathbb{R}^d$ with respect to μ, we obtain

$$\int_{\mathbb{R}^d} G(x, \gamma^x)\, \mathrm{d}\mu(x) \ge \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \psi(y)\, \mathrm{d}\gamma^x(y)\, \mathrm{d}\mu(x) - \int_{\mathbb{R}^d} \psi(x)\, \mathrm{d}\mu(x) = \langle \nu - \mu, \psi\rangle.$$

This, since $\gamma \in \Pi(\mu, \nu)$ and $\psi \in C^2_{\mathrm{b}}(\mathbb{R}^d)$ satisfying $-f^*\left(\frac{1}{2}\nabla^2\psi\right) = 0$ in $\mathbb{R}^d$, according to Theorem 13, completes our proof of Proposition 17. □

For each $x \in \mathbb{R}^d$ and for each universally measurable function $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ satisfying the estimate $|\varphi(\cdot)| \le C(1+|\cdot|^2)$ for some constant $C>0$, we define

$$\varphi^G(x) := \inf\left\{ \int_{\mathbb{R}^d} \varphi \,\mathrm{d}p + G(x,p) \;\middle|\; p \in \mathscr{P}_2(\mathbb{R}^d) \right\}. \tag{24}$$

Inasmuch as $G(x,\delta_x)=0$,

$$\varphi^G \le \varphi. \tag{25}$$

We denote by $\Phi_{\mathrm{bb},2}(\mathbb{R}^d)$ the subset of functions in $\Phi_2(\mathbb{R}^d)$ which are bounded from below.

Remark 18. If $\varphi \in \Phi_{\mathrm{bb},2}(\mathbb{R}^d)$, then φ^G is lower semianalytic (and hence universally measurable, see [9, Proposition 7.47]), bounded from below, $|\varphi^G(\cdot)| \le C(1+|\cdot|^2)$ for some constant $C>0$ and the integral $\int_{\mathbb{R}^d} \varphi^G \,\mathrm{d}p$ is well defined for all $p \in \mathscr{P}_2(\mathbb{R}^d)$. In particular, for each $x \in \mathbb{R}^d$, we can define $\varphi^{GG}(x)$.

Proposition 19. *The following assertions hold:*

(i) *H is convex on $\mathscr{P}_2(\mathbb{R}^d) \times \mathscr{P}_2(\mathbb{R}^d)$ and lower semicontinuous with respect to the weak topology on $\mathscr{P}_2(\mathbb{R}^d) \times \mathscr{P}_2(\mathbb{R}^d)$ in duality with $\Phi_2(\mathbb{R}^d) \times \Phi_2(\mathbb{R}^d)$; if $H(\mu,\nu) < +\infty$, then the weak transport problem* (23) *admits a solution;*

(ii) *for each $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$,*

$$H(\mu,\nu) = \sup\left\{ \int_{\mathbb{R}^d} \varphi^G \,\mathrm{d}\mu - \int_{\mathbb{R}^d} \varphi \,\mathrm{d}\nu \;\middle|\; \varphi \in \Phi_{\mathrm{bb},2}(\mathbb{R}^d) \right\}. \tag{26}$$

Proof. According to Proposition 16(i), $(x,p) \mapsto G(x,p)$ is convex in p and lower semicontinuous in (x,p) with respect to the product topology on $\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d)$, where the topology on $\mathbb{R}^d$ is generated by the Euclidean distance and the topology on $\mathscr{P}_2(\mathbb{R}^d)$ is generated by the 2-Wasserstein distance (see [27, Definition 6.8 and Theorem 6.9]). Then, using [6, Theorem 2.9], we deduce that H is lower semicontinuous with respect to the weak topology on $(\mathscr{P}_2(\mathbb{R}^d))^2$ in duality with $(\Phi_2(\mathbb{R}^d))^2$ and prove that (23) admits a solution whenever $H(\mu,\nu) < +\infty$. Applying [6, Theorem 3.1], we obtain the dual formulation (26), which implies the convexity of H. This completes our proof of Proposition 19. □

To develop the theory of subadditive cost functionals, which appeared earlier in [2, Section 6], where the role of $\mathbb{R}^d$ is replaced by the closure of a bounded open convex subset of $\mathbb{R}^d$, we need to introduce some additional assumptions on f, namely either the coercivity (see the example (A)), or the growth assumption (see the example (B)), or both (see the examples (C), (D)). In particular, we introduce the following assumptions:

(A3) f is coercive;

(A4) $\mathrm{dom}(f) = S_d^+$ and there exists $\kappa_1 > 0$ such that $f(A) \le \kappa_1 \operatorname{tr}(A)$ for all $A \in S_d^+$.

Remark 20. Since f satisfies (A1) and (A2), the assumption (A3) holds if and only if there exists a constant $\kappa_0 > 0$ such that $f(A) \ge \kappa_0 \operatorname{tr}(A)$ for each $A \in S_d^+$. Indeed, assume that (A3) holds. If $A \in S_d^+$ and $A \ne 0$, then, in view of (A1), $f(A) = |A| f\big(\frac{A}{|A|}\big)$. Using this, (A0) and (A2), we deduce that there exists $\widetilde{E} \in S_d^+$ such that $|\widetilde{E}| = 1$ and $f(\widetilde{E}) = \min\{f(E) \mid E \in S_d^+,\ |E| = 1\} < +\infty$. Then, defining $\kappa_0 = f(\widetilde{E})/\sqrt{d}$, we have $f(A) \ge f(\widetilde{E})|A| \ge \kappa_0 \operatorname{tr}(A)$ for each $A \in S_d^+$. Clearly, the last inequality implies the coercivity of f.

Next, we prove that G is *narrowly* lower semicontinuous if (A3) holds. Recall that $(p_n)_{n\in\mathbb{N}} \subset \mathscr{P}(\mathbb{R}^d)$ narrowly converges to $p \in \mathscr{P}(\mathbb{R}^d)$ if $\int_{\mathbb{R}^d} \varphi \,\mathrm{d}p_n \to \int_{\mathbb{R}^d} \varphi \,\mathrm{d}p$ as $n \to +\infty$ for each $\varphi \in C_{\mathrm{b}}(\mathbb{R}^d)$.

Proposition 21. *Let* (A3) *hold, $x_n \to x \in \mathbb{R}^d$, $(p_n)_{n\in\mathbb{N}} \subset \mathscr{P}_2(\mathbb{R}^d)$ and p_n narrowly converges to $p \in \mathscr{P}_2(\mathbb{R}^d)$. Then*

$$G(x,p) \le \liminf_{n\to+\infty} G(x_n,p_n).$$

Proof. Without loss of generality, there exists a constant $C > 0$ (independent of n) such that for each $n \in \mathbb{N}$, $G(x_n, p_n) \le C$. Then, by Theorem 13, the infimum in (9), where $\mu = \delta_{x_n}$ and $\nu = p_n$, is actually a minimum. Thus, for each $n \in \mathbb{N}$, there exists $\lambda_n \in \mathscr{M}(\mathbb{R}^d, S_d^+)$ such that $|\lambda_n|(\mathbb{R}^d) < +\infty$, $\operatorname{tr}\big(\frac{1}{2}\nabla^2\lambda_n\big) = p_n - \delta_{x_n}$ and $G(x_n, p_n) = \int_{\mathbb{R}^d} f\big(\frac{\mathrm{d}\lambda_n}{\mathrm{d}|\lambda_n|}\big)\,\mathrm{d}|\lambda_n|$. Using this and (A3), we obtain

$$|\lambda_n|(\mathbb{R}^d) \le \eta_0 \int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\lambda_n}{\mathrm{d}|\lambda_n|}\right)\mathrm{d}|\lambda_n| \le \eta_0 C$$

for some constant $\eta_0 > 0$ independent of n (see Remark 20). Then, according to the Banach–Alaoglu theorem, there exists $\lambda \in \mathscr{M}(\mathbb{R}^d, S_d^+)$ such that $|\lambda|(\mathbb{R}^d) < +\infty$ and λ_n converges weakly to λ. Since, for each $\varphi \in C_c^2(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} \varphi\,\mathrm{d}p_n - \varphi(x_n) = \int_{\mathbb{R}^d} \frac{1}{2}\nabla^2\varphi : \mathrm{d}\lambda_n,$$

letting n tend to $+\infty$ and using the weak convergences, we deduce that

$$\int_{\mathbb{R}^d} \varphi\,\mathrm{d}p - \varphi(x) = \int_{\mathbb{R}^d} \frac{1}{2}\nabla^2\varphi : \mathrm{d}\lambda. \tag{27}$$

By direct adaptation of the density argument in [26, Remark 2.3], (27) implies that $\operatorname{tr}\big(\frac{1}{2}\nabla^2\lambda\big) = p - \delta_x$. Thus, λ is a competitor for $G(x,p) = F(\delta_x, p)$ (see (9)), which, in view of the lower semicontinuity of the function $\sigma \mapsto \int_{\mathbb{R}^d} f\big(\frac{\mathrm{d}\sigma}{\mathrm{d}|\sigma|}\big)\,\mathrm{d}|\sigma|$ on the subset of finite measures in $\mathscr{M}(\mathbb{R}^d, S_d^+)$, yields the estimate

$$G(x,p) \le \int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\lambda}{\mathrm{d}|\lambda|}\right)\mathrm{d}|\lambda| \le \liminf_{n\to+\infty} \int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\lambda_n}{\mathrm{d}|\lambda_n|}\right)\mathrm{d}|\lambda_n| = \liminf_{n\to+\infty} G(x_n, p_n)$$

and completes our proof of Proposition 21. □

Let $\mathscr{S}_{\mathrm{bb},2}(\mathbb{R}^d)$ be the set of all lower semicontinuous functions $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ such that φ is bounded from below and for some constant $C > 0$, $\big|\varphi(x)\big| \le C\big(1 + |x|^2\big)$ for each $x \in \mathbb{R}^d$.

Proposition 22. *Let* (A3) *hold and* $\varphi \in \mathscr{S}_{\mathrm{bb},2}(\mathbb{R}^d)$. *Then the infimum for* φ *in* (24) *is actually a minimum and* $\varphi^G \in \mathscr{S}_{\mathrm{bb},2}(\mathbb{R}^d)$.

Proof. Let $x \in \mathbb{R}^d$ and $(p_n)_{n\in\mathbb{N}} \subset \mathscr{P}_2(\mathbb{R}^d)$ be a minimizing sequence for $\varphi^G(x) \in \mathbb{R}$. Then there exists a constant $C > 0$ (independent of n) such that for each $n \in \mathbb{N}$ large enough, $G(x, p_n) \le C$. Using this, together with (A3) (see Remark 20) and Corollary 15, we deduce that

$$A := \sup_{n\in\mathbb{N}} \int_{\mathbb{R}^d} |y|^2\,\mathrm{d}p_n(y) < +\infty,$$

which, in view of [4, Remark 5.1.5], implies that $(p_n)_{n\in\mathbb{N}}$ is tight. Then, by Prokhorov's theorem (see [4, Theorem 5.1.3]), there exists a probability measure p on $\mathbb{R}^d$ such that, up to a subsequence (not relabeled), p_n converges narrowly to p. Thus,

$$\int_{\mathbb{R}^d} |y|^2\,\mathrm{d}p(y) \le \liminf_{n\to+\infty} \int_{\mathbb{R}^d} |y|^2\,\mathrm{d}p_n(y) \le A$$

and $p \in \mathscr{P}_2(\mathbb{R}^d)$. By the narrow convergence (recall that $\varphi \in \mathscr{S}_{\mathrm{bb},2}(\mathbb{R}^d)$) and Proposition 21,

$$\int_{\mathbb{R}^d} \varphi\,\mathrm{d}p + G(x,p) \le \liminf_{n\to+\infty} \int_{\mathbb{R}^d} \varphi\,\mathrm{d}p_n + G(x, p_n),$$

which says that p is a minimizer in the definition of $\varphi^G(x)$ (see (24)).

Next, we prove the lower semicontinuity of φ^G. Let $x_n \to x$, $p_n \in \mathscr{P}_2(\mathbb{R}^d)$ be a minimizer in the definition of $\varphi^G(x_n)$ and $\liminf_{n\to+\infty} \varphi^G(x_n) < +\infty$. Proceeding as before, we can assume that there exists $p \in \mathscr{P}_2(\mathbb{R}^d)$ such that, up to a subsequence (not relabeled), p_n converges narrowly to p. Using the narrow convergence, the fact that $\varphi \in \mathscr{S}_{\mathrm{bb},2}(\mathbb{R}^d)$ and Proposition 21, we obtain the following

$$\varphi^G(x) \le \int_{\mathbb{R}^d} \varphi \,\mathrm{d}p + G(x,p) \le \liminf_{n\to+\infty} \int_{\mathbb{R}^d} \varphi \,\mathrm{d}p_n + G(x_n,p_n) = \liminf_{n\to+\infty} \varphi^G(x_n),$$

which proves the lower semicontinuity of φ^G and completes our proof of Proposition 22. □

Under the assumption (A3), we can, on the one hand relax and, on the other hand, strengthen the dual constraint in (26) using bounded lower semicontinuous functions.

Proposition 23. *Let $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$ and (A3) hold. Then*

$$H(\mu,\nu) = \sup\left\{ \int_{\mathbb{R}^d} \varphi^G \,\mathrm{d}\mu - \int_{\mathbb{R}^d} \varphi \,\mathrm{d}\nu \;\middle|\; \varphi \in \mathscr{S}_{\mathrm{b}}(\mathbb{R}^d) \right\}. \tag{28}$$

Proof. Let $\varphi \in \Phi_{\mathrm{bb},2}(\mathbb{R}^d)$ and $\varphi_n = \min\{n,\varphi\}$ for each $n \in \mathbb{N}$. Then $\varphi_n \in C_{\mathrm{b}}(\mathbb{R}^d)$ and $\varphi_n \nearrow \varphi$ as $n \to +\infty$. According to Proposition 22, for each $n \in \mathbb{N}$ and for each $x \in \mathbb{R}^d$, there exists $p_n \in \mathscr{P}_2(\mathbb{R}^d)$ such that $\varphi_n^G(x) = \int_{\mathbb{R}^d} \varphi_n \,\mathrm{d}p_n + G(x,p_n)$. Since $\varphi_n^G(x) \le \varphi(x) < +\infty$, arguing by the same way as in the proof of Proposition 22, we deduce that there exists $p \in \mathscr{P}_2(\mathbb{R}^d)$ such that, up to a subsequence (not relabeled), p_n converges narrowly to p. For each $k \in \mathbb{N}$, using the weak convergence and Proposition 21, we obtain

$$\int_{\mathbb{R}^d} \varphi_k \,\mathrm{d}p + G(x,p) \le \liminf_{n\to+\infty} \int_{\mathbb{R}^d} \varphi_k \,\mathrm{d}p_n + G(x,p_n) \le \liminf_{n\to+\infty} \int_{\mathbb{R}^d} \varphi_n \,\mathrm{d}p_n + G(x,p_n) = \liminf_{n\to+\infty} \varphi_n^G(x).$$

Letting k tend to $+\infty$, by the monotone convergence theorem, we have

$$\varphi^G(x) \le \int_{\mathbb{R}^d} \varphi \,\mathrm{d}p + G(x,p) = \lim_{k\to+\infty} \int_{\mathbb{R}^d} \varphi_k \,\mathrm{d}p + G(x,p) \le \liminf_{n\to+\infty} \varphi_n^G(x).$$

On the other hand, $\varphi_n^G(x) \le \varphi^G(x)$ for each $n \in \mathbb{N}$ and hence $\varphi_n^G(x) \nearrow \varphi^G(x)$ as $n \to +\infty$. Thus, by the monotone convergence theorem, $\int_{\mathbb{R}^d} \varphi_n^G \,\mathrm{d}\mu \to \int_{\mathbb{R}^d} \varphi^G \,\mathrm{d}\mu$ and $\int_{\mathbb{R}^d} \varphi_n \,\mathrm{d}\nu \to \int_{\mathbb{R}^d} \varphi \,\mathrm{d}\nu$ as $n \to +\infty$. This, together with (26), implies that

$$H(\mu,\nu) = \sup\left\{ \int_{\mathbb{R}^d} \varphi^G \,\mathrm{d}\mu - \int_{\mathbb{R}^d} \varphi \,\mathrm{d}\nu \;\middle|\; \varphi \in C_{\mathrm{b}}(\mathbb{R}^d) \right\}.$$

Since for each $\varphi \in \mathscr{S}_{\mathrm{b}}(\mathbb{R}^d)$, there exists a sequence $(\varphi_n)_{n\in\mathbb{N}} \subset C_{\mathrm{b}}(\mathbb{R}^d)$ such that $\varphi_n \nearrow \varphi$ as $n \to +\infty$, repeating the above procedure, we complete our proof of Proposition 23. □

If the G-transform $\varphi \mapsto \varphi^G$ is idempotent on $\mathscr{S}_{\mathrm{b}}(\mathbb{R}^d)$, the following dual formulation for $H(\mu,\nu)$ holds.

Proposition 24. *Let $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$, (A3) hold and $\varphi^{GG} = \varphi^G$ for each $\varphi \in \mathscr{S}_{\mathrm{b}}(\mathbb{R}^d)$. Then*

$$H(\mu,\nu) = \sup\left\{ \langle \nu - \mu, \psi \rangle \;\middle|\; \psi \in \mathscr{U}_{\mathrm{b}}(\mathbb{R}^d),\ -\psi = (-\psi)^G \right\}. \tag{29}$$

Proof. In view of Proposition 23, $H(\mu,\nu)$ is greater than or equal to the supremum in (29). On the other hand, using the estimate $\varphi^G \le \varphi$ in (28), we deduce the following

$$H(\mu,\nu) \le \sup\left\{ \langle \mu - \nu, \varphi^G \rangle \;\middle|\; \varphi \in \mathscr{S}_{\mathrm{b}}(\mathbb{R}^d) \right\} \le \sup\left\{ \langle \mu - \nu, \varphi \rangle \;\middle|\; \varphi \in \mathscr{S}_{\mathrm{b}}(\mathbb{R}^d),\ \varphi = \varphi^G \right\},$$

where the latter estimate comes from the assumption that $\varphi^{GG} = \varphi^G$ for each $\varphi \in \mathscr{S}_{\mathrm{b}}(\mathbb{R}^d)$, since in this case, taking into account Proposition 22, we have $\{\varphi^G \mid \varphi \in \mathscr{S}_{\mathrm{b}}(\mathbb{R}^d)\} \subset \{\varphi \mid \varphi \in \mathscr{S}_{\mathrm{b}}(\mathbb{R}^d),\ \varphi = \varphi^G\}$. Thus, the supremum in (29) is greater than or equal to $H(\mu,\nu)$ and the dual formulation (29) holds, which completes our proof of Proposition 24. □

The following proposition, which is a generalization of [2, Proposition 6.4], describes some situations in which the G-transform is idempotent on $\mathscr{S}_{\mathrm{b}}(\mathbb{R}^d)$, which in particular happens when H is subadditive.

Proposition 25. *Let* (A3) *hold. Then the following assertions are equivalent:*

(i) *for each choice of* $\mu, \nu, p \in \mathscr{P}_2(\mathbb{R}^d)$*,* $H(\mu,\nu) \le H(\mu,p) + H(p,\nu)$*;*

(ii) *for each* $\nu, p \in \mathscr{P}_2(\mathbb{R}^d)$ *and analytically measurable probability kernel* $y \in \mathbb{R}^d \mapsto \gamma^y \in \mathscr{P}_2(\mathbb{R}^d)$*,*

$$G(x,\nu) \le G(x,p) + \int_{\mathbb{R}^d} G(y,\gamma^y)\,\mathrm{d}p(y) \quad \textit{whenever}\ \nu = \int_{\mathbb{R}^d} \gamma^y\,\mathrm{d}p(y); \tag{30}$$

(iii) *for each* $\varphi \in \mathscr{S}_{\mathrm{b}}(\mathbb{R}^d)$*,* $\varphi^{GG} = \varphi^G$*.*

Proof. The proof of the implication (i) ⇒ (ii) follows by choosing $\mu = \delta_x$, $\nu = \int_{\mathbb{R}^d} \gamma^y\,\mathrm{d}p(y)$ and using the definition of $H(p,\nu)$ as the infimum (see (23)).

Now we prove the implication (ii) ⇒ (iii). For each $\varphi \in \mathscr{S}_{\mathrm{b}}(\mathbb{R}^d)$, in view of Proposition 22, which applies in particular to every function lying in $\mathscr{S}_{\mathrm{b}}(\mathbb{R}^d) \subset \mathscr{S}_{\mathrm{bb},2}(\mathbb{R}^d)$, and inasmuch as $\varphi^G \le \varphi$ (see (25)), we have $\varphi^G \in \mathscr{S}_{\mathrm{b}}(\mathbb{R}^d)$. Repeating this observation for $\varphi^G \in \mathscr{S}_{\mathrm{b}}(\mathbb{R}^d)$, one can see that $\varphi^{GG} \in \mathscr{S}_{\mathrm{b}}(\mathbb{R}^d)$ and $\varphi^{GG} \le \varphi^G$. Thus, it is enough to prove the estimate

$$\varphi^G(x) \le \int_{\mathbb{R}^d} \varphi^G(y)\,\mathrm{d}p(y) + G(x,p) \tag{31}$$

for each $x \in \mathbb{R}^d$ and $p \in \mathscr{P}_2(\mathbb{R}^d)$. Using the lower semicontinuity of G and the fact that $\varphi \in \mathscr{S}_{\mathrm{b}}(\mathbb{R}^d)$, we deduce that the function $(x,p) \mapsto \int_{\mathbb{R}^d} \varphi\,\mathrm{d}p + G(x,p)$ is lower semicontinuous on $\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d)$, where the topology on $\mathbb{R}^d$ is generated by the Euclidean distance and the topology on $\mathscr{P}_2(\mathbb{R}^d)$ is generated by the 2-Wasserstein distance. Then, according to [9, Proposition 7.50], for each $\varepsilon > 0$ there exists an analytically measurable probability kernel $y \in \mathbb{R}^d \mapsto \gamma^y \in \mathscr{P}_2(\mathbb{R}^d)$ such that

$$\varphi^G(y) + \varepsilon \ge \int_{\mathbb{R}^d} \varphi(z)\,\mathrm{d}\gamma^y(z) + G(y,\gamma^y).$$

Then defining $\nu(\mathrm{d}z) = \int_{\mathbb{R}^d} \gamma^y(\mathrm{d}z)\,\mathrm{d}p(y)$, integrating both sides of the above inequality with respect to $p \in \mathscr{P}_2(\mathbb{R}^d)$ and using (ii), we obtain

$$\begin{aligned} G(x,p) + \int_{\mathbb{R}^d} \varphi^G(y)\,\mathrm{d}p(y) + \varepsilon &\ge G(x,p) + \int_{\mathbb{R}^d} G(y,\gamma^y)\,\mathrm{d}p(y) + \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \varphi(z)\,\mathrm{d}\gamma^y(z)\,\mathrm{d}p(y) \\ &\ge G(x,\nu) + \int_{\mathbb{R}^d} \varphi(z)\,\mathrm{d}\nu(z) \\ &\ge \varphi^G(x), \end{aligned}$$

which yields (31) and completes our proof of the implication (ii) ⇒ (iii).

The implication (iii) ⇒ (i) is a direct consequence of Proposition 24, since for each $\psi \in \mathscr{U}_{\mathrm{b}}(\mathbb{R}^d)$ such that $-\psi = (-\psi)^G$,

$$\langle \nu - \mu, \psi\rangle \le \langle p - \mu, \psi\rangle + \langle \nu - p, \psi\rangle \le H(\mu,p) + H(p,\nu).$$

This completes our proof of Proposition 25. □

In view of the subadditivity of the function f (see (A1)) and under the assumption (A3), the functional H is subadditive, and hence the G-transform is idempotent on $\mathscr{S}_{\mathrm{b}}(\mathbb{R}^d)$.

Theorem 26. *Let* $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ *and* (A3) *hold. Then*

$$H(\mu,\nu) = \sup\big\{\langle \nu - \mu, \psi\rangle \,\big|\, \psi \in \mathscr{U}_{\mathrm{b}}(\mathbb{R}^d),\ -\psi = (-\psi)^G\big\}.$$

Proof. The proof follows from Proposition 24 as soon as $\varphi^{GG} = \varphi^G$ for each $\varphi \in \mathscr{S}_{\mathrm{b}}(\mathbb{R}^d)$. This, in view of Proposition 25, holds if for each $p \in \mathscr{P}_2(\mathbb{R}^d)$, $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ and analytically measurable probability kernel $y \in \mathbb{R}^d \mapsto \gamma^y \in \mathscr{P}_2(\mathbb{R}^d)$,

$$G(x,\nu) \le G(x,p) + \int_{\mathbb{R}^d} G(y,\gamma^y)\,\mathrm{d}p(y) \quad \text{whenever}\ \nu = \int_{\mathbb{R}^d} \gamma^y\,\mathrm{d}p(y). \tag{32}$$

Using the definition of H (see (23)) and Proposition 17, we have

$$\int_{\mathbb{R}^d} G(y,\gamma^y)\,\mathrm{d}p(y) \geq H(p,\nu) \geq F(p,\nu). \tag{33}$$

Combining Proposition 16(ii) and (33), yields (32), namely

$$G(x,\nu) \leq G(x,p) + F(p,\nu) \leq G(x,p) + \int_{\mathbb{R}^d} G(y,\gamma^y)\,\mathrm{d}p(y)$$

(see (22)). This completes our proof of Theorem 26. □

Remark 27. Let $E \subset \mathbb{R}^d$. Then in view of (22) and (24),

$$\begin{aligned}
&(-\psi)^G(x) = -\psi(x) \quad \forall\, x \in E \\
&\iff \int_{\mathbb{R}^d} -\psi\,\mathrm{d}p + G(x,p) \geq -\psi(x) \quad \forall\,(x,p) \in E \times \mathscr{P}_2(\mathbb{R}^d) \\
&\iff \int_{\mathbb{R}^d} -\psi\,\mathrm{d}p + \int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\lambda}{\mathrm{d}|\lambda|}\right)\mathrm{d}|\lambda| \geq -\psi(x) \\
&\qquad\qquad \forall\,(x,p,\lambda) \in E \times \mathscr{P}_2(\mathbb{R}^d) \times \mathscr{M}(\mathbb{R}^d, S_d^+),\ \mathrm{tr}\left(\frac{1}{2}\nabla^2\lambda\right) = p - \delta_x.
\end{aligned}$$

Assuming (A4) instead of (A3), we also obtain the dual formulation for H, where the dual competitors are invariant under the G-transform but belong to $\Phi_{\mathrm{bb},2}(\mathbb{R}^d)$ (see Theorem 32). We first prove the following key result.

Proposition 28. *Let* (A4) *hold. Then for each* $p \in \mathscr{P}_2(\mathbb{R}^d)$,

$$G\big([p],p\big) \leq \kappa_1 \operatorname{var}(p). \tag{34}$$

Furthermore, if $f = \mathrm{tr}$ *on* S_d^+, *then for each* $p \in \mathscr{P}_2(\mathbb{R}^d)$,

$$G\big([p],p\big) = \operatorname{var}(p). \tag{35}$$

Proof. In view of Proposition 9 and (A4),

$$G\big([p],p\big) \leq \kappa_1 \inf\left\{\int_{\mathbb{R}^d} \mathrm{tr}\left(\frac{\mathrm{d}\lambda}{\mathrm{d}|\lambda|}\right)\mathrm{d}|\lambda| \;\middle|\; \mathrm{tr}\left(\frac{1}{2}\nabla^2\lambda\right) = p - \delta_{[p]}\right\}.$$

Thus, it suffices to prove (35). Assume that $f = \mathrm{tr}$ on S_d^+. By Jensen's inequality, for each concave function $u\colon \mathbb{R}^d \to \mathbb{R}$,

$$\langle p - \delta_{[p]}, u\rangle \leq 0. \tag{36}$$

Since for each $A \in S_d$, $-\mathrm{tr}^*(A) = 0 \Leftrightarrow A - I_d \leq 0$ (see (67)), according to Theorem 13,

$$\begin{aligned}
G\big([p],p\big) &= \sup\left\{\langle p - \delta_{[p]}, \psi\rangle \;\middle|\; \psi \in C_{\mathrm{b}}^2(\mathbb{R}^d),\ \left(\frac{1}{2}\nabla^2\psi - I_d\right) \leq 0 \text{ on } \mathbb{R}^d\right\} \\
&= \big\langle p - \delta_{[p]}, |\cdot|^2\big\rangle + \sup\big\{\langle p - \delta_{[p]}, u\rangle \;\big|\; u \in C_{\mathrm{b}}^2(\mathbb{R}^d),\ u \text{ is concave on } \mathbb{R}^d\big\} \\
&= \operatorname{var}(p),
\end{aligned}$$

where the last equality comes from (36). This completes our proof of Proposition 28. □

Corollary 29. *Let* (A4) *hold. Then for each lower semianalytic function* $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ *bounded from below such that* $\varphi = \varphi^G$, *the function* $\varphi(\cdot) + \kappa_1|\cdot|^2$ *is convex and locally Lipschitz on* $\mathbb{R}^d$.

Proof of Corollary 29. Let $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ be lower semianalytic, bounded from below and $\varphi = \varphi^G$. According to Remark 27, for each $p \in \mathscr{P}_2(\mathbb{R}^d)$,

$$\varphi\big([p]\big) \leq \int_{\mathbb{R}^d} \varphi\,\mathrm{d}p + G\big([p],p\big) \leq \int_{\mathbb{R}^d} \varphi\,\mathrm{d}p + \kappa_1\left(\int_{\mathbb{R}^d} |y|^2\,\mathrm{d}p(y) - \big|[p]\big|^2\right),$$

where the latter estimate comes from the estimate (34) of Proposition 28. Thus,

$$\varphi\big([p]\big)+\kappa_1\big|[p]\big|^2 \le \int_{\mathbb{R}^d}\big(\varphi(y)+\kappa_1|y|^2\big)\,\mathrm{d}p(y). \tag{37}$$

Since

$$\frac{x+y}{2}=\left[\frac{\delta_x}{2}+\frac{\delta_y}{2}\right] \quad \forall\, x,y\in\mathbb{R}^d,$$

(37) yields

$$\varphi\Big(\frac{x+y}{2}\Big)+\kappa_1\Big|\frac{x+y}{2}\Big|^2 \le \frac{\varphi(x)+\varphi(y)}{2}+\frac{\kappa_1\big(|x|^2+|y|^2\big)}{2},$$

which, together with the local boundedness of φ, implies the convexity of $\varphi(\cdot)+\kappa_1|\cdot|^2$ on $\mathbb{R}^d$. For the fact that a convex function is locally Lipschitz on the interior of its proper domain, the reader may consult [23, Theorem 10.4]. This completes our proof of Corollary 29. □

Proposition 30. *Let $\mu,\nu\in\mathscr{P}_2(\mathbb{R}^d)$,* (A4) *hold and $\varphi^{GG}=\varphi^G$ for each $\varphi\in\Phi_{\mathrm{bb},2}(\mathbb{R}^d)$. Then*

$$H(\mu,\nu)=\sup\big\{\langle\nu-\mu,\psi\rangle \,\big|\, -\psi\in\Phi_{\mathrm{bb},2}(\mathbb{R}^d),\ -\psi=(-\psi)^G\big\}. \tag{38}$$

Proof. By Proposition 19(ii), $H(\mu,\nu)$ is greater than or equal to the supremum in (38). In view of Remark 18 and Corollary 29, for each $\varphi\in\Phi_{\mathrm{bb},2}(\mathbb{R}^d)$, the functions φ^G and φ^{GG} are well defined and $\varphi^G\in\Phi_{\mathrm{bb},2}(\mathbb{R}^d)$ whenever $\varphi^{GG}=\varphi^G$. Then $\big\{\varphi^G \,\big|\, \varphi\in\Phi_{\mathrm{bb},2}(\mathbb{R}^d)\big\}\subset\big\{\varphi \,\big|\, \varphi\in\Phi_{\mathrm{bb},2}(\mathbb{R}^d),\ \varphi=\varphi^G\big\}$, since $\varphi^{GG}=\varphi^G$ for each $\varphi\in\Phi_{\mathrm{bb},2}(\mathbb{R}^d)$ by our assumption. Using this and the estimate

$$\int_{\mathbb{R}^d}\varphi^G\,\mathrm{d}\mu-\int_{\mathbb{R}^d}\varphi\,\mathrm{d}\nu\le\int_{\mathbb{R}^d}\varphi^G\,\mathrm{d}\mu-\int_{\mathbb{R}^d}\varphi^G\,\mathrm{d}\nu$$

for each $\varphi\in\Phi_{\mathrm{bb},2}(\mathbb{R}^d)$ in (26), we deduce that $H(\mu,\nu)$ is less than or equal to the supremum in (38). This completes our proof of Proposition 30. □

The next proposition is a counterpart of Proposition 25, where we replace the assumption (A3) by (A4) and describe some situations in which the G-transform is idempotent on $\Phi_{\mathrm{bb},2}(\mathbb{R}^d)$.

Proposition 31. *Let* (A4) *hold. Then the following assertions are equivalent:*

(i) *For each choice of $\mu,\nu,p\in\mathscr{P}_2(\mathbb{R}^d)$, $H(\mu,\nu)\le H(\mu,p)+H(p,\nu)$.*

(ii) *For each $\nu,p\in\mathscr{P}_2(\mathbb{R}^d)$ and analytically measurable probability kernel $y\in\mathbb{R}^d\mapsto\gamma^y\in\mathscr{P}_2(\mathbb{R}^d)$,*

$$G(x,\nu)\le G(x,p)+\int_{\mathbb{R}^d}G(y,\gamma^y)\,\mathrm{d}p(y) \quad \textit{whenever}\ \nu=\int_{\mathbb{R}^d}\gamma^y\,\mathrm{d}p(y). \tag{39}$$

(iii) *For each $\varphi\in\Phi_{\mathrm{bb},2}(\mathbb{R}^d)$, $\varphi^{GG}=\varphi^G$.*

Proof. The proof follows by reproducing the arguments of the proof of Proposition 25 with minor modifications, in particular, using Proposition 30 in the proof of the implication (iii) ⇒ (i). □

Theorem 32. *Let $\mu,\nu\in\mathscr{P}_2(\mathbb{R}^d)$ and* (A4) *hold. Then*

$$H(\mu,\nu)=\sup\big\{\langle\nu-\mu,\psi\rangle \,\big|\, -\psi\in\Phi_{\mathrm{bb},2}(\mathbb{R}^d),\ -\psi=(-\psi)^G\big\}.$$

Proof. Proceeding in a similar manner to the proof of Theorem 26, we deduce that the assertions (i)–(iii) of Proposition 31 hold. Then, applying Proposition 30, we complete our proof of Theorem 32. □

If (A3) and (A4) hold simultaneously, we have the following dual formulation.

Theorem 33. *Let $\mu,\nu\in\mathscr{P}_2(\mathbb{R}^d)$ and* (A3), (A4) *hold. Then*

$$H(\mu,\nu)=\sup\big\{\langle\nu-\mu,\psi\rangle \,\big|\, \psi\in C_{\mathrm{b}}(\mathbb{R}^d),\ -\psi=(-\psi)^G\big\}. \tag{40}$$

Proof. The proof follows from Theorem 26 and Corollary 29. □

3.4. *Duality for F in terms of invariant functions under G-transform*

The following theorem refines the dual formulation for $F(\mu,\nu)$ obtained in Theorem 13 in terms of invariant functions under G-transform $\varphi \mapsto \varphi^G$ (see (24)) in line with [2,17].

Theorem 34. *For each $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$, we have*

$$F(\mu,\nu) = \sup\left\{\langle \nu-\mu,\psi\rangle \;\middle|\; \psi \in C^2_{\mathrm{b}}(\mathbb{R}^d),\ -\psi = (-\psi)^G\right\}. \tag{41}$$

Remark 35. The connection between the formulae (40) and (41) is striking, with the only difference being that the competitors for $F(\mu,\nu)$ are additionally twice continuously differentiable with respect to the competitors for $H(\mu,\nu)$. Moreover, Theorem 50 indicates the equality of these formulae.

The proof of Theorem 34 is a direct consequence of Theorem 13 and the next result.

Proposition 36. *Let $\psi \in C^2_{\mathrm{b}}(\mathbb{R}^d)$. Then $-f^*\left(\frac{1}{2}\nabla^2\psi\right) = 0$ in $\mathbb{R}^d$ if and only if $-\psi = (-\psi)^G$.*

Proof. According to Theorem 13, for each $x \in \mathbb{R}^d$, $p \in \mathscr{P}_2(\mathbb{R}^d)$ and $\psi \in C^2_{\mathrm{b}}(\mathbb{R}^d)$ such that $-f^*\left(\frac{1}{2}\nabla^2\psi\right) = 0$ in $\mathbb{R}^d$,

$$G(x,p) \geq \int_{\mathbb{R}^d} \psi \,\mathrm{d}p - \psi(x)$$

and hence

$$(-\psi)^G(x) = \inf\left\{\int_{\mathbb{R}^d} -\psi\,\mathrm{d}p + G(x,p) \;\middle|\; p \in \mathscr{P}_2(\mathbb{R}^d)\right\} \geq -\psi(x).$$

Thus, $-\psi = (-\psi)^G$ (see (25)).

Let us now assume that $\psi \in C^2_{\mathrm{b}}(\mathbb{R}^d)$ and $-\psi = (-\psi)^G$. Fix arbitrary $A = (a_{ij})_{i,j=1}^d \in S_d^{++}$ and $x_0 \in \mathbb{R}^d$. Let $g_A(x,y)$ be the Green function of the elliptic operator $L = -\sum_{i,j=1}^d a_{ij}\partial_{ij} = -\operatorname{div}(A\nabla)$ on $B_r(x_0)$, namely, for each $y \in B_r(x_0)$, $g_A(\cdot,y) \in W_0^{1,p}\big(B_r(x_0)\big)$ whenever $p < d/(d-1)$ and

$$-A : \nabla^2 g_A = \delta_y \quad \text{in } \mathscr{D}'\big(B_r(x_0)\big),$$

which means that

$$-\int_{B_r(x_0)} A : \nabla^2\varphi(x) g_A(x,y)\,\mathrm{d}x = \varphi(y) \quad \forall\, \varphi \in C^2_{\mathrm{c}}\big(B_r(x_0)\big)$$

(see, for instance, [20]). Since $A \in S_d^{++}$, $A = P\operatorname{diag}(\lambda_1,\dots,\lambda_d)P^{\mathrm{T}}$ for some orthogonal real $d\times d$ matrix P and positive numbers $\lambda_i > 0$. Also $A^{-1} \in S_d^{++}$ and there exists the unique matrix $B \in S_d^{++}$ such that $A = B^{-2}$, namely $B = P\operatorname{diag}\big(1/\sqrt{\lambda_1},\dots,1/\sqrt{\lambda_d}\big)P^{\mathrm{T}}$. Fix an arbitrary $v \in C^2_{\mathrm{b}}(\mathbb{R}^d)$ and define $u(\cdot) = v(B^{-1}\cdot)$ so that $u \in C^2_{\mathrm{b}}(\mathbb{R}^d)$. Then $A : \nabla^2 v(x) = \Delta u(Bx)$ for each $x \in \mathbb{R}^d$. If g is the Green function of the Laplace operator on $BB_r(x_0) = \big\{Bx \;\big|\; x \in B_r(x_0)\big\}$, then $g_A(x,y) = \det(B)g(Bx,By)$. To lighten the notation, define $U = BB_r(x_0)$. Next, using the Green representation formula, changing the variables and using that $A : \nabla^2 v(x) = \Delta u(Bx)$ and $g_A(x,x_0) = \det(B)g(Bx,Bx_0)$, we have

$$\begin{aligned}
v(x_0) &= u(Bx_0)\\
&= -\int_U \Delta u(x) g(x,Bx_0)\,\mathrm{d}x - \int_{\partial U} u(x)\nabla g(x,Bx_0)\cdot \nu_{\partial U}(x)\,\mathrm{d}\mathscr{H}^{d-1}(x)\\
&= -\det(B)\int_{B_r(x_0)} \Delta u(Bx) g(Bx,Bx_0)\,\mathrm{d}x\\
&\qquad - \det(B)\int_{\partial B_r(x_0)} u(Bx)\nabla g(Bx,Bx_0)\cdot\nu_{\partial U}(Bx)\big|B^{-\mathrm{T}}\nu(x)\big|\,\mathrm{d}\mathscr{H}^{d-1}(x)\\
&= -\int_{B_r(x_0)} A : \nabla^2 v(x) g_A(x,x_0)\,\mathrm{d}x - \int_{\partial B_r(x_0)} v(x) A\nabla g_A(x,x_0)\cdot\nu(x)\,\mathrm{d}\mathscr{H}^{d-1}(x),
\end{aligned} \tag{42}$$

where $\nu_{\partial U}$ and ν denote the outward pointing unit normal vector fields along ∂U and $\partial B_r(x_0)$, respectively.

By Hopf's lemma (see [18,21]), $-A\nabla g_A(x,x_0)\cdot\nu(x) > 0$ for each $x \in \partial B_r(x_0)$. Then, using (42) with $v = 1$, we obtain $p = -A\nabla g_A(\cdot,x_0)\cdot\nu(\cdot)\mathscr{H}^{d-1} \llcorner \partial B_r(x_0) \in \mathscr{P}_2(\mathbb{R}^d)$.

Altogether, due to (42) and the fact that $v \in C^2_{\mathrm{b}}(\mathbb{R}^d)$ was arbitrarily chosen, we have

$$\operatorname{tr}\left(\frac{1}{2}\nabla^2\lambda\right) = p - \delta_{x_0}, \tag{43}$$

where $\lambda = 2Ag_A(\cdot,x_0)\mathscr{L}^d \llcorner B_r(x_0) \in \mathscr{M}(\mathbb{R}^d, S^+_d)$ and $|\lambda|(\mathbb{R}^d) < +\infty$. Since $-\psi(x_0) = (-\psi)^G(x_0)$, according to Remark 27,

$$\int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\lambda}{\mathrm{d}|\lambda|}\right)\mathrm{d}|\lambda| \geq \int_{\mathbb{R}^d}\psi\,\mathrm{d}p - \psi(x_0).$$

This, together with the positive 1-homogeneity of f (see (A1)), (43) and the fact that $\psi \in C^2_{\mathrm{b}}(\mathbb{R}^d)$, implies that

$$\int_{B_r(x_0)} f(A)g_A(x,x_0)\,\mathrm{d}x \geq \int_{B_r(x_0)} A : \frac{1}{2}\nabla^2\psi(x)g_A(x,x_0)\,\mathrm{d}x. \tag{44}$$

Assume by contradiction that $A : \frac{1}{2}\nabla^2\psi(x_0) - f(A) > 0$. Then there exist $\varepsilon, r > 0$ such that for each $x \in B_r(x_0)$, $A : \frac{1}{2}\nabla^2\psi(x) - f(A) \geq \varepsilon$. But this contradicts (44), since $\int_{B_r(x_0)} g_A(x,x_0)\,\mathrm{d}x > 0$. Thus, $A : \frac{1}{2}\nabla^2\psi(x_0) - f(A) \leq 0$, which implies that $\frac{1}{2}\nabla^2\psi(x_0) \in \operatorname{dom}(f^*)$, because $A \in S^{++}_d$ was arbitrarily chosen and $S^{++}_d \cap \operatorname{dom}(f)$ is dense in $\operatorname{dom}(f)$ (see (A0)). Since $x_0 \in \mathbb{R}^d$ was arbitrary, $\frac{1}{2}\nabla^2\psi(\mathbb{R}^d) \subset \operatorname{dom}(f^*)$, which holds if and only if $-f^*\left(\frac{1}{2}\nabla^2\psi\right) = 0$ in $\mathbb{R}^d$. This completes our proof of Proposition 36. □

4. *F* versus *H*

4.1. *F* = *H* under approximation assumptions

Our first type of approximation assumption is related to the assumption (A3).

Theorem 37. *Let $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$ and* (A3) *hold. Assume that for each $\psi \in \mathscr{U}_{\mathrm{b}}(\mathbb{R}^d)$ such that $-\psi = (-\psi)^G$ there exists $(\psi_n)_{n\in\mathbb{N}} \subset C^2_{\mathrm{b}}(\mathbb{R}^d)$ such that $-\psi_n = (-\psi_n)^G$ for each $n \in \mathbb{N}$ and $\langle \nu-\mu,\psi_n\rangle \to \langle \nu-\mu,\psi\rangle$ as $n \to +\infty$. Then $F(\mu,\nu) = H(\mu,\nu)$.*

Proof. By Proposition 17, $F(\mu,\nu) \leq H(\mu,\nu)$. Next, using the assumption of Theorem 37, together with Theorems 26 and 34, we have $H(\mu,\nu) \leq F(\mu,\nu)$. This completes our proof of Theorem 37. □

Example (A). For each $A \in S_d$,

$$f(A) = \begin{cases} t & \text{if } A = tI_d \text{ for some } t \geq 0, \\ +\infty & \text{otherwise.} \end{cases} \tag{45}$$

Clearly, f satisfies (A0)–(A3). Then for each $A \in S_d$,

$$f^*(A) = \sup\{A : M - f(M) \mid M \in \operatorname{dom}(f)\} = \begin{cases} 0 & \text{if } \operatorname{tr}(A) \leq 1, \\ +\infty & \text{otherwise.} \end{cases} \tag{46}$$

Given $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$, using Theorem 13 and (46), we have

$$\begin{aligned} F(\mu,\nu) &= \sup\left\{\langle \nu-\mu,\psi\rangle \,\middle|\, \psi \in C^2_{\mathrm{b}}(\mathbb{R}^d),\ \operatorname{tr}\left(\frac{1}{2}\nabla^2\psi(x)\right) \leq 1 \ \ \forall\, x \in \mathbb{R}^d\right\} \\ &= \sup\left\{\langle \nu-\mu,\psi\rangle \,\middle|\, \psi \in C^2_{\mathrm{b}}(\mathbb{R}^d),\ \Delta\left[\frac{1}{2}\psi(x) - \frac{|x|^2}{2d}\right] \leq 0 \ \ \forall\, x \in \mathbb{R}^d\right\} \\ &= \left\langle \nu-\mu, \frac{1}{d}|\cdot|^2\right\rangle + \sup\{\langle \nu-\mu,\varphi\rangle \mid \varphi \in C^2_{\mathrm{b}}(\mathbb{R}^d),\ \Delta\varphi(x) \leq 0 \ \ \forall\, x \in \mathbb{R}^d\} \end{aligned}$$

$$= \begin{cases} \frac{1}{d}\operatorname{var}(\nu) - \frac{1}{d}\operatorname{var}(\mu) & \text{if } \mu \leq_{\text{sh}} \nu, \\ +\infty & \text{otherwise} \end{cases} \tag{47}$$

$$= \begin{cases} f\big(\int_{\mathbb{R}^d} x \otimes x \,\mathrm{d}\big(\nu(x) - \mu(x)\big)\big) & \text{if } \mu \leq_{\text{sh}} \nu, \\ +\infty & \text{otherwise,} \end{cases} \tag{48}$$

where to obtain (47) we have used the fact that if $\mu \leq_{\text{sh}} \nu$ (see Definition 5), then $[\mu] = [\nu]$ and hence $\langle \nu - \mu, \frac{1}{d}|\cdot|^2\rangle = \frac{1}{d}\operatorname{var}(\nu) - \frac{1}{d}\operatorname{var}(\mu)$. To obtain (48) we have used the following. Notice that a competitor λ for (9) such that $\int_{\mathbb{R}^d} f\big(\frac{\mathrm{d}\lambda}{\mathrm{d}|\lambda|}\big)\,\mathrm{d}|\lambda| < +\infty$ with f defined by (45) has the form $\lambda = u\, I_d\, m$, where m is a nonnegative measure on $\mathbb{R}^d$, $u \geq 0$ m-a.e. on $\mathbb{R}^d$ and $u \in L^1(\mathbb{R}^d, \mathrm{d}m)$. Furthermore, such a measure λ solves $\operatorname{tr}\big(\frac{1}{2}\nabla^2\lambda\big) = \nu - \mu$ if and only if $\Delta\big(\frac{1}{2}um\big) = \nu - \mu$ in the weak sense (see (8)). Thus, for each $i, j \in \{1, \dots, d\}$ such that $i \neq j$, we have

$$\int_{\mathbb{R}^d} (x_i^2 - x_j^2)\,\mathrm{d}\nu(x) - \int_{\mathbb{R}^d} (x_i^2 - x_j^2)\,\mathrm{d}\mu(x) = \int_{\mathbb{R}^d} \frac{1}{2}\Delta[x_i^2 - x_j^2]\,u(x)\,\mathrm{d}m(x) = 0 \tag{49}$$

and

$$\int_{\mathbb{R}^d} x_i x_j\,\mathrm{d}\nu(x) - \int_{\mathbb{R}^d} x_i x_j\,\mathrm{d}\mu(x) = \int_{\mathbb{R}^d} \frac{1}{2}\Delta[x_i x_j]\,u(x)\,\mathrm{d}m(x) = 0. \tag{50}$$

If $\mu \leq_{\text{sh}} \nu$, by (47), $F(\mu,\nu) < +\infty$. Then, (49) and (50) imply that $\int_{\mathbb{R}^d} x \otimes x\,\mathrm{d}\big(\nu(x) - \mu(x)\big) \in S_d^+$ is a diagonal matrix equal to $\int_{\mathbb{R}^d} \frac{1}{d}|x|^2\,\mathrm{d}\big(\nu(x) - \mu(x)\big) I_d$ and hence

$$f\left(\int_{\mathbb{R}^d} x \otimes x\,\mathrm{d}\big(\nu(x) - \mu(x)\big)\right) = \int_{\mathbb{R}^d} \frac{1}{d}|x|^2\,\mathrm{d}\big(\nu(x) - \mu(x)\big) = \frac{1}{d}\operatorname{var}(\nu) - \frac{1}{d}\operatorname{var}(\mu).$$

Thus, for each $a \in \mathbb{R}^d$, for each $\xi \in \mathbb{R}^d$ and for each $A \in S_d$ such that $\operatorname{tr}(A) = 0$, the function

$$\psi(x) = a + x \cdot \xi + \left(A + \frac{I_d}{d}\right) : x \otimes x$$

is a dual optimizer for $F(\mu,\nu)$, which is understood in the setting of (16), when $F(\mu,\nu) < +\infty$ and f is defined by (45).

Notice that the equality $F(\mu,\nu) = H(\mu,\nu)$ when f is defined by (45) can be proved based on the results from [14], as well as using Theorem 37. Below we present both approaches allowing the reader to verify the assumptions of Theorem 37 in the context of the example (A).

- By (47), if $F(\mu,\nu) < +\infty$, then μ and ν are in subharmonic order. It follows by [14, Theorem 1.5 and Remark 1.7] that in this case there exists at least one transport plan $\gamma \in \Pi(\mu,\nu)$ such that $\gamma = \gamma^x \otimes \mu$ with $\delta_x \leq_{\text{sh}} \gamma^x$ for μ-a.e. $x \in \mathbb{R}^d$. For this transport, one gets

$$\begin{aligned}
\int_{\mathbb{R}^d} G(x,\gamma^x)\,\mathrm{d}\mu(x) &= \int_{\mathbb{R}^d} F(\delta_x,\gamma^x)\,\mathrm{d}\mu(x) \\
&= \int_{\mathbb{R}^d} \frac{1}{d}\big(\operatorname{var}(\gamma^x) - \operatorname{var}(\delta_x)\big)\,\mathrm{d}\mu(x) \\
&= \frac{1}{d}\int_{\mathbb{R}^d} \operatorname{var}(\gamma^x)\,\mathrm{d}\mu(x) \\
&= \frac{1}{d}\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d} |y|^2\,\mathrm{d}\gamma^x(y) - \left|\int_{\mathbb{R}^d} y\,\mathrm{d}\gamma^x(y)\right|^2\right)\mathrm{d}\mu(x) \\
&= \frac{1}{d}\left(\int_{\mathbb{R}^d} |y|^2\,\mathrm{d}\nu(y) - \int_{\mathbb{R}^d} |x|^2\,\mathrm{d}\mu(x)\right) \\
&= F(\mu,\nu),
\end{aligned}$$

where we have used the facts that $x = [\gamma^x]$ and $[\mu] = [\nu]$. The equality $F(\mu,\nu) = H(\mu,\nu)$ then directly follows from Proposition 17.

- We check the assumptions of Theorem 37. For each $r > 0$ and $y \in \mathbb{R}^d$, we define the measure $\lambda_{y,r} = 2I_d\, g(\cdot, y)\mathscr{L}^d \llcorner B_r(y) \in \mathscr{M}(\mathbb{R}^d, S_d^+)$, where g is the Green function of the Laplacian on $B_r(y)$. Let $\psi \in \mathscr{U}_{\mathrm{b}}(\mathbb{R}^d)$ and $-\psi = (-\psi)^G$. Using the facts that
$$\mathrm{tr}\left(\frac{1}{2}\nabla^2 \lambda_{y,r}\right) = \mathscr{H}^{d-1}\big(\partial B_r(y)\big)^{-1}\mathscr{H}^{d-1} \llcorner \partial B_r(y) - \delta_y$$
and
$$\int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\lambda_{y,r}}{\mathrm{d}|\lambda_{y,r}|}\right) \mathrm{d}|\lambda_{y,r}| = \int_{B_r(y)} 2g(x,y)\,\mathrm{d}x = \frac{r^2}{d},$$
according to Remark 27, we have
$$⨍_{\partial B_r(y)} -\psi(x)\,\mathrm{d}\mathscr{H}^{d-1}(x) + \frac{r^2}{d} \geq -\psi(y).$$
This implies that
$$⨍_{\partial B_r(y)} \left(-\psi(x) + \frac{|x|^2}{d}\right) \mathrm{d}\mathscr{H}^{d-1}(x) \geq -\psi(y) + \frac{|y|^2}{d}.$$
Then, by Definition 4, $\Psi(\cdot) := -\psi(\cdot) + \frac{1}{d}|\cdot|^2 \in \mathscr{SH}(\mathbb{R}^d)$. Fix an arbitrary $\varepsilon > 0$. Defining for each $x \in \mathbb{R}^d$,
$$\Psi_\varepsilon(x) = \int_{\mathbb{R}^d} \Psi(y)\eta_\varepsilon(x-y)\,\mathrm{d}y,$$
where $\eta_\varepsilon(\cdot) = \varepsilon^{-d}\eta(\cdot/\varepsilon) \in C_{\mathrm{c}}^\infty(\mathbb{R}^d)$ and η is a standard mollifier as in Lemma 46, we observe that $\Psi_\varepsilon \in C_{\mathrm{b}}^2(\mathbb{R}^d)$ is subharmonic on $\mathbb{R}^d$ (the reader may consult the proof of Lemma 46). For each $p \in \mathscr{P}_2(\mathbb{R}^d)$ and $y \in \mathbb{R}^d$ such that $\Delta\big(\frac{1}{2}u\,m\big) = p - \delta_y$ for some nonnegative measure m on $\mathbb{R}^d$ and $u \in L^1(\mathbb{R}^d, \mathrm{d}m)$ such that $u \geq 0$ m-a.e. on $\mathbb{R}^d$, it holds $\delta_y \leq_{\mathrm{sh}} p$. Thus,
$$\int_{\mathbb{R}^d} \Psi_\varepsilon(x)\,\mathrm{d}p\,(x) \geq \Psi_\varepsilon(y), \tag{51}$$
since $\Psi_\varepsilon \in C_{\mathrm{b}}^2(\mathbb{R}^d) \cap \mathscr{SH}(\mathbb{R}^d)$ (see Definition 5). Next, observing that $\frac{1}{d}|\cdot|_\varepsilon^2 \in C_{\mathrm{b}}^2(\mathbb{R}^d) \cap \mathscr{SH}(\mathbb{R}^d)$, where
$$\frac{|x|_\varepsilon^2}{d} = \int_{\mathbb{R}^d} \frac{|y|^2}{d}\eta_\varepsilon(x-y)\,\mathrm{d}y \quad \text{and} \quad \Delta\frac{|x|_\varepsilon^2}{d} = 2 \quad \forall\, x \in \mathbb{R}^d,$$
if $\Delta\big(\frac{1}{2}um\big) = p - \delta_y$, we have
$$\int_{\mathbb{R}^d} u\,\mathrm{d}m = \int_{\mathbb{R}^d} \frac{1}{2}\Delta\left(\frac{|x|_\varepsilon^2}{d}\right) u(x)\,\mathrm{d}m(x) = \int_{\mathbb{R}^d} \frac{|x|_\varepsilon^2}{d}\,\mathrm{d}p\,(x) - \frac{|y|_\varepsilon^2}{d}.$$
Using (51), the above formula and the fact that
$$\int_{\mathbb{R}^d} u\,\mathrm{d}m = \int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\lambda}{\mathrm{d}|\lambda|}\right)\mathrm{d}|\lambda|,$$
where $\mathrm{d}\lambda = uI_d\,\mathrm{d}m$, we get
$$\int_{\mathbb{R}^d} \left(\Psi_\varepsilon(x) - \frac{|x|_\varepsilon^2}{d}\right)\mathrm{d}p\,(x) + \int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\lambda}{\mathrm{d}|\lambda|}\right)\mathrm{d}|\lambda| \geq \Psi_\varepsilon(y) - \frac{|y|_\varepsilon^2}{d}.$$
This, in view of Remark 27, implies that $\Psi_\varepsilon - \frac{1}{d}|\cdot|_\varepsilon^2 = \big(\Psi_\varepsilon - \frac{1}{d}|\cdot|_\varepsilon^2\big)^G$ on $\mathbb{R}^d$. Fix now a sequence of sufficiently small positive numbers $(\varepsilon_n)_{n\in\mathbb{N}}$ such that $\varepsilon_n \to 0+$ as $n \to +\infty$. Observe that
$$\Psi_{\varepsilon_n}(x) - \frac{|x|_{\varepsilon_n}^2}{d} \longrightarrow -\psi(x) \quad \forall\, x \in \mathbb{R}^d$$
(here we use that the convolution of a subharmonic function converges to this function everywhere, in view of the monotonicity condition of subharmonic functions; see, for instance, [22, Section 2.9]). Altogether, we have defined the approximation sequence $(\psi_n)_{n\in\mathbb{N}} = \big(-\Psi_{\varepsilon_n} + \frac{1}{d}|\cdot|_{\varepsilon_n}^2\big)_{n\in\mathbb{N}}$ for ψ in the sense of Theorem 37. Therefore, according to Theorem 37, $F = H$ when f is defined by (45).

Remark 38. Applying Theorem 37 when f is defined by (45) allows us to recover the result of [14], namely the fact that whenever μ and ν are in subharmonic order, then there exists at least one transport $\gamma \in \Pi(\mu,\nu)$ such that $\gamma = \gamma^x \otimes \mu$ with $\delta_x \le_{\mathrm{sh}} \gamma^x$ for μ-a.e. $x \in \mathbb{R}^d$.

Our second type of approximation assumption is related to the assumption (A4).

Theorem 39. *Let $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$ and* (A4) *hold. Assume that for each $\varphi \in \Phi_{\mathrm{bb},2}(\mathbb{R}^d)$ such that $\varphi = \varphi^G$ there exists a sequence $(\varphi_n)_{n\in\mathbb{N}} \subset C_{\mathrm{b}}^2(\mathbb{R}^d)$ such that $\varphi_n = \varphi_n^G$ for each $n \in \mathbb{N}$ and $\langle \nu-\mu,\varphi_n\rangle \to \langle \nu-\mu,\varphi\rangle$ as $n \to +\infty$. Then $F(\mu,\nu) = H(\mu,\nu)$.*

Proof. By Proposition 17, $F(\mu,\nu) \le H(\mu,\nu)$. Next, using the assumption of Theorem 39, together with Theorems 32 and 34, we have $H(\mu,\nu) \le F(\mu,\nu)$. This completes our proof of Theorem 39. □

Example (B). For some $B \in S_d^+$ and for each $A \in S_d$,

$$f(A) = \begin{cases} A : B & \text{if } A \in S_d^+, \\ +\infty & \text{otherwise.} \end{cases} \tag{52}$$

Then f satisfies (A0)–(A2), (A4) and for each $A \in S_d$,

$$f^*(A) = \sup\{(A-B) : M \mid M \in S_d^+\} = \begin{cases} 0 & \text{if } A - B \le 0, \\ +\infty & \text{otherwise.} \end{cases} \tag{53}$$

Given $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$, using Theorem 13 and (53), we obtain

$$\begin{aligned} F(\mu,\nu) &= \sup\left\{\langle \nu-\mu,\psi\rangle \;\middle|\; \psi \in C_{\mathrm{b}}^2(\mathbb{R}^d),\ \frac{1}{2}\nabla^2\psi(x) - B \le 0 \ \ \forall\, x \in \mathbb{R}^d\right\} \\ &= \int_{\mathbb{R}^d} B : x\otimes x\,\mathrm{d}\big(\nu(x)-\mu(x)\big) + \sup\{\langle \nu-\mu,\varphi\rangle \mid \varphi \in C_{\mathrm{b}}^2(\mathbb{R}^d),\ \varphi \text{ is concave on } \mathbb{R}^d\} \\ &= \begin{cases} \int_{\mathbb{R}^d} B : x\otimes x\,\mathrm{d}\big(\nu(x)-\mu(x)\big) & \text{if } \mu \le_{\mathrm{c}} \nu, \\ +\infty & \text{otherwise} \end{cases} \qquad (54) \\ &= \begin{cases} f\big(\int_{\mathbb{R}^d} x\otimes x\,\mathrm{d}\big(\nu(x)-\mu(x)\big)\big) & \text{if } \mu \le_{\mathrm{c}} \nu, \\ +\infty & \text{otherwise,} \end{cases} \qquad (55) \end{aligned}$$

where to obtain (54) we have used Remark 3. Notice that for each $a \in \mathbb{R}$ and for each $\zeta \in \mathbb{R}^d$, the function

$$\psi(x) = a + x\cdot\zeta + B : x\otimes x$$

is a dual optimizer for $F(\mu,\nu)$, which is understood in the setting of (16), when $F(\mu,\nu) < +\infty$ and f is defined by (52).

The equality $F(\mu,\nu) = H(\mu,\nu)$ when f is defined by (52) can be proved based on the Strassen theorem, as well as using Theorem 39. We present both approaches, which allows the reader to verify the assumptions of Theorem 39 in the context of the example (B).

- By (54), if $F(\mu,\nu) < +\infty$, then $\mu \le_{\mathrm{c}} \nu$. It follows by the Strassen theorem (see [24]) that there exists $\gamma \in \Pi(\mu,\nu)$ such that $\delta_x \le_{\mathrm{c}} \gamma^x$ for μ-a.e. $x \in \mathbb{R}^d$ for which

$$\begin{aligned} \int_{\mathbb{R}^d} G(x,\gamma^x)\,\mathrm{d}\mu(x) &= \int_{\mathbb{R}^d} F(\delta_x,\gamma^x)\,\mathrm{d}\mu(x) \\ &= \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d} B : y\otimes y\,\mathrm{d}\gamma^x(y) - B : x\otimes x\right)\mathrm{d}\mu(x) \\ &= F(\mu,\nu). \end{aligned}$$

The equality $F(\mu,\nu) = H(\mu,\nu)$ then directly follows from Proposition 17.

- We check the assumptions of Theorem 39. Let $\varphi \in \Phi_2(\mathbb{R}^d)$. We claim that $\varphi = \varphi^G$ if and only if the function $\varphi_B \in \Phi_2(\mathbb{R}^d)$ defined by $\varphi_B(x) = \varphi(x) + B : x \otimes x$ is convex on $\mathbb{R}^d$. Indeed, if φ_B is convex, using Jensen's inequality and (54), for each $p \in \mathscr{P}_2(\mathbb{R}^d)$, we have

$$\varphi\big([p]\big) \le \int_{\mathbb{R}^d} \big(\varphi(y) + B : y \otimes y\big)\,\mathrm{d}p(y) - B : [p] \otimes [p] = \int_{\mathbb{R}^d} \varphi\,\mathrm{d}p + G\big([p], p\big), \tag{56}$$

which, in view of Remark 27 and (54), yields $\varphi^G = \varphi$ on $\mathbb{R}^d$. The same argument yields that if $\varphi = \varphi^G$ on $\mathbb{R}^d$, then for each $p \in \mathscr{P}_2(\mathbb{R}^d)$, (56) holds and hence

$$\varphi_B\big([p]\big) \le \int_{\mathbb{R}^d} \varphi_B\,\mathrm{d}p. \tag{57}$$

For each $x, y \in \mathbb{R}^d$, choosing $p = \frac{1}{2}(\delta_x + \delta_y)$ in (57), we obtain

$$\varphi_B\left(\frac{x+y}{2}\right) \le \frac{\varphi_B(x)}{2} + \frac{\varphi_B(y)}{2},$$

which, since φ_B is continuous on $\mathbb{R}^d$, implies that φ_B is convex on $\mathbb{R}^d$. This completes the proof of our claim.

Let $\varphi \in \Phi_{\mathrm{bb},2}(\mathbb{R}^d)$ satisfy $\varphi = \varphi^G$. Then φ_B is convex on $\mathbb{R}^d$. Furthermore, there exists a sequence $(\psi_k)_{k\in\mathbb{N}} \subset C^2_{\mathrm{b}}(\mathbb{R}^d)$ of convex functions such that $\int_{\mathbb{R}^d} \psi_k\,\mathrm{d}\mu \to \int_{\mathbb{R}^d} \varphi_B\,\mathrm{d}\mu$ and $\int_{\mathbb{R}^d} \psi_k\,\mathrm{d}\nu \to \int_{\mathbb{R}^d} \varphi_B\,\mathrm{d}\nu$ as $k \to +\infty$ (we refer to Remark 3). For each $k \in \mathbb{N}$, define $\varphi_k(x) = \psi_k(x) - B : x \otimes x$ for each $x \in \mathbb{R}^d$. Since $\psi_k \in C^2_{\mathrm{b}}(\mathbb{R}^d)$ is convex, $\varphi_k = \varphi_k^G \in C^2_{\mathrm{b}}(\mathbb{R}^d)$. This defines the approximation sequence for φ in the sense of Theorem 39, since

$$\int_{\mathbb{R}^d} \varphi_k\,\mathrm{d}\mu \longrightarrow \int_{\mathbb{R}^d} (\varphi_B(x) - B : x \otimes x)\,\mathrm{d}\mu(x) = \int_{\mathbb{R}^d} \varphi\,\mathrm{d}\mu$$

and

$$\int_{\mathbb{R}^d} \varphi_k\,\mathrm{d}\nu \longrightarrow \int_{\mathbb{R}^d} (\varphi_B(x) - B : x \otimes x)\,\mathrm{d}\nu(x) = \int_{\mathbb{R}^d} \varphi\,\mathrm{d}\nu$$

as $k \to +\infty$. Therefore, according to Theorem 39, $F = H$ when f is defined by (52).

Remark 40. Applying Theorem 39 when f is defined by (52) allows us to recover the Strassen theorem, namely the fact that whenever μ and ν are in convex order, there exists at least one transport $\gamma \in \Pi(\mu, \nu)$ such that $\gamma = \gamma^x \otimes \mu$ with $\delta_x \le_{\mathrm{c}} \gamma^x$ for μ-a.e. $x \in \mathbb{R}^d$. The same observation holds in the context of Theorem 50 when f is defined by (66).

4.2. *Viscosity solutions*

In this subsection, under the assumption (A3), we characterize the functions $\psi \in C_{\mathrm{b}}(\mathbb{R}^d)$ such that $-\psi = (-\psi)^G$ as viscosity solutions of the Hamilton–Jacobi–Bellman equation $-f^*\big(\frac{1}{2}\nabla^2 u\big) = 0$ in $\mathbb{R}^d$ (notice that f^* is discontinuous and takes its values in $\{0, +\infty\}$, see (6)).

Remark 41. Let (A3) hold. Since $\mathrm{dom}(f) \cap S_d^{++}$ is dense in $\mathrm{dom}(f) \ne \emptyset$ (see (A0)), for each $A \in S_d$,

$$f^*(A) = \sup_{t>0} t(\mathscr{T}(A) - 1) = \begin{cases} 0 & \text{if } \mathscr{T}(A) \le 1, \\ +\infty & \text{otherwise,} \end{cases} \tag{58}$$

where

$$\mathscr{T}(A) = \sup\big\{A : E \,\big|\, E \in S_d^{++},\ f(E) = 1\big\}. \tag{59}$$

In view of Proposition 36 and (58), if $\psi \in C^2_{\mathrm{b}}(\mathbb{R}^d)$, then $-\psi = (-\psi)^G$ on $\mathbb{R}^d$ if and only if $1 - \mathscr{T}\big(\frac{1}{2}\nabla^2\psi\big) \ge 0$ in $\mathbb{R}^d$. However, a function $\psi \in C_{\mathrm{b}}(\mathbb{R}^d)$ such that $-\psi = (-\psi)^G$ may not be regular enough to define $\nabla^2\psi$ in the classical sense. Using the theory of viscosity solutions, we can define $\nabla^2\psi$ in the viscosity sense. Taking into account Theorems 33 and 34, to derive the equality between F and H, we first show that each $\psi \in C_b(\mathbb{R}^d)$ such that $-\psi = (-\psi)^G$ is a viscosity supersolution of the equation $1 - \mathscr{T}\left(\frac{1}{2}\nabla^2 u\right) = 0$ in $\mathbb{R}^d$ (see Definition 43).

Following [13], we shall say that a function $\mathscr{F}\colon S_d \to \mathbb{R}$ is *proper* (also called *degenerate elliptic*) if

$$\mathscr{F}(A_2) \le \mathscr{F}(A_1) \quad \text{whenever } A_1 \le A_2. \tag{60}$$

Remark 42. For each $c \in \mathbb{R}$ and $A \in S_d^{++}$, the function $\mathscr{F}(B) = c - A : B$, $B \in S_d$ is proper. Furthermore, for each $B_1, B_2 \in S_d$ satisfying $B_1 \le B_2$, we have

$$\sup\left\{\frac{1}{2}B_2 : E \;\middle|\; E \in S_d^{++},\ f(E) = 1\right\} \ge \sup\left\{\frac{1}{2}B_1 : E \;\middle|\; E \in S_d^{++},\ f(E) = 1\right\}$$

and hence $c - \mathscr{T}\left(\frac{1}{2}B_2\right) \le c - \mathscr{T}\left(\frac{1}{2}B_1\right)$, where $\mathscr{T}\colon S_d \to \mathbb{R}$ is defined in (59). Thus, $\mathscr{F}(\cdot) = c - \mathscr{T}\left(\frac{1}{2}\cdot\right)$ is proper for each $c \in \mathbb{R}$.

For the reader's convenience, we recall (the reader may consult [13]) the next definition.

Definition 43. *Let $\mathscr{F}\colon S_d \to \mathbb{R}$ be proper (see* (60)*). A lower semicontinuous function $u\colon \mathbb{R}^d \to \mathbb{R}$ is a viscosity supersolution of $\mathscr{F} = 0$ (a viscosity solution of $\mathscr{F} \ge 0$) in $\mathbb{R}^d$ provided that if $\varphi \in C^2(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$ is a local minimum of $u - \varphi$, then $\mathscr{F}\left(\nabla^2\varphi(x)\right) \ge 0$.*

The definitions of a viscosity subsolution and a solution of $\mathscr{F} = 0$ are likewise, we refer to [13]. Notice that the function $\mathscr{F}$ in (60) and in Definition 43 can be discontinuous. Even more, allowing $\mathscr{F}$ to become infinite (see, for instance, [13, Example 1.11]), we observe that $\mathscr{F} = -f^*\colon S_d \to \{0, -\infty\}$ is proper. In particular, since $-f^* \le 0$, it follows that any upper semicontinuous function ψ is a viscosity subsolution of $-f^*\left(\frac{1}{2}\nabla^2 u\right) = 0$.

Proposition 44. *Let $\psi \in C_b(\mathbb{R}^d)$ satisfy $-\psi = (-\psi)^G$ on $\mathbb{R}^d$. Then ψ is a viscosity supersolution of $1 - \mathscr{T}\left(\frac{1}{2}\nabla^2 u\right) = 0$ in $\mathbb{R}^d$.*

Proof. Let $x_0 \in \mathbb{R}^d$, $\varphi \in C^2(\mathbb{R}^d)$ and x_0 be a local minimum of $\psi - \varphi$. Then there exists $r > 0$ such that $\psi(x) - \psi(x_0) \ge \varphi(x) - \varphi(x_0)$ for each $x \in \overline{B}_r(x_0)$. Fix an arbitrary $A = (a_{ij})_{i,j=1}^d \in S_d^{++}$ such that $f(A) = 1$. Let g_A be the Green function of the elliptic operator $L = -\sum_{i,j=1}^d a_{ij}\partial_{ij}$ on $B_r(x_0)$. Then, defining $\lambda = 2Ag_A(\cdot, x_0)\mathscr{L}^d \llcorner B_r(x_0) \in \mathscr{M}(\mathbb{R}^d, S_d^+)$ and $p = -A\nabla g_A(\cdot, x_0)\cdot\nu(\cdot)\mathscr{H}^{d-1} \llcorner \partial B_r(x_0) \in \mathscr{P}_2(\mathbb{R}^d)$, where $\nu(x)$ is the outward pointing unit normal to $\partial B_r(x_0)$ at x, we know that

$$\operatorname{tr}\left(\frac{1}{2}\nabla^2\lambda\right) = p - \delta_{x_0} \tag{61}$$

(see the proof of Proposition 36). Since $\varphi - \varphi(x_0) \le \psi - \psi(x_0)$ on $\overline{B}_r(x_0)$,

$$\int_{\mathbb{R}^d} \varphi\, \mathrm{d}p - \varphi(x_0) \le \int_{\mathbb{R}^d} \psi\, \mathrm{d}p - \psi(x_0) \le G(x_0, p), \tag{62}$$

where the last estimate comes from the fact that $-\psi = (-\psi)^G$ (see Remark 27). On the other hand, in view of (61), λ is a competitor in the definition of $G(x_0, p) = F(\delta_{x_0}, p)$ (see Proposition 9). Using this, (A1) and the fact that $f(A) = 1$, we obtain

$$G(x_0, p) \le \int_{\mathbb{R}^d} f\left(\frac{\mathrm{d}\lambda}{\mathrm{d}|\lambda|}\right)\mathrm{d}|\lambda| = 2\int_{B_r(x_0)} f(A) g_A(x, x_0)\,\mathrm{d}x = 2\int_{B_r(x_0)} g_A(x, x_0)\,\mathrm{d}x.$$

This, together with (62) and the fact that we can actually use φ as a test function for (61) (since the supports of the measures λ, p and δ_{x_0} are contained in $\overline{B}_r(x_0)$ and we can multiply φ by a cutoff function equal to 1 on $B_{2r}(x_0)$), implies that

$$\int_{B_r(x_0)} A : \nabla^2\varphi(x) g_A(x, x_0)\,\mathrm{d}x \le 2\int_{B_r(x_0)} g_A(x, x_0)\,\mathrm{d}x. \tag{63}$$

Then $A : \frac{1}{2}\nabla^2\varphi(x_0) \le 1$, because otherwise we could choose $r > 0$ small enough such that for some $\varepsilon > 0$ and for each $x \in B_r(x_0)$ we would have $A : \frac{1}{2}\nabla^2\varphi(x) - 1 \ge \varepsilon$, which would lead to a contradiction with (63), since $\int_{B_r(x_0)} g_A(x,x_0)\,\mathrm{d}x > 0$. Thus, $1 - A : \frac{1}{2}\nabla^2\varphi(x_0) \geq 0$ for each $A \in S_d^{++}$ satisfying $f(A) = 1$. Therefore,

$$1 - \mathscr{T}\left(\frac{1}{2}\nabla^2\varphi(x_0)\right) = 1 - \sup\left\{A : \frac{1}{2}\nabla^2\varphi(x_0) \;\middle|\; A \in S_d^{++},\ f(A) = 1\right\} \geq 0.$$

This, according to Definition 43, completes our proof of Proposition 44. □

Next, we perform a smoothing procedure by convolution with a mollifier for a viscosity solution of $1 - \mathscr{T}\left(\frac{1}{2}\nabla^2 u\right) \geq 0$ in $\mathbb{R}^d$ to obtain the classical solution.

Remark 45. The same argument as in the proof of Proposition 44 implies that for each $c \in \mathbb{R}$ and for each lower semicontinuous function $w\colon \mathbb{R}^d \to \mathbb{R}$, w is a viscosity supersolution of $c - \mathscr{T}\left(\frac{1}{2}\nabla^2 u\right) = 0$ in $\mathbb{R}^d$ if and only if for each $A \in S_d^{++}$ such that $f(A) = 1$, w is a viscosity supersolution of $c - A : \frac{1}{2}\nabla^2 u = 0$ in $\mathbb{R}^d$.

Lemma 46. *Let $A \in S_d^{++}$, $c \in \mathbb{R}$ and $w \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$ be lower semicontinuous on $\mathbb{R}^d$. Let $\eta \in C_c^\infty(\mathbb{R}^d)$, $\mathrm{supp}(\eta) = \overline{B}_1(0)$, $\eta \geq 0$, $\eta(x) = \eta(-x)$, $\int_{\mathbb{R}^d}\eta\,\mathrm{d}x = 1$ and $\eta_\varepsilon(\cdot) = \varepsilon^{-d}\eta(\cdot/\varepsilon)$ for each $\varepsilon > 0$. Assume that w is a viscosity supersolution of $c - \mathscr{T}\left(\frac{1}{2}\nabla^2 u\right) = 0$ in $\mathbb{R}^d$, where $\mathscr{T}$ is defined in* (59). *Then for each $\varepsilon > 0$, $c - \mathscr{T}\left(\frac{1}{2}\nabla^2 w_\varepsilon\right) \geq 0$ in $\mathbb{R}^d$, where $w_\varepsilon(\cdot) = \int_{\mathbb{R}^d} w(y)\eta_\varepsilon(\cdot - y)\,\mathrm{d}y$.*

Proof. Let $A \in S_d^{++}$ be such that $f(A) = 1$. Let $B \in S_d^{++}$ be the unique matrix such that $A = B^{-2}$ and for each $x \in \mathbb{R}^d$, define $v(x) = w(B^{-1}x)$ and $h(x) = v(x) - \frac{c}{d}|x|^2$.

Step 1. We prove that the following assertions are equivalent:

(i) w is a viscosity supersolution of $c - A : \frac{1}{2}\nabla^2 u = 0$ in $\mathbb{R}^d$;
(ii) v is a viscosity supersolution of $c - \frac{1}{2}\Delta u = 0$ in $\mathbb{R}^d$;
(iii) h is a viscosity supersolution of $-\frac{1}{2}\Delta u = 0$ in $\mathbb{R}^d$.

The equivalence (i) ⇔ (ii) directly follows from $A : \frac{1}{2}\nabla^2\psi(x) = \frac{1}{2}\Delta\varphi(Bx)$ whenever $\psi(x) = \varphi(Bx)$ for each $x \in \mathbb{R}^d$ and $\varphi \in C^2(\mathbb{R}^d)$, while the equivalence (ii) ⇔ (iii) is straightforward.

Step 2. We prove that the inequality $c - \frac{1}{2}\Delta v_\varepsilon \geq 0$ holds in $\mathbb{R}^d$, where $v_\varepsilon(\cdot) = \int_{\mathbb{R}^d} v(y)\widetilde{\eta}_\varepsilon(\cdot - y)\,\mathrm{d}y$ and $\widetilde{\eta}_\varepsilon(\cdot) = \left(\det(B)\right)^{-1}\eta_\varepsilon(B^{-1}\cdot)$. By definition, $\widetilde{\eta}_\varepsilon \in C_c^\infty(\mathbb{R}^d)$, $\mathrm{supp}(\widetilde{\eta}_\varepsilon) \subset \left\{Bx \;\middle|\; x \in \overline{B}_\varepsilon(0)\right\}$, $\widetilde{\eta}_\varepsilon \geq 0$, $\widetilde{\eta}_\varepsilon(x) = \widetilde{\eta}_\varepsilon(-x)$ and $\int_{\mathbb{R}^d}\widetilde{\eta}_\varepsilon(x)\,\mathrm{d}x = 1$. Since w is a viscosity supersolution of $c - A : \frac{1}{2}\nabla^2 u = 0$ in $\mathbb{R}^d$ (see Remark 45), by the equivalence (i) ⇔ (iii) proved in Step 1, h is a viscosity supersolution of $-\frac{1}{2}\Delta u = 0$ in $\mathbb{R}^d$. Let us show that $h_\varepsilon(\cdot) = \int_{\mathbb{R}^d} h(y)\widetilde{\eta}_\varepsilon(\cdot - y)\,\mathrm{d}y$ is superharmonic in $\mathbb{R}^d$, which is equivalent to the property $c - \frac{1}{2}\Delta v_\varepsilon \geq 0$ in $\mathbb{R}^d$. The proof is based on the fact that h is viscosity superharmonic in $\mathbb{R}^d$ if and only if $⨍_{\partial B_r(x_0)} h(x)\,\mathrm{d}\mathscr{H}^{d-1}(x) \leq h(x)$ for each $x_0 \in \mathbb{R}^d$ and $r > 0$, namely h is superharmonic in $\mathbb{R}^d$ (see Definition 4). Thus, changing the variables and applying Fubini's theorem, we obtain

$$\begin{aligned}
⨍_{\partial B_r(x)} h_\varepsilon(y)\,\mathrm{d}\mathscr{H}^{d-1}(y) &= ⨍_{\partial B_r(x)} \mathrm{d}\mathscr{H}^{d-1}(y)\int_{\mathbb{R}^d} h(z)\widetilde{\eta}_\varepsilon(y-z)\,\mathrm{d}z \\
&= \int_{\mathbb{R}^d}\mathrm{d}z ⨍_{\partial B_r(x)} h(y-z)\widetilde{\eta}_\varepsilon(z)\,\mathrm{d}\mathscr{H}^{d-1}(y) \\
&\leq \int_{\mathbb{R}^d} h(x-z)\widetilde{\eta}_\varepsilon(z)\,\mathrm{d}z \\
&= h_\varepsilon(x),
\end{aligned}$$

where we have also used that $⨍_{\partial B_r(x-z)} h(y)\,\mathrm{d}\mathscr{H}^{d-1}(y) \leq h(x-z)$, since h is superharmonic in $\mathbb{R}^d$. Then, by Definition 4, h_ε is superharmonic in $\mathbb{R}^d$. This, since $h_\varepsilon \in C^2(\mathbb{R}^d)$, implies that $-\frac{1}{2}\Delta h_\varepsilon \geq 0$ in $\mathbb{R}^d$ and hence $c - \frac{1}{2}\Delta v_\varepsilon \geq 0$ in $\mathbb{R}^d$ as desired.

Step 3. Using the result of Step 2, the fact that $A:\nabla^2\eta_\varepsilon(x) = \det(B)\Delta\widetilde{\eta}_\varepsilon(Bx)$ for each $x \in \mathbb{R}^d$ and changing the variables, for each $x \in \mathbb{R}^d$ we deduce the following:

$$\begin{aligned}
0 &\le c - \frac{1}{2}\Delta v_\varepsilon(Bx)\\
&= c - \frac{1}{2}\int_{\mathbb{R}^d} v(y)\Delta\widetilde{\eta}_\varepsilon(Bx-y)\,\mathrm{d}y\\
&= c - \frac{\big(\det(B)\big)^{-1}}{2}\int_{\mathbb{R}^d} v(y)A:\nabla^2\eta_\varepsilon(x-B^{-1}y)\,\mathrm{d}y\\
&= c - \frac{1}{2}\int_{\mathbb{R}^d} v(By)A:\nabla^2\eta_\varepsilon(x-y)\,\mathrm{d}y\\
&= c - A:\frac{1}{2}\nabla^2 w_\varepsilon(x).
\end{aligned}$$

This, in view of Remark 45 and the fact that $A \in S_d^{++}$ satisfying $f(A) = 1$ was arbitrarily chosen (observe that a *classical* supersolution is a viscosity supersolution, and a viscosity supersolution of class C^2 is a *classical* supersolution), completes our proof of Lemma 46. □

Now we characterize the dual competitors in (40) as viscosity solutions of $-f^*\big(\frac{1}{2}\nabla^2 u\big) = 0$ in $\mathbb{R}^d$.

Proposition 47. *Let $\psi \in C_\mathrm{b}(\mathbb{R}^d)$ and* (A3) *hold. Then $-\psi = (-\psi)^G$ on $\mathbb{R}^d$ if and only if ψ is a viscosity solution of*

$$-f^*\left(\frac{1}{2}\nabla^2 u\right) = 0$$

in $\mathbb{R}^d$.

Remark 48. In [25, Theorem 4.2], in the Markovian case, it was proved that the dynamic value function is a viscosity solution of the Hamilton–Jacobi–Bellman equation. It is worth noting that, unlike [25], in our case f^* is not continuous and $\mathrm{dom}(f^*)$ is a closed convex subset of S_d. Furthermore, we provide an analytical proof of Proposition 47 that is different from the proof of [25, Theorem 4.2], showing in addition that if ψ is a viscosity solution of the Hamilton–Jacobi–Bellman equation $-f^*\big(\frac{1}{2}\nabla^2 u\big) = 0$ in $\mathbb{R}^d$, then $-\psi$ is invariant under the G-transform.

Proof. By Proposition 44, (58) and Definition 43 (where we allow $\mathscr{F}$ to be discontinuous and, even more, to become infinite), if $\psi \in C_\mathrm{b}(\mathbb{R}^d)$ and $-\psi = (-\psi)^G$ on $\mathbb{R}^d$, then ψ is a viscosity solution of $-f^*\big(\frac{1}{2}\nabla^2 u\big) = 0$ in $\mathbb{R}^d$.

Conversely, if $\psi \in C_\mathrm{b}(\mathbb{R}^d)$ is a viscosity solution of $-f^*\big(\frac{1}{2}\nabla^2 u\big) = 0$ in $\mathbb{R}^d$, by (58) and Definition 43, ψ is a viscosity supersolution of $1-\mathscr{T}\big(\frac{1}{2}\nabla^2 u\big) = 0$ in $\mathbb{R}^d$. Next, applying Lemma 46, we deduce that for each $\varepsilon > 0$, it holds $1-\mathscr{T}\big(\frac{1}{2}\nabla^2\psi_\varepsilon\big) \ge 0$ in $\mathbb{R}^d$, where $\psi_\varepsilon(\cdot) = \int_{\mathbb{R}^d}\psi(y)\eta_\varepsilon(\cdot - y)\,\mathrm{d}y$ and η is the mollifier of Lemma 46. Hence $-f^*\big(\frac{1}{2}\nabla^2\psi_\varepsilon\big) = 0$ in $\mathbb{R}^d$ (see (58)). This, according to Proposition 36, yields the equality $-\psi_\varepsilon = (-\psi_\varepsilon)^G$ on $\mathbb{R}^d$ for each $\varepsilon > 0$. According to Proposition 22, for each $x \in \mathbb{R}^d$ and $\varepsilon > 0$, there exist $p, p_\varepsilon \in \mathscr{P}_2(\mathbb{R}^d)$ such that

$$(-\psi_\varepsilon)^G(x) = \int_{\mathbb{R}^d} -\psi_\varepsilon\,\mathrm{d}p_\varepsilon + G(x, p_\varepsilon) \tag{64}$$

and

$$(-\psi)^G(x) = \int_{\mathbb{R}^d} -\psi\,\mathrm{d}p + G(x, p). \tag{65}$$

Using (65), the fact that $\psi \in C_{\rm b}(\mathbb{R}^d)$, Lebesgue's dominated convergence theorem, (64) and the equality $-\psi_\varepsilon = (-\psi_\varepsilon)^G$, for each $x \in \mathbb{R}^d$, we have

$$\begin{aligned}(-\psi)^G(x) &= \int_{\mathbb{R}^d} -\psi \, \mathrm{d}p + G(x,p)\\ &= \lim_{\varepsilon\to 0+} \int_{\mathbb{R}^d} -\psi_\varepsilon \, \mathrm{d}p + G(x,p)\\ &\geq \lim_{\varepsilon\to 0+} \int_{\mathbb{R}^d} -\psi_\varepsilon \, \mathrm{d}p_\varepsilon + G(x,p_\varepsilon)\\ &= \lim_{\varepsilon\to 0+} -\psi_\varepsilon(x)\\ &= -\psi(x).\end{aligned}$$

This, in view of the fact that $-\psi \geq (-\psi)^G$ on $\mathbb{R}^d$, proves the equality $-\psi = (-\psi)^G$ on $\mathbb{R}^d$ and completes our proof of Proposition 47. □

The next corollary is a direct consequence of Theorem 33 and Proposition 47.

Corollary 49. *Let $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ and* (A3), (A4) *hold. Then*

$$H(\mu,\nu) = \sup\left\{ \langle \nu-\mu, \psi\rangle \;\middle|\; \psi \in C_{\rm b}(\mathbb{R}^d) \text{ is a viscosity solution of } -f^*\left(\frac{1}{2}\nabla^2 u\right) = 0 \text{ in } \mathbb{R}^d \right\}.$$

4.3. *F = H under the assumptions* (A3) *and* (A4)

Theorem 50. *Let* (A3) *and* (A4) *hold. Then $F = H$ on $\mathscr{P}_2(\mathbb{R}^d) \times \mathscr{P}_2(\mathbb{R}^d)$.*

Proof. Let $\psi \in C_{\rm b}(\mathbb{R}^d)$ be a viscosity solution of $-f^*\left(\frac{1}{2}\nabla^2 u\right) = 0$ in $\mathbb{R}^d$. Then, performing the smoothing procedure as in the proof of Proposition 47, we obtain a sequence of functions $\psi_n \in C^2_{\rm b}(\mathbb{R}^d)$ such that $-f^*\left(\frac{1}{2}\nabla^2 \psi_n\right) = 0$ in $\mathbb{R}^d$, $\int_{\mathbb{R}^d} \psi_n \, \mathrm{d}\mu \to \int_{\mathbb{R}^d} \psi \, \mathrm{d}\mu$ and $\int_{\mathbb{R}^d} \psi_n \, \mathrm{d}\nu \to \int_{\mathbb{R}^d} \psi \, \mathrm{d}\nu$ as $n \to +\infty$. Using this, Corollary 49 and Theorem 13, we obtain $H(\mu,\nu) \leq F(\mu,\nu)$. Therefore, $H(\mu,\nu) = F(\mu,\nu)$, since $H(\mu,\nu) \geq F(\mu,\nu)$ by Proposition 17. This completes our proof of Theorem 50. □

Example (C). For each $A \in S_d$,

$$f(A) = \begin{cases} \operatorname{tr}(A) & \text{if } A \in S_d^+,\\ +\infty & \text{otherwise.}\end{cases} \tag{66}$$

Clearly, f satisfies (A0)–(A4). Then for each $A \in S_d$,

$$\mathscr{T}(A) \leq 1 \iff A - I_d \leq 0. \tag{67}$$

Indeed, $A : M \leq 1$ for each $M \in S_d^{++}$ such that $\operatorname{tr}(M) = 1$ if and only if $(A - I_d) : M \leq 0$ for each $M \in S_d^+$, which is equivalent to saying that $A - I_d \leq 0$. Using this and (59), one deduces (67).

Given $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$, applying Theorem 13, (58) and (67), we compute

$$\begin{aligned} F(\mu,\nu) &= \sup\left\{ \langle \nu-\mu, \psi\rangle \;\middle|\; \psi \in C^2_{\rm b}(\mathbb{R}^d),\ \frac{1}{2}\nabla^2\psi(x) - I_d \leq 0 \ \forall\, x \in \mathbb{R}^d \right\}\\ &= \left\langle \nu-\mu, |\cdot|^2 \right\rangle + \sup\left\{ \langle \nu-\mu, u\rangle \;\middle|\; u \in C^2_{\rm b}(\mathbb{R}^d),\ u \text{ is concave on } \mathbb{R}^d \right\}\\ &= \begin{cases} \operatorname{var}(\nu) - \operatorname{var}(\mu) & \text{if } \mu \leq_{\rm c} \nu,\\ +\infty & \text{otherwise}\end{cases} \qquad (68)\\ &= \begin{cases} f\left(\int_{\mathbb{R}^d} x\otimes x \, \mathrm{d}\big(\nu(x) - \mu(x)\big)\right) & \text{if } \mu \leq_{\rm c} \nu,\\ +\infty & \text{otherwise,}\end{cases} \qquad (69)\end{aligned}$$

where we have used that $\left\langle \nu-\mu, |\cdot|^2 \right\rangle = \operatorname{var}(\nu) - \operatorname{var}(\mu)$ if $\mu \leq_{\rm c} \nu$ (in this case $[\mu] = [\nu]$). For each $a \in \mathbb{R}$ and $\xi \in \mathbb{R}^d$, the function $\psi(x) = a + \xi\cdot x + |x|^2$ is a dual optimizer for $F(\mu,\nu)$, which is understood in the setting of (16), when $F(\mu,\nu)<+\infty$ and f is defined by (66). By Theorem 50, $F=H$ when f is defined by (66).

Example (D). For each $A\in S_d$,

$$f(A)=\begin{cases}\lambda_{\max}(A) & \text{if } A\in S_d^+,\\ +\infty & \text{otherwise,}\end{cases} \tag{70}$$

where $\lambda_{\max}(A)=\max\{Ax\cdot x \mid |x|=1\}$ is the largest eigenvalue of A. Clearly, f satisfies (A0)–(A4). Every matrix $A\in S_d$ has d real eigenvalues $\lambda_1,\dots,\lambda_d$ such that $|\lambda_1|\ge|\lambda_2|\ge\cdots\ge|\lambda_d|$ with corresponding eigenvectors $v_1,\dots,v_d\in\mathbb{R}^d$ forming an orthonormal basis of $\mathbb{R}^d$. Denote $J_+=\{j\in\{1,\dots,d\}\mid \lambda_j\ge 0\}$, $J_-=\{1,\dots,d\}\setminus J_+$, $A^+=\sum_{j\in J_+}\lambda_j v_j\otimes v_j$, $A^-=\sum_{j\in J_-}\lambda_j v_j\otimes v_j$ so that $A=A^++A^-$. We claim that

$$\mathcal{T}(A)\le 1 \iff \sum_{j\in J_+}\lambda_j\le 1. \tag{71}$$

Indeed, if $A\in S_d$, then $A=PDP^{\mathrm{T}}$, where $D=\operatorname{diag}(\lambda_1,\dots,\lambda_d)$ and $P=(v_1\cdots v_d)\in S_d$ is the orthonormal matrix with columns $v_1,\dots,v_d$. Let $\varepsilon\in(0,1)$ be fairly small. Define $M=PEP^{\mathrm{T}}$, where $E=(e_{ij})_{i,j=1}^d\in S_d^{++}$ is the diagonal matrix satisfying $e_{ii}=\varepsilon$ if $\lambda_i<0$ and $e_{ii}=1$ otherwise. Next, we compute

$$A:M=\operatorname{tr}(PDP^{\mathrm{T}}PEP^{\mathrm{T}})=\operatorname{tr}(DE)=\sum_{j\in J_+}\lambda_j+\varepsilon\sum_{j\in J_-}\lambda_j.$$

Thus, assuming that $\mathcal{T}(A)\le 1$ and letting $\varepsilon\to 0+$, we have $\sum_{j\in J_+}\lambda_j\le 1$. Conversely, suppose that $\sum_{j\in J_+}\lambda_j\le 1$. For each $M\in S_d^{++}$ such that $\lambda_{\max}(M)=1$,

$$A:M=(A^++A^-):M\le\operatorname{tr}(A^+M)=\sum_{j\in J_+}\lambda_j Mv_j\cdot v_j\le\sum_{j\in J_+}\lambda_j\lambda_{\max}(M)\le 1.$$

This proves our claim.

Given $\mu,\nu\in\mathscr{P}_2(\mathbb{R}^d)$, using Theorem 13, (58) and (71), we deduce that

$$F(\mu,\nu)=\sup\left\{\langle\nu-\mu,\psi\rangle \;\middle|\; \psi\in C_{\mathrm{b}}^2(\mathbb{R}^d),\ \sum_{j\in J_+}\lambda_j\left(\frac12\nabla^2\psi(x)\right)\le 1\ \ \forall\, x\in\mathbb{R}^d\right\}. \tag{72}$$

Since $\frac1d\operatorname{tr}(A)\le\lambda_{\max}(A)\le\operatorname{tr}(A)$ for each $A\in S_d^+$, we have $\frac1d\widetilde F(\mu,\nu)\le F(\mu,\nu)\le\widetilde F(\mu,\nu)$, where

$$\widetilde F(\mu,\nu)=\inf\left\{\int_{\mathbb{R}^d}\operatorname{tr}\left(\frac{\mathrm{d}\lambda}{\mathrm{d}|\lambda|}\right)\mathrm{d}|\lambda| \;\middle|\; \operatorname{tr}\left(\frac12\nabla^2\lambda\right)=\nu-\mu\right\}.$$

In particular, if $d=1$, then $\widetilde F(\mu,\nu)=F(\mu,\nu)$ and the expression of $F(\mu,\nu)$ in terms of μ and ν comes from (68), (69). Then, in view of (68),

$$F(\mu,\nu)<+\infty \iff \mu\le_{\mathrm{c}}\nu.$$

Assume that $\mu\le_{\mathrm{c}}\nu$. Define

$$\alpha=\sup\{\langle\nu-\mu,\psi\rangle \mid \psi\in C_{\mathrm{b}}^2(\mathbb{R}^d)\text{ is convex and }\Delta\psi=2\text{ in }\mathbb{R}^d\}.$$

For each $\xi\in\partial B_1(0)$, $x\in\mathbb{R}^d\mapsto|x\cdot\xi|^2$ belongs to $C_{\mathrm{b}}^2(\mathbb{R}^d)$, is convex and $\Delta|x\cdot\xi|^2=2$ in $\mathbb{R}^d$. Then

$$\begin{aligned} f\left(\int_{\mathbb{R}^d}x\otimes x\,\mathrm{d}\big(\nu(x)-\mu(x)\big)\right)&=\max_{\xi\in\partial B_1(0)}\xi\otimes\xi:\int_{\mathbb{R}^d}x\otimes x\,\mathrm{d}\big(\nu(x)-\mu(x)\big)\\ &=\max_{\xi\in\partial B_1(0)}\int_{\mathbb{R}^d}|x\cdot\xi|^2\,\mathrm{d}\big(\nu(x)-\mu(x)\big)\\ &\le\alpha. \end{aligned} \tag{73}$$

Let $\psi\in C_{\mathrm{b}}^2(\mathbb{R}^d)$ be a convex function such that $\Delta\psi=2$ in $\mathbb{R}^d$. Define $\varphi=\psi-\frac1d|\cdot|^2\in C_{\mathrm{b}}^2(\mathbb{R}^d)$. Then φ is harmonic in $\mathbb{R}^d$. Since $\varphi\in\Phi_2(\mathbb{R}^d)$, using [15, Theorem 2.10], we deduce that there exists $A\in S_d$ such that $\operatorname{tr}(A)=0$, $A+\frac{I_d}{d}\ge 0$ and $\nabla^2\varphi(x)=2A$ for each $x\in\mathbb{R}^d$. Then there exist $a\in\mathbb{R}$ and

$\zeta \in \mathbb{R}^d$ such that $\psi(x) = a + \zeta \cdot x + \left(A + \frac{I_d}{d}\right) : x \otimes x$ for each $x \in \mathbb{R}^d$. Thus, taking into account that $\mu \le_{\mathrm{c}} \nu$, we have

$$\alpha = \sup\left\{ \int_{\mathbb{R}^d} \left(A + \frac{I_d}{d}\right) : x \otimes x \,\mathrm{d}\big(\nu(x) - \mu(x)\big) \;\middle|\; A + \frac{I_d}{d} \in S_d^+,\ \mathrm{tr}(A) = 0 \right\}. \tag{74}$$

For each $A + \frac{I_d}{d} \in S_d^+$, where $\mathrm{tr}(A) = 0$, there exist eigenvalues $\lambda_1 \ge \cdots > \lambda_d \ge 0$ and corresponding eigenvectors $v_1, \ldots, v_d$ forming an orthonormal basis of $\mathbb{R}^d$ such that we have $A + \frac{I_d}{d} = \sum_{i=1}^d \lambda_i v_i \otimes v_i$ and $\sum_{i=1}^d \lambda_i = 1$. This, together with (72), (73) and (74), yields

$$f\left(\int_{\mathbb{R}^d} x \otimes x \,\mathrm{d}\big(\nu(x) - \mu(x)\big) \right) = \alpha \le F(\mu, \nu).$$

Next, let $\psi \in C_{\mathrm{b}}^2(\mathbb{R}^d)$ be such that $\sum_{j \in J_+} \lambda_j \left(\frac{1}{2} \nabla^2 \psi(x)\right) \le 1$ for each $x \in \mathbb{R}^d$. Then there exists a subharmonic function $u \in C_{\mathrm{b}}^2(\mathbb{R}^d)$ such that $\Delta u = 2 - \Delta \psi \ge 0$. Using [15, Theorem 2.10], we obtain a matrix $A \in S_d$ such that $\mathrm{tr}(A) = 0$ and $\nabla^2 u(x) + \nabla^2 \psi(x) = 2A + \frac{2I_d}{d}$ for each $x \in \mathbb{R}^d$. Then there exist $a \in \mathbb{R}$ and $\zeta \in \mathbb{R}^d$ such that $\psi(x) = a + \zeta \cdot x + \left(A + \frac{I_d}{d}\right) : x \otimes x - u(x)$ for each $x \in \mathbb{R}^d$. Taking into account the constraint in (72) and the assumption $\mu \le_{\mathrm{c}} \nu$, we obtain

$$F(\mu, \nu) = \sup\left\{ \int_{\mathbb{R}^d} \left(\left(A + \frac{I_d}{d}\right) : x \otimes x - u(x) \right) \mathrm{d}\big(\nu(x) - \mu(x)\big) \;\middle|\; \begin{array}{l} A \in S_d,\ \mathrm{tr}(A) = 0, \\ u \in C_{\mathrm{b}}^2(\mathbb{R}^d),\ \Delta u \ge 0 \text{ in } \mathbb{R}^d \\ \text{and } \sum_{j \in J_+} \lambda_j \left(\left(A + \frac{I_d}{d}\right) - \frac{1}{2} \nabla^2 u(x) \right) \le 1 \\ \forall\, x \in \mathbb{R}^d \end{array} \right\}.$$

Conjecture. *If $\mu \le_{\mathrm{c}} \nu$, then*

$$f\left(\int_{\mathbb{R}^d} x \otimes x \,\mathrm{d}\big(\nu(x) - \mu(x)\big) \right) = F(\mu, \nu).$$

As in the example (C), in our opinion, a dual optimizer for $F(\mu, \nu)$ should be sought among convex functions when f is defined by (70) and $\mu \le_{\mathrm{c}} \nu$. We expect that such an optimizer exists and that its Laplacian is equal to 2 in $\mathbb{R}^d$. In dimension 1 and in general when $\mu \le_{\mathrm{sh}} \nu$, the above conjecture is true and

$$F(\mu, \nu) = \frac{1}{d} \mathrm{var}(\nu) - \frac{1}{d} \mathrm{var}(\mu),$$

where for each $a \in \mathbb{R}$, for each $\zeta \in \mathbb{R}^d$ and for each $A \in S_d$ such that $A + \frac{I_d}{d} \ge 0$ and $\mathrm{tr}(A) = 0$, the function $\psi(x) = a + \zeta \cdot x + \left(A + \frac{I_d}{d}\right) : x \otimes x$ is a dual optimizer for $F(\mu, \nu)$, which is understood in the setting of (16).

By Theorem 50, $F = H$ when f is defined by (70).

Acknowledgments

I would like to warmly thank Guy Bouchitté for introducing me to this problem, for stimulating discussions, valuable comments on the manuscript and also for bringing [19] to my attention. Some discussions with Thierry Champion and Jean-Jacques Alibert were useful. I am also grateful to the anonymous referee for carefully reading the paper and for comments and suggestions that helped improve the paper.

Declaration of interests

The author does not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and has declared no affiliations other than their research organizations.

References

[1] A. Alfonsi, J. Corbetta and B. Jourdain, “Sampling of probability measures in the convex order by Wasserstein projection”, *Ann. Inst. Henri Poincaré, Probab. Stat.* **56** (2020), no. 3, pp. 1706–1729.

[2] J.-J. Alibert, G. Bouchitté and T. Champion, “A new class of costs for optimal transport planning”, *Eur. J. Appl. Math.* **30** (2019), no. 6, pp. 1229–1263.

[3] L. Ambrosio, N. Fusco and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford Mathematical Monographs, Clarendon Press, 2000.

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

[5] J. D. Backhoff-Veraguas, M. Beiglböck, M. Huesmann and S. Källblad, “Martingale Benamou–Brenier: a probabilistic perspective”, *Ann. Probab.* **48** (2020), no. 5, pp. 2258–2289.

[6] J. D. Backhoff-Veraguas, M. Beiglböck and G. Pammer, “Existence, duality, and cyclical monotonicity for weak transport costs”, *Calc. Var. Partial Differ. Equ.* **58** (2019), no. 6, article no. 203 (28 pages).

[7] J. D. Backhoff-Veraguas and G. Pammer, “Applications of weak transport theory”, *Bernoulli* **28** (2022), no. 1, pp. 370–394.

[8] M. Beiglböck and N. Juillet, “Shadow couplings”, *Trans. Am. Math. Soc.* **374** (2021), pp. 4973–5002.

[9] D. P. Bertsekas and S. E. Shreve, *Stochastic optimal control*, Mathematics in Science and Engineering, vol. 139, Academic Press Inc., 1978.

[10] G. Bouchitté, “Convex analysis and duality methods”, in *Encyclopedia of mathematical physics* (J.-P. Françoise, G. L. Naber and S. T. Tsou, eds.), Academic Press Inc., 2006, pp. 642–652.

[11] L. Brasco, “A survey on dynamical transport distances”, *J. Math. Sci., New York* **181** (2012), no. 6, pp. 755–781.

[12] H. R. Brezis, *Functional analysis, Sobolev spaces and partial differential equations*, Universitext, Springer, 2011.

[13] M. G. Crandall, H. Ishii and P.-L. Lions, “User’s guide to viscosity solutions of second order partial differential equations”, *Bull. Am. Math. Soc.* **27** (1992), no. 1, pp. 1–67.

[14] N. A. Ghoussoub, Y.-H. Kim and T. Lim, “Optimal Brownian stopping when the source and target are radially symmetric distributions”, *SIAM J. Control Optim.* **58** (2020), no. 5, pp. 2765–2789.

[15] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, Classics in Mathematics, Springer, 2001.

[16] N. Gozlan and N. Juillet, “On a mixture of Brenier and Strassen theorems”, *Proc. Lond. Math. Soc.* **120** (2020), no. 3, pp. 434–463.

[17] N. Gozlan, C. Roberto, P.-M. Samson and P. Tetali, “Kantorovich duality for general transport costs and applications”, *J. Funct. Anal.* **273** (2017), no. 11, pp. 3327–3405.

[18] E. Hopf, “A remark on linear elliptic differential equations of second order”, *Proc. Am. Math. Soc.* **3** (1952), pp. 791–793.

[19] M. Huesmann and D. Trevisan, “A Benamou–Brenier formulation of martingale optimal transport”, *Bernoulli* **25** (2019), no. 4A, pp. 2729–2757.

[20] W. Littman, G. Stampacchia and H. F. Weinberger, “Regular points for elliptic equations with discontinuous coefficients”, *Ann. Sc. Norm. Super. Pisa, Cl. Sci.* **17** (1963), no. 1–2, pp. 43–77.

[21] O. A. Oleinik, “On properties of solutions of certain boundary problems for equations of elliptic type”, *Mat. Sb., N. Ser.* **72** (1952), no. 3, pp. 695–702.

[22] T. Radó, *Subharmonic functions*, Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 5, Springer, 1937.

[23] R. T. Rockafellar, *Convex analysis*, Princeton Mathematical Series, vol. 28, Princeton University Press, 1970.

[24] V. Strassen, “The existence of probability measures with given marginals”, *Ann. Math. Stat.* **36** (1965), pp. 423–439.

[25] X. Tan and N. Touzi, “Optimal transportation under controlled stochastic dynamics”, *Ann. Probab.* **41** (2013), pp. 3201–3240.

[26] D. Trevisan, “Well-posedness of multidimensional diffusion processes with weakly differentiable coefficients”, *Electron. J. Probab.* **21** (2016), article no. 22 (41 pages).

[27] C. Villani, *Optimal transport. Old and new*, Grundlehren der Mathematischen Wissenschaften, vol. 338, Springer, 2009.