HAL
open science

On Optimal Transport Of Matrix-Valued Measures

Yann Brenier, Dmitry Vorotnikov

▶ **To cite this version:**

Yann Brenier, Dmitry Vorotnikov. On Optimal Transport Of Matrix-Valued Measures. SIAM Journal on Mathematical Analysis, 2020, 52 (3), pp.2849-2873. 10.1137/19M1274857 . hal-02389318

HAL Id: hal-02389318

https://hal.science/hal-02389318

Submitted on 2 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ON OPTIMAL TRANSPORT OF MATRIX-VALUED MEASURES

YANN BRENIER AND DMITRY VOROTNIKOV

Abstract. We suggest a new way of defining optimal transport of positive-semidefinite matrix-valued measures. It is inspired by a recent rendering of the incompressible Euler equations and related conservative systems as concave maximization problems. The main object of our attention is the Kantorovich-Bures metric space, which is a matricial analogue of the Wasserstein and Hellinger-Kantorovich metric spaces. We establish some topological, metric and geometric properties of this space.

1. Introduction

Positive-definite-matrix-valued densities arise in signal processing, geometry (Riemannian metrics) and other applications. Due to the success of the Monge-Kantorovich optimal transport theory, there have been recent attempts to introduce the matrix-valued optimal transport in a relevant way [14, 10, 11, 13, 12, 15, 33, 28, 21]. It is reasonable to try to achieve this goal via a dynamical formulation in line with [4]. This requires a kind of transport equation for matricial densities. The listed references either employ the Lindblad equaton [22] and related ideas, or have a static Monge-Kantorovich outlook. Our approach is totally different, and we believe that it is promising for the applications because of its relative simplicity from the numerical perspective.

The first author has recently observed in [7] that the incompressible Euler equation can be recast as concave maximization problem. The method is actually applicable to various conservative PDEs, cf. [7, 39]. The procedure of [7, 39] naturally produces variational problems involving matricial densities. These problems are very similar to the dynamical optimal transport and to the mean-field games, and may serve a heuristic for constructing matricial optimal transport problems. The problem that we introduce in this paper is based on the transport-like operator

$$-(\nabla q)^{Sym} \tag{1.1}$$

that is related to the concave maximization rendering [7] of the incompressible Euler equation. Here q is a suitable momentum-like field. One can generate (1.1) even more straigtforwardly by considering the Burgers-like problem

$$\partial_t v + \operatorname{div}(v \otimes v) = 0. \tag{1.2}$$

2010 *Mathematics Subject Classification.* 28A33, 47A56, 49Q20, 51F99, 58B20.

Key words and phrases. Monge-Kantorovich transport, Bures distance, geodesic space, metric cone.

This is perhaps the most elementary vectorial PDE that fits into the framework of "abstract Euler" equations introduced in [39]. The corresponding concave maximization problem, cf. [39], may be formally written as

$$\sup_{q,B} \int_{[0,T]\times\mathbb{R}^d} v_0 \cdot q - \frac{1}{4} G^{-1} q \cdot q \tag{1.3}$$

where the vector fields q and the positive-definite matrix fields G are subject to the constraints

$$\partial_t G_t = -(\nabla q_t)^{Sym}, \tag{1.4}$$

$$G_T \equiv \frac{1}{2} I_d. \tag{1.5}$$

However, the operator (1.1) that appears in (1.4) has a nontrivial cokernel, which means that we cannot join any two matrix-valued densities with a path directed by the tangents of the form (1.1). This is somewhat similar to the impossibility of joining two measures of different mass in the classical Monge-Kantorovich transport. The latter issue can be fixed in the framework of the unbalanced optimal transport [24, 26, 16, 27, 17, 34] by interpolating between the classical optimal transport and the Hellinger (also known as Fisher-Rao) metric related to the information geometry [3, 29, 23]. The matricial counterpart of the Hellinger metric is the Bures metric [18, 36]. The latter one is usually defined for constant densities but can be naturally generalized to non-constant densities, cf. [15]. Then we can interpolate between this quantum information metric and the matricial transport driven by (1.4). This procedure generates an additional reactive term in the transport equation, and we can join any two positive-definite-matrix-valued measures by a suitable continuous path. The same correction term was recently used in [15] for the Lindblad equation. The resulting dynamical transportation problem generates a distance on the space of positive-definite-matrix-valued measures, which we call the Kantorovich-Bures distance. This distance is a matricial generalization the Wasserstein distance [37] and of the recently introduced Hellinger-Kantorovich distance [24, 26, 16, 27, 17]. The Kantorovich-Bures distance is frame-indifferent in the spirit of rational mechanics [35]. The Kantorovich-Bures space is a geodesic metric space. It has a conic structure comparable to the one that was recently discovered [25] for the Hellinger-Kantorovich space. The Bures space of constant positive-definite matrices may be viewed as a totally geodesic submanifold in the Kantorovich-Bures space.

The paper is organized as follows. In the remaining part of the Introduction, we present basic notation and preliminary facts. In Section 2, we define the Bures-Kantorovich distance using a dynamical variational construction. In Section 3, we explore some topological, metric and geometric properties of the Bures-Kantorovich metric space. In Section 4, we study the metric cone structure of the Bures-Kantorovich space. In the Appendices, we discuss the frame-indifference of the distance and formal Riemannian geometry of the Bures-Kantorovich space, and prove several technical lemmas.

Notation and preliminaries.

- We will use the following basic notation:

- $\mathbb{R}^{d\times d}$ is the space of $d\times d$ matrices, equipped with the Frobenius product $\Phi:\Psi = Tr(\Phi\Psi^\top)$ and the norm $|\Phi| = \sqrt{\Phi:\Phi}$,
- $A^{Sym} := \frac{1}{2}(A + A^\top)$ will denote the symmetric part of $A \in \mathbb{R}^{d\times d}$,
- $\mathcal{S}$ is the subspace of symmetric $d\times d$ matrices,
- $\mathcal{P}^+$ is the subspace of symmetric positive-semidefinite $d\times d$ matrices,
- $\mathcal{P}^{++}$ is the subspace of symmetric positive-definite $d\times d$ matrices,
- $\mathcal{P}^1$ is the subspace of symmetric positive-definite $d\times d$ matrices of unit trace,
- $\mathbb{P}^+$ is the set of $\mathcal{P}^+$-valued Radon measures P on $\mathbb{R}^d$ with finite $Tr(\mathrm{d}P(\mathbb{R}^d))$,
- $\mathbb{P}^{++}$ is the set of absolutely continuous (w.r.t. the Lebesgue measure $\mathcal{L}^d$) $\mathcal{P}^{++}$-valued Radon measures P on $\mathbb{R}^d$ with finite $Tr(\mathrm{d}P(\mathbb{R}^d))$,
- $\mathbb{P}^1$ is the set of $\mathcal{P}^+$-valued Radon measures P on $\mathbb{R}^d$ with $Tr(\mathrm{d}P(\mathbb{R}^d))$=1.

- We will use the following simple inequalities

$$PA : A \le TrP|A|^2,\ Pq\cdot q \le TrP|q|^2,\ P\in\mathcal{P}^+, A\in\mathbb{R}^{d\times d}, q\in\mathbb{R}^d. \tag{1.6}$$

$$A : B \ge 0,\ A, B \in \mathcal{P}^+. \tag{1.7}$$

- We use the following notation for sets of functions:

$\mathcal{C}_b$:	bounded continuous with $\|\phi\|_\infty = \sup|\phi|$;
$\mathcal{C}_b^1$:	bounded $\mathcal{C}^1$ with bounded first derivatives;
$\mathcal{C}_c^\infty$:	smooth compactly supported;
$\mathcal{C}_0$:	continuous and decaying at infinity;
Lip :	bounded and Lipschitz continuous with $\|\phi\|_{\mathrm{Lip}} = \|\nabla\phi\|_\infty + \|\phi\|_\infty$.

- Given a sequence $\{G^k\}_{k\in\mathbb{N}} \subset \mathbb{P}^+$ and $G\in\mathbb{P}^+$ we say that:
  (i) G^k *converges narrowly* to G if there holds

$$\forall\,\phi\in\mathcal{C}_b(\mathbb{R}^d):\qquad \lim_{k\to\infty}\int_{\mathbb{R}^d}\phi(x)\mathrm{d}G^k(x) = \int_{\mathbb{R}^d}\phi(x)\mathrm{d}G(x).$$

  (ii) G^k *converges weakly-*$*$ to G if there holds

$$\forall\,\phi\in\mathcal{C}_0(\mathbb{R}^d):\qquad \lim_{k\to\infty}\int_{\mathbb{R}^d}\phi(x)\mathrm{d}G^k(x) = \int_{\mathbb{R}^d}\phi(x)\mathrm{d}G(x).$$

- Given a measure $G_0\in\mathbb{P}^+$ and a continuous function $F:\mathbb{R}^d\to\mathbb{R}^d$, the measure $F\#G_0$ is the pushforward of G_0 by F, determined by

$$\int_{\mathbb{R}^d}\phi\,\mathrm{d}(F\#G_0) = \int_{\mathbb{R}^d}\phi\circ F\,\mathrm{d}G_0$$

  for all test functions $\phi\in\mathcal{C}_b(\mathbb{R}^d)$.
- For curves $t\in[0,1]\mapsto G_t\in\mathbb{P}^+$ we write $G\in\mathcal{C}_w([0,1];\mathbb{P}^+)$ for the continuity with respect to the narrow topology.
- Given a non-identically-zero measure $G\in\mathbb{P}^+$ we will denote by $L^2(\mathrm{d}G) = L^2(\mathrm{d}G;\mathcal{S}\times\mathbb{R}^d)$ the Hilbert space obtained by completion of the quotient by the seminorm

kernel of the space $\mathcal{C}_b^1(\mathbb{R}^d;\mathcal{S}\times\mathbb{R}^d)$ equipped with the Hilbert seminorm

$$\|\mathfrak{U}\|^2_{L^2(\mathrm{d}G)} = \int_{\mathbb{R}^d} \mathrm{d}G(x)u\cdot u + \int_{\mathbb{R}^d} \mathrm{d}G(x)U(x) : U(x).$$

Here $\mathfrak{U} := (U,u)$ stands for a generic element in $\mathcal{C}_b^1(\mathbb{R}^d;\mathcal{S}\times\mathbb{R}^d)$.

It is not difficult to see that the elements of $L^2(\mathrm{d}G)$ can be rendered as pairs $\mathfrak{U} = (U,u) \in L^2(\mathrm{d}G;\mathcal{S})\times L^2(\mathrm{d}G;\mathbb{R}^d)$, where the latter two spaces are defined in the conventional way (as, for instance, in [20]).

- In a similar fashion, given a narrowly continuous curve $G \in \mathcal{C}_w([0,1];\mathbb{P}^+)$, we can define the space $L^2(0,1;L^2(\mathrm{d}G_t))$. The Hilbert norm in $L^2(0,1;L^2(\mathrm{d}G_t))$ is

$$\|\mathfrak{U}\|^2_{L^2(0,1;L^2(\mathrm{d}G_t))} = \int_0^1 \left(\int_{\mathbb{R}^d} \mathrm{d}G_t(x)u_t(x)\cdot u_t(x) + \int_{\mathbb{R}^d} \mathrm{d}G_t(x)U_t(x) : U_t(x) \right) \mathrm{d}t.$$

- The *bounded-Lipschitz distance* (*BL*) between two matrix measures $G_0, G_1 \in \mathbb{P}^+$ is

$$d_{BL}(G_0,G_1) = \sup_{\|\Phi\|_{\mathrm{Lip}}\le 1} \left| \int_{\mathbb{R}^d} \Phi : (\mathrm{d}G_1 - \mathrm{d}G_0) \right|.$$

  The distance d_{BL} metrizes the narrow convergence on $\mathbb{P}^+$. A sketch of the proof in the case of matrix measures on an interval can be found in [30]. In our situation the claim can still be shown by mimicking the proof strategy for the scalar-valued Radon measures [6, 19]. The key observation [30] is that $\mathcal{S}$-valued bounded continuous functions can be approximated by monotone (in the sense of positive semi-definiteness) sequences of bounded Lipschitz ones. We also point out is that the supremum can be restricted to smooth compactly supported functions. This follows from the tightness of a set consisting of two matricial measures of finite mass.
- By *geodesics* we always mean constant-speed, minimizing metric geodesics.
- C is a generic positive constant.

2. The Kantorovich-Bures distance

The starting point for our considerations is

Definition 2.1 (Kantorovich-Bures distance)**.** *Given two matrix measures $G_0, G_1 \in \mathbb{P}^+$, we define*

$$d^2_{KB}(G_0,G_1) := \inf_{\mathcal{A}(G_0,G_1)} \|\mathfrak{U}\|^2_{L^2(0,1;L^2(\mathrm{d}G_t))}, \tag{2.1}$$

where the admissible set $\mathcal{A}(G_0,G_1)$ consists of all couples $(G_t,\mathfrak{U}_t)_{t\in[0,1]}$, $\mathfrak{U}_t = (U_t,u_t)$, such that

$$\begin{cases} G \in \mathcal{C}_w([0,1];\mathbb{P}^+), \\ G|_{t=0} = G_0; \quad G|_{t=1} = G_1, \\ \mathfrak{U} \in L^2(0,T;L^2(\mathrm{d}G_t)), \\ \partial_t G_t = \{-\nabla(G_t u_t) + G_t U_t\}^{Sym} \quad \textit{in the weak sense, i.e.,} \end{cases}$$

$$\int_{\mathbb{R}^d} \Phi : (\mathrm{d}G_t - \mathrm{d}G_s) - \int_s^t \int_{\mathbb{R}^d} (\mathrm{d}G_\tau : \partial_\tau \Phi_\tau)\mathrm{d}\tau$$
$$= \int_s^t \int_{\mathbb{R}^d} (\mathrm{d}G_\tau u_\tau \cdot \operatorname{div}\Phi_\tau + \mathrm{d}G_\tau U_\tau : \Phi_\tau)\mathrm{d}\tau \quad (2.2)$$

for all test functions $\Phi \in \mathcal{C}_b^1([0,1]\times\mathbb{R}^d;\mathcal{S})$ *and* $t,s \in [0,1]$.

We could have formally started from minimizing a more general Lagrangian, namely,

$$d_{KB}^2(G_0,G_1) := \inf_{\mathcal{B}(G_0,G_1)} \int_0^1 \left(\int_{\mathbb{R}^d} G_t^{-1}(x) q_t(x)\cdot q_t(x) + G_t^{-1}(x) R_t(x) : R_t(x)\,\mathrm{d}x \right)\mathrm{d}t, \quad (2.3)$$

where the admissible set $\mathcal{B}(G_0,G_1)$ consists of tuples (G_t,q_t,R_t), where $G_t(x)\in\mathcal{P}^{++}$, $q_t(x)\in\mathbb{R}^d$ and $R_t(x)\in\mathbb{R}^{d\times d}$, such that

$$\begin{cases} G|_{t=0} = G_0; \quad G|_{t=1} = G_1, \\ \partial_t G_t = \{-\nabla q_t + R_t\}^{Sym}. \end{cases}$$

This complies with (1.3), (1.4) and the discussion in the Introduction. The reactive part is a generalization of the Bures metric, as will be evident in Remark 3.5, see also Remark 2.7. In contrast with (1.3), we opted for dropping the factor 1/4 in the right-hand sides of (2.1) and (2.3) for a purely aesthetic reason, although this factor seems to be rather fundamental. Indeed, keeping it would halve the distance d_{KB}, which is in good agreement with Theorem 6 (ii). It would also eliminate the factor 4 in Theorem 5, Proposition 3.2, Corollary 3.4, etc.

Perturbing an alleged minimizer of (2.3) by adding $(\delta q,\delta R)$ for which

$$L(\delta q,\delta R) := \{-\nabla(\delta q_t) + \delta R_t\}^{Sym} = 0, (\delta q,\delta R)|_{t=0,1} = 0,$$

we see that the minimizer formally satisfies

$$\int_0^1 \left(\int_{\mathbb{R}^d} G_t^{-1}(x) q_t(x)\cdot \delta q_t(x) + G_t^{-1}(x) R_t(x) : \delta R_t(x)\,\mathrm{d}x \right)\mathrm{d}t = 0,$$

for all pertubations $(\delta q,\delta R)$ from $Ker\,L$. This implies that such a minimizer can be written in the form $q = G\operatorname{div}U$, $R = GU$, for some $U_t(x)\in\mathcal{S}$, hence (2.3) yields (2.1) via setting $u := \operatorname{div}U$.

Remark 2.2 (Transport of Hermitian matrices). It seems that our results can be easily extended onto the case of matrix functions with complex entries; we opted for describing the real-valued case just for maintaining a more transparent connection with the classical Monge-Kantorovich transport.

Remark 2.3 (Torus). All the considerations of the paper are valid, mutatis mutandis, if the measures in question are defined on the torus $\mathbb{T}^d$ instead of $\mathbb{R}^d$.

We shall prove shortly that

Theorem 1. *d_{KB} is a distance on $\mathbb{P}^+$.*

We first need a preliminary technical bound:

Lemma 2.4. *Let $G \in \mathcal{C}_w([0,1];\mathbb{P}^+)$ be a narrowly continuous curve, assume that the constraint (2.2) is satisfied for some potential $\mathfrak{U} \in L^2(0,T;L^2(\mathrm{d}G_t))$ with finite energy*

$$E = E[G;\mathfrak{U}] = \|\mathfrak{U}\|^2_{L^2(0,T;L^2(\mathrm{d}G_t))}$$

and let $M := 2(\max\{m_0, m_1\} + E)$ with $m_i = Tr\,\mathrm{d}G_i(\mathbb{R}^d)$. Then the masses are bounded uniformly in time, $m_t = Tr\,\mathrm{d}G_t(\mathbb{R}^d) \le M$ and

$$\forall\, \Phi \in \mathcal{C}^1_b(\mathbb{R}^d;\mathcal{S}): \qquad \left|\int_{\mathbb{R}^d} \Phi : (\mathrm{d}G_t - \mathrm{d}G_s)\right| \le (\|\operatorname{div}\Phi\|_\infty + \|\Phi\|_\infty)\sqrt{ME}|t-s|^{1/2} \tag{2.4}$$

for all $0 \le s \le t \le 1$.

Proof. By narrow continuity of $t \mapsto G_t$ the masses m_t are uniformly bounded and $m = \max_{t\in[0,t]} m_t$ is finite. A Cauchy-Schwarz-like argument applied to the weak constraint (2.2), together with (1.6), imply

$$\begin{aligned}
&\left|\int_{\mathbb{R}^d} \Phi : (\mathrm{d}G_t - \mathrm{d}G_s)\right| \\
&\qquad = \left|\int_s^t \left(\int_{\mathbb{R}^d} \mathrm{d}G_\tau(x) u_\tau(x)\cdot \operatorname{div}\Phi(x) + \int_{\mathbb{R}^d} \mathrm{d}G_\tau(x) U_\tau(x) : \Phi(x)\right)\mathrm{d}\tau\right| \\
&\qquad \le \left(\int_s^t \left(\int_{\mathbb{R}^d} \mathrm{d}G_\tau(x) \operatorname{div}\Phi(x)\cdot \operatorname{div}\Phi(x) + \int_{\mathbb{R}^d} \mathrm{d}G_\tau(x) \Phi(x) : \Phi(x)\right)\mathrm{d}\tau\right)^{1/2} \\
&\qquad\quad \times \left(\int_s^t \left(\int_{\mathbb{R}^d} \mathrm{d}G_\tau(x) u_\tau(x)\cdot u_\tau(x) + \int_{\mathbb{R}^d} \mathrm{d}G_\tau(x) U_\tau(x) : U_\tau(x)\right)\mathrm{d}\tau\right)^{1/2} \\
&\qquad\qquad\qquad\qquad \le (\|\operatorname{div}\Phi\|_\infty + \|\Phi\|_\infty)\sqrt{m}\cdot |t-s|^{1/2} E^{1/2},
\end{aligned}$$

and it is enough to estimate $m \le M = 2(\max\{m_0, m_1\} + E)$ as in our statement. Choosing $\Phi \equiv I$ we obtain from the previous estimate $|m_t - m_s| \le \sqrt{mE}|t-s|^{1/2}$. Let $t_0 \in [0,1]$ be any time when $m_{t_0} = m$: choosing $t = t_0$ and $s = 0$ we immediately get $m \le m_0 + \sqrt{mE}|t_0 - 0|^{1/2} \le m_0 + \sqrt{mE}$, and some elementary algebra bounds $m \le 2(m_0 + E)$. Exchanging the roles of G_0, G_1 we get similarly $m \le 2(m_1 + E)$, and finally $m \le M$. □

Proof of Theorem 1. Let us first show that $d_{KB}(G_0, G_1)$ is always finite for any $G_0, G_1 \in \mathbb{P}^+$. Indeed for any $P_0 \in \mathbb{P}^+$ it is easy to see that $P_t = (1-t)^2 P_0$ and $\mathfrak{U}_t = \left(-\frac{2}{1-t} I, 0\right)$ give a narrowly continuous curve $t \mapsto P_t \in \mathbb{P}^+$ connecting P_0 to zero, and an easy computation shows that this path has finite energy $E = 4Tr(\mathrm{d}P_0(\mathbb{R}^d)) < \infty$ (this curve is actually the geodesic between P_0 and 0, see Corollary 3.4 below). Rescaling time, it is then easy to connect any two measures $G_0, G_1 \in \mathbb{P}^+$ in time $t \in [0,1]$ by first connecting G_0 to 0 in time $t \in [0,1/2]$ and then connecting 0 to G_1 in time $t \in [1/2,1]$ with cost exactly $E = 2(4\mathrm{d}G_0(\mathbb{R}^d) + 4\mathrm{d}G_1(\mathbb{R}^d)) < \infty$.

In order to show that d_{KB} is really a distance, observe first that the symmetry $d_{KB}(G_0,G_1) = d_{KB}(G_1,G_0)$ is obvious by definition.

For the indiscernability, assume that $G_0, G_1 \in \mathbb{P}^+$ are such that $d_{KB}(G_0,G_1) = 0$. Let $\left(G_t^k, \mathfrak{U}_t^k\right)_{t\in[0,1]}$ be any minimizing sequence in (2.1), i.e., $\lim_{k\to\infty} E[G^k;\mathfrak{U}^k] = d_{KB}^2(G_0,G_1) = 0$. By Lemma 2.4 we see that the masses $m_t^k = Tr\, \mathrm{d}G_t^k(\mathbb{R}^d)$ are uniformly bounded, $\sup_{t\in[0,1],k\in\mathbb{N}} m_t^k \leq M$. For any fixed $\Phi \in \mathcal{C}_c^\infty(\mathbb{R}^d;\mathcal{S})$ the fundamental estimate (2.4) gives

$$\left|\int_{\mathbb{R}^d} \Phi : (\mathrm{d}G_1 - \mathrm{d}G_0)\right| \leq (\|\operatorname{div}\Phi\|_\infty + \|\Phi\|_\infty)\sqrt{ME[G^k;\mathfrak{U}^k]}.$$

Since $\lim_{k\to\infty} E[G^k;\mathfrak{U}^k] = 0$ we conclude that $\int_{\mathbb{R}^d} \Phi : (\mathrm{d}G_1 - \mathrm{d}G_0)$ for all $\Phi \in \mathcal{C}_c^\infty(\mathbb{R}^d;\mathcal{S})$, thus $G_1 = G_0$ as desired.

As for the triangular inequality, fix any $G_0, G_1, P \in \mathbb{P}^+$ and let us prove that $d_{KB}(G_0,G_1) \leq d_{KB}(G_0,P) + d_{KB}(P,G_1)$. We can assume that all three distances are nonzero, otherwise the triangular inequality trivially holds by the previous point. Let now $(\underline{G}_t^k, \underline{\mathfrak{U}}_t^k)_{t\in[0,1]}$ be a minimizing sequence in the definition of $d_{KB}^2(G_0,P) = \lim_{k\to\infty} E[\underline{G}^k;\underline{\mathfrak{U}}^k]$, and let similarly $(\overline{G}_t^k, \overline{\mathfrak{U}}_t^k)_{t\in[0,1]}$ be such that $d_{KB}^2(P,G_1) = \lim_{k\to\infty} E[\overline{G}^k;\overline{\mathfrak{U}}^k]$. For fixed $\tau \in (0,1)$ let $(G_t, \mathfrak{U}_t)$ be the continuous path obtained by first following $\left(\underline{G}^k, \frac{1}{\tau}\underline{\mathfrak{U}}^k\right)$ from G_0 to P in time τ, and then following $\left(\overline{G}^k, \frac{1}{1-\tau}\overline{\mathfrak{U}}^k\right)$ from P to G_1 in time $1-\tau$. Then $\left(G_t^k, \mathfrak{U}_t^k\right)_{t\in[0,1]}$ is an admissible path connecting G_0 to G_1, hence by definition of our distance and the explicit time scaling we get that

$$d_{KB}^2(G_0,G_1) \leq E[G^k;\mathfrak{U}^k] = \frac{1}{\tau}E[\overline{G}^k;\overline{\mathfrak{U}}^k] + \frac{1}{1-\tau}E[\underline{G}^k;\underline{\mathfrak{U}}^k].$$

Letting $k \to \infty$ we obtain for any fixed $\tau \in (0,1)$

$$d_{KB}^2(G_0,G_1) \leq \frac{1}{\tau}d_{KB}^2(G_0,P) + \frac{1}{1-\tau}d_{KB}^2(P,G_1).$$

Finally choosing $\tau = \frac{d_{KB}(G_0,P)}{d_{KB}(G_0,P)+d_{KB}(P,G_1)} \in (0,1)$ yields

$$d_{KB}^2(G_0,G_1) \leq \frac{1}{\tau}d_{KB}^2(G_0,P) + \frac{1}{1-\tau}d_{KB}^2(P,G_1) = (d_{KB}(G_0,P) + d_{KB}(P,G_1))^2$$

and the proof is complete. □

Corollary 2.5. *The elements of a bounded set in* $(\mathbb{P}^+, d_{KB})$ *have uniformly bounded mass. Conversely, subsets of* $\mathbb{P}^+$ *with uniformly bounded mass are bounded in* $(\mathbb{P}^+, d_{KB})$.

Proof. The first statement is an immediate consequence of Lemma 2.5. The converse one follows from the observation that the squared distance from any element $P_0 \in \mathbb{P}^+$ to zero is controlled by $4m(P_0)$, see the proof of Theorem 1. □

Another simple property is

Lemma 2.6. *If $(G_t, \mathfrak{U}_t)_{t\in[0,1]}$ is a narrowly continuous curve with total energy E then $t \mapsto G_t$ is 1/2-Hölder continuous w.r.t. d_{KB}, and more precisely*

$$\forall\, t_0, t_1 \in [0,1]: \qquad d_{KB}(G_{t_0}, G_{t_1}) \le \sqrt{E}|t_0 - t_1|^{1/2}.$$

Proof. Rescaling in time and connecting G_{t_0} to G_{t_1} by the path $(G_s, (t_1 - t_0)\mathfrak{U}_s)_{s\in[0,1]}$ with $t = t_0 + (t_1 - t_0)s$, the resulting energy scales as $d^2_{KB}(G_{t_0}, G_{t_1}) \le E[\overline{G}; \overline{\mathfrak{U}}] \le E|t_0 - t_1|$. □

Remark 2.7. In Definition 2.1 it is possible to restrict ourselves to the admissible paths which satisfy the additional constraint $u \equiv 0$. This leads to another distance d_H on $\mathbb{P}^+$, which is a matricial analogue of the Hellinger distance. All the results of this paper remain true for d_H, and the proofs are literally the same. However, this distance is much stronger than d_{KB}, which might be less relevant in applications. For example, $d_{KB}(\delta_{x_k} I, \delta_0 I) \to 0$ as $x_k \to 0$, but $d_H(\delta_{x_k} I, \delta_0 I) = 2\sqrt{2}$.

3. Properties of the distance and existence of geodesics

We begin with some topological properties of the Kantorovich-Bures space.

Theorem 2 (Comparison with narrow converence)**.** *The convergence of matrix measures w.r.t. the distance d_{KB} implies narrow convergence, and any Cauchy sequence in $(\mathbb{P}^+, d_{KB})$ is Cauchy in $(\mathbb{P}^+, d_{BL})$. Moreover, for any pair $G_0, G_1 \in \mathbb{P}^+$ with masses m_0, m_1 there holds*

$$d_{BL}(G_0, G_1) \le C_d \sqrt{(m_0 + m_1)}\, d_{KB}(G_0, G_1) \tag{3.1}$$

with some uniform C_d depending only on the dimension.

Proof. Fix G_0, G_1, and let $(G_t, \mathfrak{U}_t)$ be any admissible path from G_0 to G_1 with finite energy E. Taking the supremum over Φ with $\|\Phi\|_{\mathrm{Lip}} \le 1$ in (2.4) we get $d_{BL}(G_0, G_1) \le C\sqrt{ME}$, where $M = 2(\max\{m_0, m_1\} + E)$ as in Lemma 2.4. Choosing now a minimizing sequence instead of an arbitrary path and taking the limit we essentially obtain the same estimate with $E = \lim E[G^k; \mathfrak{U}^k] = d^2_{KB}(G_0, G_1)$, whence

$$d_{BL}(G_0, G_1) \le C\sqrt{2(\max\{m_0, m_1\} + d^2_{KB}(G_0, G_1))}\, d_{KB}(G_0, G_1).$$

By the triangle inequality and Corollary 3.4 we control $d^2_{KB}(G_0, G_1) \le 2(d^2_{KB}(G_0, 0) + d^2_{KB}(0, G_1)) = 8(m_0 + m_1)$, which immediately yields (3.1).

If G^k is a Cauchy sequence in $(\mathbb{P}^+, d_{KB})$ with mass $m^k = Tr\, \mathrm{d}G^k(\mathbb{R}^d)$. Since Cauchy sequences are bounded we control $4m^k = d^2_{KB}(G^k, 0) \le C$ uniformly in k, thus from (3.1) we see that

$$d_{BL}(G^p, G^q) \le C d_{KB}(G^p, G^q).$$

Thus, G^k is d_{BL}-Cauchy. Similarly, if a sequence is d_{KB}-converging, it is d_{BL}- and hence narrowly converging (to the same limit). □

Definition 3.1. *Let (X,ρ) be a metric space, σ be a Hausdorff topology on X. We say that the distance ρ is sequentially lower semicontinuous with respect to σ if for all σ-converging sequences $x_k \xrightarrow{\sigma} x$, $y_k \xrightarrow{\sigma} y$ one has*

$$\rho(x,y) \le \liminf_{k\to\infty} \rho(x_k,y_k).$$

Theorem 3 (Lower-semicontinuity)**.** *The distance d_{KB} is sequentially lower semicontinuous with respect to the weak-$*$ topology on $\mathbb{P}^+$.*

Proof. Consider any two converging sequences

$$G_0^k \underset{k\to\infty}{\to} G_0, \qquad G_1^k \underset{k\to\infty}{\to} G_1 \qquad \text{weakly-}*$$

of finite Radon measures from $\mathbb{P}^+(\mathbb{R}^d)$. For each k, the endpoints G_0^k and G_1^k can be joined by an admissible narrowly continuous path $(G_t^k, \mathfrak{U}_t^k)_{t\in[0,1]}$ with energy

$$E[G^k;\mathfrak{U}^k] \le d_{KB}^2(G_0^k,G_1^k) + k^{-1}.$$

Due to weak-$*$ compactness, the masses $m_0^k = Tr\, dG_0^k(\mathbb{R}^d)$ and $m_1^k = Tr\, dG_1^k(\mathbb{R}^d)$ are bounded uniformly in $k\in\mathbb{N}$. By Corollary 2.5 the set $\cup_{k\in\mathbb{N}}\{G_0^k,G_1^k\}$ is bounded in $(\mathbb{P}^+,d_{KB})$, thus the energies $E[G^k;\mathfrak{U}^k]$ and the masses $m_t^k = Tr\,\mathrm{d}G_t^k(\mathbb{R}^d)$ are bounded uniformly in $k\in\mathbb{N}$ and $t\in[0,1]$

$$m_t^k \le M \quad \text{and} \quad E[G^k;\mathfrak{U}^k] \le \overline{E}.$$

By the (classical) Banach-Alaoglu theorem with $\mathbb{P}^+ \subset (\mathcal{C}_0)^*$, all the curves $(G_t^k)_{t\in[0,1]}$ lie in a fixed weak-$*$ sequentially relatively compact set $\mathcal{K}_M = \{G\in\mathbb{P}^+ : Tr\, dG(\mathbb{R}^d) \le M\}$ uniformly in k,t. By the fundamental estimate (2.4) we get

$$\left|\int_{\mathbb{R}^d} \Phi : (\mathrm{d}G_t^k - \mathrm{d}G_s^k)\right| \le \sqrt{M\overline{E}}|t-s|^{1/2}(\|\operatorname{div}\Phi\|_\infty + \|\Phi\|_\infty) \le C|t-s|^{1/2}(\|\nabla\Phi\|_\infty + \|\Phi\|_\infty)$$

for all $\phi\in\mathcal{C}_b^1$, which implies

$$\forall\, t,s\in[0,1], \forall k\in\mathbb{N}: \qquad d_{BL}(G_s^k,G_t^k) \le C|t-s|^{1/2}.$$

Invoking the above uniform 1/2-Hölder continuity w.r.t. d_{BL}, the sequential lower semicontinuity of d_{BL} with respect to the weak-$*$ convergence (Lemma B.2 in Appendix B), and the fact that $G_t^k\in\mathcal{K}_M$, we conclude by a refined version of Arzelà-Ascoli theorem (Lemma B.4 in the Appendix B) that there exists a d_{BL} (thus narrowly) continuous curve $(G_t)_{t\in[0,1]}$ connecting G_0 and G_1 such that

$$\forall t\in[0,1]: \qquad G_t^k \to G_t \quad \text{weakly-}* \tag{3.2}$$

along some subsequence $k\to\infty$ (not relabeled here). Let $Q := (0,1)\times\mathbb{R}^d$ and μ^k be the matricial measure on Q defined by duality as

$$\forall\,\phi\in\mathcal{C}_c(Q): \qquad \int_Q \phi(t,x)\mathrm{d}\mu^k(t,x) = \int_0^1\left(\int_{\mathbb{R}^d}\phi(t,.)\mathrm{d}G_t^k\right)\mathrm{d}t.$$

Exploiting the pointwise convergence (3.2) and the uniform bound on the masses $m_t^k \le M$, a simple application of Lebesgue's dominated convergence guarantees that

$$\mu^k \to \mu^0 \qquad \text{weakly-}* \text{ in } \mathbb{P}^+(Q),$$

where the finite measure $\mu^0 \in \mathbb{P}^+(Q)$ is defined by duality in terms of the weak-$*$ limit $G_t = \lim G_t^k$ (as was μ^k in terms of G_t^k).

We are going to apply a variant of the Banach-Alaoglu theorem, Proposition B.1 in Appendix B, in the space

$$X = \mathcal{C}_c^1(Q; \mathcal{S} \times \mathbb{R}^d).$$

Namely, we set

$$\|(\Phi, \phi)\| = \|\phi\|_{L^\infty(Q)} + \|\Phi\|_{L^\infty(Q)},$$

$$\|(\Phi, \phi)\|_k = \left(\int_Q \mathrm{d}\mu^k \phi \cdot \phi + \mathrm{d}\mu^k \Phi : \Phi \right)^{1/2}, \qquad k = 0, 1, \ldots,$$

and define the linear forms

$$\varphi_k(\Phi, \phi) = \int_Q \mathrm{d}\mu^k u^k \cdot \phi + \mathrm{d}\mu^k U^k : \Phi, \qquad k = 1, 2, \ldots.$$

The separability of $\mathcal{C}_c^1(Q; \mathcal{S} \times \mathbb{R}^d)$, the weak-$*$ convergence of μ^k, uniform boundedness of the masses of $Tr\, \mu^k(Q) \le M$, and the Cauchy-Schwarz inequality imply that the hypotheses of our Proposition B.1 are met with

$$c_k := \|\varphi_k\|_{(X, \|\cdot\|_k)^*} \le \|\mathfrak{U}^k\|_{L^2(0,1; L^2(\mathrm{d}G^k))} = \sqrt{E[G^k; \mathfrak{U}^k]} \le \sqrt{d_{KB}^2(G_0^k, G_1^k) + k^{-1}}.$$

Consequently, there exists a continuous functional φ_0 on the space $(X, \|\cdot\|_0)$ such that up to a subsequence

$$\forall (\Phi, \phi) \in \mathcal{C}_c^1(Q; \mathcal{S} \times \mathbb{R}^d): \qquad \int_0^1 \left(\int_{\mathbb{R}^d} \mathrm{d}G_t^k u_t^k \cdot \phi_t + \mathrm{d}G_t^k U_t^k : \Phi_t \right) \mathrm{d}t \underset{k \to \infty}{\to} \varphi_0(\Phi)$$

with moreover

$$\|\varphi_0\|_{(X, \|\cdot\|_0)^*} \le \liminf_{k \to \infty} d_{KB}(G_0^k, G_1^k). \tag{3.3}$$

Let $N_0 \subset X$ be the kernel of the seminorm $\|\cdot\|_0$. By the Riesz representation theorem, the dual $(X, \|\cdot\|_0)^* = (X/N_0, \|\cdot\|_0)^*$ can be isometrically identified with the completion $\overline{X/N_0}$ of X/N_0 with respect to $\|\cdot\|_0$, which is exactly $L^2(0,1; L^2(\mathrm{d}G_t))$. As a consequence there exists $\mathfrak{U} = (U, u) \in L^2(0, T; L^2(\mathrm{d}G_t))$ such that

$$\varphi_0(\Phi, \phi) = \int_Q \mathrm{d}\mu^0 u \cdot \phi + \mathrm{d}\mu^0 U : \Phi = \int_0^1 \left(\int_{\mathbb{R}^d} \mathrm{d}G_t u_t \cdot \phi_t + \mathrm{d}G_t U_t : \Phi_t \right) \mathrm{d}t$$

and

$$\|\mathfrak{U}\|_{L^2(0,1; L^2(\mathrm{d}G_t))} = \|\varphi_0\|_{(X, \|\cdot\|_0)^*},$$

and it is straightforward to check that $(G, \mathfrak{U})$ is an admissible curve joining G^0, G^1, because the above convergence is enough to pass to the limit in the constraint (2.2). Recalling (3.3), it remains to take into account that by the definition of our distance

$$d^2_{KB}(G_0, G_1) \le E[G; \mathfrak{U}] = \|\mathfrak{U}\|^2_{L^2(0,1;L^2(\mathrm{d}G_t))} = \|\varphi_0\|^2_{(X,\|\cdot\|_0)^*} \le \liminf_{k\to\infty} d^2_{KB}(G^k_0, G^k_1).$$

□

During the proof of Theorem 2 we observed the upper bound

$$d^2_{KB}(G_0, G_1) \le 8(m_0 + m_1).$$

Let us show that it can improved.

Proposition 3.2 (Upper bound of the distance)**.** *For every pair $G_0, G_1 \in \mathbb{P}^+$ with masses m_0, m_1 one has*

$$d^2_{KB}(G_0, G_1) \le 4(m_0 + m_1). \tag{3.4}$$

Proof. Since $\mathbb{P}^{++}$ is dense in $\mathbb{P}^+$ in the weak-$*$ topology (one can simply use the standard mollifiers), in view of Theorem 3 we can assume that $G_0, G_1 \in \mathbb{P}^{++}$. Consider the curve

$$\mathrm{d}G_t = \left(t\sqrt{G_1} + (1-t)\sqrt{G_0}\right)^2 \mathrm{d}\mathcal{L}^d.$$

The corresponding potential $\mathfrak{U}_t \in L^2(0,1;L^2(\mathrm{d}G))$ can be defined by the by Riesz duality as

$$\langle \mathfrak{U}, (\Phi, \phi)\rangle_{L^2(0,1;L^2(\mathrm{d}G))} = 2\int_0^1 \int_{\mathbb{R}^d} (\sqrt{G_1} - \sqrt{G_0}) : \left(t\sqrt{G_1} + (1-t)\sqrt{G_0}\right)\Phi_t \,\mathrm{d}\mathcal{L}^d \,\mathrm{d}t \tag{3.5}$$

for all $(\Phi, \phi) \in L^2(0,1;L^2(\mathrm{d}G))$. It is not difficult to see that the constraint (2.2) is satisfied. By definition, the energy of this path is $\|\mathfrak{U}\|^2_{L^2(0,1;L^2(\mathrm{d}G))}$. By the Cauchy-Schwarz inequality and (1.7),

$$\begin{aligned}
&\langle \mathfrak{U}, (\Phi, \phi)\rangle^2_{L^2(0,1;L^2(\mathrm{d}G))} \\
&\qquad \le 4\int_0^1 \int_{\mathbb{R}^d} (\sqrt{G_1} - \sqrt{G_0})\,\mathrm{d}\mathcal{L}^d \int_{\mathbb{R}^d} \mathrm{d}\left(t\sqrt{G_1} + (1-t)\sqrt{G_0}\right)\Phi_t \\
&\qquad\qquad : \left(t\sqrt{G_1} + (1-t)\sqrt{G_0}\right)\Phi_t \,\mathrm{d}\mathcal{L}^d \,\mathrm{d}t \\
&\qquad \le 4\|\Phi, 0\|^2_{L^2(0,1;L^2(\mathrm{d}G))} \int_0^1 \int_{\mathbb{R}^d} (\sqrt{G_1} - \sqrt{G_0}) : (\sqrt{G_1} - \sqrt{G_0})\,\mathrm{d}\mathcal{L}^d \,\mathrm{d}t \\
&\qquad\qquad \le 4\|\Phi, \phi\|^2_{L^2(0,1;L^2(\mathrm{d}G))} \int_0^1 \int_{\mathbb{R}^d} (Tr\, G_1 + Tr\, G_0)\,\mathrm{d}\mathcal{L}^d \,\mathrm{d}t.
\end{aligned}$$

Thus, the energy $E[G_t, \mathfrak{U}_t]$ is less than or equal to the right-hand side of (3.4). □

Definition 3.3 (cf. [9])**.** *We say that two points x, y in a metric space (X, ρ) almost admit a midpoint if there exists a sequence $\{z_k\} \subset X$ such that*

$$|\rho(x,y) - 2\rho(x,z_k)| \le k^{-1}, \quad |\rho(x,y) - 2\rho(y,z_k)| \le k^{-1}.$$

Theorem 4 (Existence of geodesics)**.** *$(\mathbb{P}^+, d_{KB})$ is a geodesic space, and for all $G_0, G_1 \in \mathbb{P}^+$ the infimum in (2.1) is always a minimum. Moreover this minimum is attained for a d_{KB}-Lipschitz curve G such that $d_{KB}(G_t, G_s) = |t-s| d_{KB}(G_0, G_1)$ and a potential $\mathfrak{U} \in L^2(0,1; L^2(\mathrm{d}G_t))$ such that $\|\mathfrak{U}_t\|_{L^2(\mathrm{d}G_t)} = cst = d_{KB}(G_0, G_1)$ for a.e. $t \in [0,1]$.*

Proof. We first observe from the definition of our distance that any two points in $\mathbb{P}^+$ almost admit a midpoint. By Corollary 2.5 and the (classical) Banach-Alaoglu theorem, d_{KB}-bounded sequences contain weakly-$*$ converging subsequences. Now Lemma B.3 (analogue of the Hopf-Rinow theorem for non-complete metric spaces) together with Theorem 3 imply that $(\mathbb{P}^+, d_{KB})$ is a geodesic space. The existence and claimed properties of a minimizing admissible path in (2.1) follow by mimicking the argument from the proof of Theorem 3 for the sequence of almost minimizing paths, and by evoking the general properties of metric geodesics [2, 9]. □

The next theorem gives some insight into the geometry of the Kantorovich-Bures space.

Theorem 5 (Explicit geodesics)**.** *Fix any element $G_* \in \mathbb{P}^+$ and define the map $g : \mathcal{P}^+ \to \mathbb{P}^+$ by*

$$g(A) = A G_* A.$$

Then for any pair of commuting matrices $A_0, A_1 \in \mathcal{P}^+$ one has

$$d_{KB}^2(g(A_0), g(A_1)) = 4 \int_{\mathbb{R}^d} \mathrm{d}G_*(A_1 - A_0) : (A_1 - A_0), \tag{3.6}$$

and a geodesic between $g(A_0)$ and $g(A_1)$ is explicitly given by

$$\bar{G}_t := g(tA_1 + (1-t)A_0). \tag{3.7}$$

Proof. Step 1. Define a potential $\bar{\mathfrak{U}}_t \in L^2(0,1; L^2(\mathrm{d}\bar{G}))$ by Riesz duality as

$$\langle \bar{\mathfrak{U}}, (\Phi, \phi) \rangle_{L^2(0,1;L^2(\mathrm{d}\bar{G}))} = 2 \int_0^1 \int_{\mathbb{R}^d} \mathrm{d}G_*(A_1 - A_0) : (tA_1 + (1-t)A_0)\Phi_t \, \mathrm{d}t \tag{3.8}$$

for all $(\Phi, \phi) \in L^2(0,1; L^2(\mathrm{d}\bar{G}))$. A straightforward computation shows that $(\bar{G}_t, \bar{\mathfrak{U}}_t)$ satisfies the constraint (2.2). The energy of this path coincides with $\|\bar{\mathfrak{U}}\|^2_{L^2(0,1;L^2(\mathrm{d}\bar{G}))}$. By the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\langle \bar{\mathfrak{U}}, (\Phi, \phi) \rangle^2_{L^2(0,1;L^2(\mathrm{d}\bar{G}))} \\
&\quad \le 4 \int_0^1 \int_{\mathbb{R}^d} \mathrm{d}G_*(A_1 - A_0) : (A_1 - A_0) \int_{\mathbb{R}^d} \mathrm{d}G_*(tA_1 + (1-t)A_0)\Phi_t : (tA_1 + (1-t)A_0)\Phi_t \, \mathrm{d}t \\
&\qquad\qquad \le 4 \|\Phi, \phi\|^2_{L^2(0,1;L^2(\mathrm{d}\bar{G}))} \int_0^1 \int_{\mathbb{R}^d} \mathrm{d}G_*(A_1 - A_0) : (A_1 - A_0) \, \mathrm{d}t.
\end{aligned}$$

Thus, the energy $E[\bar{G}_t, \bar{\mathfrak{U}}_t]$ is less than or equal to the right-hand side of (3.6).

Step 2. In view of the previous step, it suffices to prove that the square of the distance is bounded from below by the right-hand side of (3.6). We first observe that without loss

of generality we may assume that $A_0 \in \mathcal{P}^{++}$. Indeed, the general case $A_0 \in \mathcal{P}^{+}$ would immediately follow by letting $\epsilon \to 0_+$ in the triangle inequality

$$d_{KB}(g(A_0), g(A_1)) \geq d_{KB}(g(A_0 + \epsilon I), g(A_1)) - d_{KB}(g(A_0 + \epsilon I), g(A_0)).$$

Step 3. Consider any admissible path $(G_t, \mathfrak{U}_t)_{t\in[0,1]}$ connecting $G_0 := g(A_0)$ to $G_1 := g(A_1)$. Let λ be any scalar probability measure on $\mathbb{R}^d$. Set $\tilde{G}_t := \lambda \int_{\mathbb{R}^d} \mathrm{d}G_t$, and define $\tilde{\mathfrak{U}}_t \in L^2(0,1; L^2(\mathrm{d}\tilde{G}))$ by duality as

$$\langle \tilde{\mathfrak{U}}, (\Phi, \phi) \rangle_{L^2(0,1;L^2(\mathrm{d}\tilde{G}))} = \int_0^1 \int_{\mathbb{R}^d} \mathrm{d}G_t U_t \int_{\mathbb{R}^d} \Phi_t \mathrm{d}\lambda \, \mathrm{d}t \tag{3.9}$$

for all $(\Phi, \phi) \in L^2(0,1; L^2(\mathrm{d}\tilde{G}))$. Then$(\tilde{G}, \tilde{\mathfrak{U}})$ is an admissible path (joining $A_0 \lambda \int_{\mathbb{R}^d} \mathrm{d}G_* A_0$ and $A_1 \lambda \int_{\mathbb{R}^d} \mathrm{d}G_* A_1$). We claim that it has lesser energy than $(G, \mathfrak{U})$. To prove the claim, we approximate this path with the sequence $\tilde{G}_t^k := \lambda \int_{\mathbb{R}^d} (k^{-1} I + \mathrm{d}G_t)$; the corresponding potentials are

$$\langle \tilde{\mathfrak{U}}^k, (\Phi, \phi) \rangle_{L^2(0,1;L^2(\mathrm{d}\tilde{G}^k))} = \int_0^1 \int_{\mathbb{R}^d} \mathrm{d}G_t U_t \int_{\mathbb{R}^d} \Phi_t \mathrm{d}\lambda \, \mathrm{d}t. \tag{3.10}$$

Let us equip the linear space $\mathbb{R}^{d\times d}$ with the scalar product

$$(B, B)_{k,t} = k^{-1} B : B + \int_{\mathbb{R}^d} \mathrm{d}G_t B : B,$$

and let $\Pi_{k,t}$ be the orthogonal projection onto the subspace $\mathcal{S}$. One explicitly computes that

$$\tilde{U}_t^k := \Pi_{k,t}\left(\left(k^{-1} I + \int_{\mathbb{R}^d} \mathrm{d}G_t \right)^{-1} \int_{\mathbb{R}^d} \mathrm{d}G_t U_t \right), \qquad u_t^k \equiv 0.$$

Then

$$
\begin{aligned}
E[\tilde{G}^k;\tilde{\mathfrak{U}}^k] &= \int_0^1 \int_{\mathbb{R}^d} \mathrm{d}\lambda \left(k^{-1}I + \int_{\mathbb{R}^d} \mathrm{d}G_t\right) \tilde{U}_t^k : \tilde{U}_t^k \,\mathrm{d}t = \int_0^1 (\tilde{U}_t^k, \tilde{U}_t^k)_{k,t} \,\mathrm{d}t \\
&\le \int_0^1 \left(\left(k^{-1}I + \int_{\mathbb{R}^d} \mathrm{d}G_t\right)^{-1} \int_{\mathbb{R}^d} \mathrm{d}G_t U_t, \left(k^{-1}I + \int_{\mathbb{R}^d} \mathrm{d}G_t\right)^{-1} \int_{\mathbb{R}^d} \mathrm{d}G_t U_t \right)_{k,t} \mathrm{d}t \\
&\le \int_0^1 \left(k^{-1}I + \int_{\mathbb{R}^d} \mathrm{d}G_t\right) \left(\left(k^{-1}I + \int_{\mathbb{R}^d} \mathrm{d}G_t\right)^{-1} \int_{\mathbb{R}^d} \mathrm{d}G_t U_t \right) \\
&\quad : \left(\left(k^{-1}I + \int_{\mathbb{R}^d} \mathrm{d}G_t\right)^{-1} \int_{\mathbb{R}^d} \mathrm{d}G_t U_t \right) \mathrm{d}t \\
&\quad + \int_0^1 \int_{\mathbb{R}^d} \mathrm{d}G_t \left(U_t - \left(k^{-1}I + \int_{\mathbb{R}^d} \mathrm{d}G_t\right)^{-1} \int_{\mathbb{R}^d} \mathrm{d}G_t U_t \right) \\
&\quad : \left(U_t - \left(k^{-1}I + \int_{\mathbb{R}^d} \mathrm{d}G_t\right)^{-1} \int_{\mathbb{R}^d} \mathrm{d}G_t U_t \right) \mathrm{d}t \\
&= \int_0^1 \int_{\mathbb{R}^d} \mathrm{d}G_t U_t : U_t \,\mathrm{d}t \\
&\quad - k^{-1} \int_0^1 \left(\left(k^{-1}I + \int_{\mathbb{R}^d} \mathrm{d}G_t\right)^{-1} \int_{\mathbb{R}^d} \mathrm{d}G_t U_t \right) : \left(\left(k^{-1}I + \int_{\mathbb{R}^d} \mathrm{d}G_t\right)^{-1} \int_{\mathbb{R}^d} \mathrm{d}G_t U_t \right) \mathrm{d}t \\
&\le E[G;\mathfrak{U}].
\end{aligned}
$$

Arguing as in the proof of Theorem 3 we can pass to the limit inferior as $k \to \infty$ to show that $E[\tilde{G}_t, \tilde{\mathfrak{U}}_t] \le E[G;\mathfrak{U}]$ as claimed.

Step 4. Obviously, the right-hand side of (3.6) does not change if we replace G_* by λD, where $D := \int_{\mathbb{R}^d} \mathrm{d}G_* \in \mathcal{P}^+$, and λ is as above. Thus, by the previous steps it is enough to check that the energies of the admissible paths of the form $G_t = \lambda F_t$ with $F_t \in \mathcal{P}^+$, $F_i = A_i D A_i$, $i = 0, 1$, $A_0 \in \mathcal{P}^{++}$, with constant-in-space potentials $\mathfrak{U}_t = (U_t, 0) \in L^2(0,1;L^2(\mathrm{d}G))$ are bounded from below by the right-hand side of (3.6). Some finite-dimensional calculus of variations shows that the minimum of those energies is achieved for $F_t = (tA_1 + (1-t)A_0)D(tA_1 + (1-t)A_0)$ with $U_t = 2(tA_1 + (1-t)A_0))^{-1}(A_1 - A_0)$, $t \ne 1$. The corresponding energy is exactly $4D(A_1 - A_0) : (A_1 - A_0)$. □

Corollary 3.4 (Geodesic to zero). *For any $G_* \in \mathbb{P}^+$, $d_{KB}^2(G_*, 0) = 4\,Tr\,\mathrm{d}G_*(\mathbb{R}^d)$, and $(1-t)^2 G_*$ is a geodesic between G_* and* 0.

Remark 3.5 (Bures manifolds). The set $\mathcal{P}^{++} \subset \mathcal{S}$ has a natural structure of a smooth manifold, and the tangent space $T_P\mathcal{P}^{++}$ at every point $P \in \mathcal{P}^{++}$ can be identified with $\mathcal{S}$. For each $\Xi \in T_P\mathcal{P}^{++}$, let $U_\Xi \in \mathcal{S}$ be the unique solution to the Lyapunov equation [5]

$$2\Xi = PU_\Xi + U_\Xi P.$$

Then

$$\langle \Xi_1, \Xi_2 \rangle_P := P U_{\Xi_1} : U_{\Xi_2} \tag{3.11}$$

is a Riemannian metric on $\mathcal{P}^{++}$ that is known as the Bures metric [18]. The induced Riemannian metric on the submanifold $\mathcal{P}^1 \subset \mathcal{P}^{++}$ is also called the Bures metric [36]. Actually, $\mathcal{P}^{++}$ is a metric cone over $\mathcal{P}^1$. In the next section we will see that $\mathbb{P}^+$ has a similar cone structure. The geodesics between $P_0, P_1 \in \mathcal{P}^{++}$ can be constructed as follows. Let $X \in \mathcal{P}^{++}$ be the unique solution to the Riccati equation [5]

$$X P_0 X = P_1.$$

Then the geodesic is

$$((1-t)I + tX) P_0 ((1-t)I + tX).$$

Let d_R denote the corresponding Riemannian distance on $\mathcal{P}^{++}$. Fix any probability measure λ on $\mathbb{R}^d$. Since I and X commute, by Theorem 5 the embedding

$$P \mapsto P\lambda$$

from $\mathcal{P}^{++}$ into $\mathbb{P}^+$ is a totally geodesic map (in the sense of [31]). Moreover, in view of Remark 2.7,

$$d_R(P_0, P_1) = d_{KB}(P_0\lambda, P_1\lambda) = d_H(P_0\lambda, P_1\lambda).$$

The midpoint $\frac{1}{4}(I+X)P_0(I+X)$ can serve to define a matrix mean $P_0 \hat{} P_1$, which may be dubbed the Bures mean. More conventional matrix means are discussed in [5].

4. The spherical distance and the conic structure

In this section we are going to explore the conic structure of $(\mathbb{P}^+, d_{KB})$. We start by defining a similar distance on $\mathbb{P}^1$ (analogue of probability measures) by a straightforward trick:

Definition 4.1 (Spherical Kantorovich-Bures distance)**.** *Given two matrix measures $G_0, G_1 \in \mathbb{P}^1$ we define*

$$d^2_{SKB}(G_0, G_1) := \inf_{\mathcal{A}_1(G_0,G_1)} \int_0^1 \left(\int_{\mathbb{R}^d} \mathrm{d}G_t(x) u_t(x) \cdot u_t(x) + \int_{\mathbb{R}^d} \mathrm{d}G_t(x) U_t(x) : U_t(x) \right) \mathrm{d}t. \tag{4.1}$$

where the admissible set $\mathcal{A}_1(G_0, G_1)$ consists of all couples $(G_t, \mathfrak{U}_t)_{t\in[0,1]}$ such that

$$\begin{cases} G \in \mathcal{C}_w([0,1]; \mathbb{P}^1), \\ G|_{t=0} = G_0; \quad G|_{t=1} = G_1, \\ \mathfrak{U} \in L^2(0,T; L^2(\mathrm{d}G_t)), \\ \partial_t G_t = \{-\nabla(G_t u_t) + G_t U_t\}^{Sym} \quad \textit{in the weak sense.} \end{cases}$$

Proposition 4.2. *d_{SKB} is a distance on $\mathbb{P}^1$.*

The proof is similar to the one of Theorem 1. Note that the indiscernability is obvious since by construction $d_{SKB} \geq d_{KB}$ on $\mathbb{P}^1$.

Remark 4.3 (Equivalent definition). It is easy to see that Definition 4.1 can be equivalently written in the following way: given two matrix measures $G_0, G_1 \in \mathbb{P}^1$ we define

$$d^2_{SKB}(G_0,G_1) := \inf_{\mathcal{A}_2(G_0,G_1)} \int_0^1 \left(\int_{\mathbb{R}^d} \mathrm{d}G_t(x) u_t(x)\cdot u_t(x) + \int_{\mathbb{R}^d} \mathrm{d}G_t(x) U_t(x) : U_t(x) - \left(\int_{\mathbb{R}^d} \mathrm{d}G_t(x) : U_t(x) \right)^2 \right) \mathrm{d}t. \quad (4.2)$$

where the admissible set $\mathcal{A}_2(G_0,G_1)$ consists of all couples $(G_t, \mathfrak{U}_t)_{t\in[0,1]}$ such that

$$\begin{cases} G \in \mathcal{C}_w([0,1];\mathbb{P}^1), \\ G|_{t=0} = G_0; \quad G|_{t=1} = G_1, \\ \mathfrak{U} \in L^2(0,T;L^2(\mathrm{d}G_t)), \\ \partial_t G_t + G_t \int_{\mathbb{R}^d} \mathrm{d}G_t : U_t = \{-\nabla(G_t u_t) + G_t U_t\}^{Sym} \quad \text{in the weak sense.} \end{cases}$$

Indeed, $\mathcal{A}_1(G_0,G_1) = \mathcal{A}_2(G_0,G_1) \cap \left[\int_{\mathbb{R}^d} \mathrm{d}G_t : U_t \equiv 0\right]$, hence the distance (4.1) is larger than or equal to (4.2). On the other hand, the inverse inequality is also true since for any path $(G_t, U_t, u_t) \in \mathcal{A}_2(G_0,G_1)$ we can find a path in $\mathcal{A}_1(G_0,G_1)$ of the same energy: one just takes $(G_t, \mathfrak{V}_t)$, where $\mathfrak{V}_t \in L^2(0,T;L^2(\mathrm{d}G_t))$ is defined by duality via

$$\langle \mathfrak{V}, (\Phi,\phi)\rangle_{L^2(0,1;L^2(\mathrm{d}G))} = \int_0^1 \left(\int_{\mathbb{R}^d} \mathrm{d}G_t(x) u_t(x)\cdot \phi_t(x) + \int_{\mathbb{R}^d} \mathrm{d}G_t(x) \left[U_t(x) - \int_{\mathbb{R}^d} \mathrm{d}G_t(y) : U_t(y) \right] : \Phi_t(x) \right) \mathrm{d}t. \quad (4.3)$$

We recall [8, 9] that, given a metric space (X, d_X) of diameter $\le \pi$, one can define another metric space $(\mathfrak{C}(X), d_{\mathfrak{C}(X)})$, called a *cone* over X, in the following manner. Consider the quotient $\mathfrak{C}(X) := X \times [0,\infty)/X\times\{0\}$, that is, all points of the fiber $X\times\{0\}$ constitute a single point of the cone which is called the apex. Now set

$$d^2_{\mathfrak{C}(X)}([x_0,r_0],[x_1,r_1]) := r_0^2 + r_1^2 - 2r_0 r_1 \cos(d_X(x_0,x_1)). \quad (4.4)$$

The cones enjoy neat scaling and other nice geometric properties [25]. A particularly regular situation appears when the diameter of X is strictly less than π, since in this case there is a one-to-one correspondence between the geodesics in X and $\mathfrak{C}(X)$. Given a cone $Y = \mathfrak{C}(X)$, X may be referred to as the *sphere* in Y.

Lemma 4.4. *If X is a length space, and $Y = \mathfrak{C}(X)$, then the distance $d_X(x_0,x_1)$ coincides with the infimum of Y-lengths of curves $[x_t,1]$ which join $[x_0,1]$ and $[x_1,1]$ and lie within $X\times\{1\}$.*

Proof. Denote by $I(x_0,x_1)$ the infimum of Y-lengths of curves $[x_t,1]$ as in the statement of the lemma. Observe from (4.4) that $d_X(x_+,x_-) \ge d_{\mathfrak{C}(X)}([x_+,1],[x_-,1])$ for any $x_+,x_- \in X$. Hence, the Y-length of any curve $[x_t,1]$ is less than or equal to the X-length of x_t. We claim that they are actually equal. It suffices to prove

$$L_Y([x_t,1]) \ge q L_X(x_t) \quad (4.5)$$

for any $q < 1$. By linearity of (4.5), it is enough to prove it for curves of sufficiently small length. From [9, Ex. 3.6.14] we infer that

$$L_Y([x_t, 1]) \geq 2\sin(L_X(x_t)/2),$$

which yields (4.5) for short curves. Since X is a length space, we immediately conclude that $I(x_0, x_1) = d_X(x_0, x_1)$. □

We are going to show that the cone over the metric space $(\mathbb{P}^1, d_{SKB}/2)$ coincides with $(\mathbb{P}^+, d_{KB}/2)$. In other words, $(\mathbb{P}^1, d_{SKB}/2)$ is a sphere in the cone $(\mathbb{P}^+, d_{KB}/2)$, hence the name "spherical distance". Firstly, for any element $G \in \mathbb{P}^+$, we set

$$r = r(G) := \sqrt{m(G)} = \sqrt{Tr\, \mathrm{d}G(\mathbb{R}^d)}.$$

Then we can identify G with a pair $[G/r^2, r] \in \mathfrak{C}(\mathbb{P}^1)$.

Theorem 6 (Conic structure)**.** *i) The space $(\mathbb{P}^1, d_{SKB})$ is a geodesic space of diameter $\leq \pi$; ii) $(\mathbb{P}^+, d_{KB}/2)$ is a metric cone over $(\mathbb{P}^1, d_{SKB}/2)$, where $\mathbb{P}^+$ is identified with $\mathfrak{C}(\mathbb{P}^1)$ via $G \leftrightarrow [G/r^2, r]$.*

Proof. Step 1. We first observe that it suffices to show that $(\mathbb{P}^+, d_{KB}/2)$ is a metric cone over some metric space (which, due to the identification above, is nothing but $\mathbb{P}^1$ equipped with some distance d). Indeed, by Proposition 3.2, for any two matrix measures $G_0, G_1 \in \mathbb{P}^1$ one has

$$d_{KB}(G_0, G_1)/2 \leq \sqrt{2}. \tag{4.6}$$

Since $(\mathbb{P}^+, d_{KB}/2)$ is a cone over $(\mathbb{P}^1, d)$, (4.6) and (4.4) imply that $\cos(d(G_0, G_1)) \geq 0$, whence the diameter of $(\mathbb{P}^1, d)$ is controlled from above by $\pi/2 < \pi$. By Theorem 4 and [8, Corollary 5.11], $(\mathbb{P}^1, d)$ is a geodesic space. Evoking Lemma 4.4 and Definition 4.1, we see that d actually coincides with $d_{SKB}/2$.

Step 2. In view of (4.6) and [25, Theorem 2.2], it suffices to show the following scaling property which characterizes the cones:

$$d^2_{KB}(r_0^2 G_0, r_1^2 G_1) = r_0 r_1 d^2_{KB}(G_0, G_1) + 4(r_0 - r_1)^2, \tag{4.7}$$

for all $G_0, G_1 \in \mathbb{P}^1$, $r_0, r_1 \geq 0$. Note that we have already proved it in the case $r_0 r_1 = 0$ (see Corollary 3.4), so we can assume that $r_0 r_1 > 0$. Consider the scalar function $a(t) = \frac{r_1 t}{r_0 + (r_1 - r_0)t}$. Then

$$a(0) = 0,\ a(1) = 1,\ a'(t)(r_0 + (r_1 - r_0)t)^2 = r_0 r_1.$$

We will also need its inverse function $t(a)$.

Let (G_t, U_t, u_t) be any admissible path joining $G_0, G_1 \in \mathbb{P}^1$. Then the path $(\tilde{G}_t, \tilde{U}_t, \tilde{u}_t)$, where

$$\tilde{G}_t = (r_0 + (r_1 - r_0)t)^2 G_{a(t)},$$
$$\tilde{U}_t = a'(t) U_{a(t)} + \frac{2(r_1 - r_0)}{r_0 + (r_1 - r_0)t} I,$$
$$\tilde{u}_t = a'(t) u_{a(t)},$$

connects $r_0^2 G_0$ and $r_1^2 G_1$. A straightforward computation shows that $(\tilde{G}_t, \tilde{U}_t, \tilde{u}_t)$ satisfies the constraint (2.2). Let us compute the energy of this path, using (2.2) with $\Phi = \Phi_a = (r_0 + (r_1 - r_0)t(a))I$:

$$
\begin{aligned}
E[\tilde{G}_t; \tilde{U}_t, \tilde{u}_t] &= \int_0^1 \left(\int_{\mathbb{R}^d} \mathrm{d}\tilde{G}_t \tilde{u}_t \cdot \tilde{u}_t + \int_{\mathbb{R}^d} \mathrm{d}\tilde{G}_t \tilde{U}_t : \tilde{U}_t \right) \mathrm{d}t \\
&= r_0 r_1 \int_0^1 a'(t) \left(\int_{\mathbb{R}^d} \mathrm{d}G_{a(t)} u_{a(t)} \cdot u_{a(t)} + \int_{\mathbb{R}^d} \mathrm{d}G_{a(t)} U_{a(t)} : U_{a(t)} \right) \mathrm{d}t \\
&\quad + 4(r_1 - r_0) \int_0^1 a'(t)(r_0 + (r_1 - r_0)t) \int_{\mathbb{R}^d} \mathrm{d}G_{a(t)} : U_{a(t)} \, \mathrm{d}t \\
&\quad + 4(r_1 - r_0)^2 \int_0^1 \int_{\mathbb{R}^d} \mathrm{d}G_{a(t)} : I \, \mathrm{d}t \\
&= r_0 r_1 \int_0^1 \left(\int_{\mathbb{R}^d} \mathrm{d}G_a u_a \cdot u_a + \int_{\mathbb{R}^d} \mathrm{d}G_a U_a : U_a \right) \mathrm{d}a \\
&\quad + 4(r_1 - r_0) \int_0^1 (r_0 + (r_1 - r_0)t(a)) \int_{\mathbb{R}^d} \mathrm{d}G_a : U_a \, \mathrm{d}a \\
&\quad + 4(r_1 - r_0)^2 \int_0^1 t'(a) \int_{\mathbb{R}^d} \mathrm{d}G_a : I \, \mathrm{d}a \\
&= r_0 r_1 E[G_t; U_t, u_t] \\
&\quad + 4(r_1 - r_0)(r_0 + (r_1 - r_0)t(1)) \int_{\mathbb{R}^d} \mathrm{d}G_1 : I \\
&\quad - 4(r_1 - r_0)(r_0 + (r_1 - r_0)t(0)) \int_{\mathbb{R}^d} \mathrm{d}G_0 : I \\
&= r_0 r_1 E[G_t; U_t, u_t] + 4(r_0 - r_1)^2. \qquad (4.8)
\end{aligned}
$$

Consequently, $d_{KB}^2(r_0^2 G_0, r_1^2 G_1) \le r_0 r_1 d_{KB}^2(G_0, G_1) + 4(r_0 - r_1)^2$. The opposite inequality is proved in a similar fashion. □

Appendix A. Frame-indifference

The principle of material frame-indifference [35] is one of the main principles of rational mechanics, which expresses the fact that the properties of a material do not depend on the choice of an observer. An observer in rational mechanics is identified with a *frame*, which is a correspondence between the spatial points and the elements x of the space $\mathbb{R}^d$, as well as between the moments of time and the elements t of the scalar axis $\mathbb{R}$. The metrics in $\mathbb{R}^d$ and in the scalar axis, as well as the time direction, are assumed to be frame-invariant. Then the most general change of coordinates is

$$
\begin{aligned}
\mathrm{t}^* &= \mathrm{t} - \mathrm{t}_0, \\
x^* &= c^*(\mathrm{t}) + Q_{\mathrm{t}} x,
\end{aligned}
$$

where $\mathfrak{t}_0 \in \mathbb{R}$, $c^* : \mathbb{R} \to \mathbb{R}^d$, $Q_\mathfrak{t}$ is a time-dependent orthogonal matrix.

Consider any vector which exists in the space irrespectively of the observer. In the initial frame, it is represented by some $w \in \mathbb{R}^d$. Then in the new frame it is $w^* = Q_\mathfrak{t} w$. A *frame-indifferent tensor* is a linear automorphism of such vectors. The representations of a frame-indifferent tensor function in the two frames are related as

$$T^*(\mathfrak{t}^*, x^*) = Q_\mathfrak{t} T(\mathfrak{t}, x) Q_\mathfrak{t}^\top.$$

We claim that our distance d_{KB} complies with the frame indifference:

$$d_{KB}(T_0^*(\mathfrak{t}^*, x^*), T_1^*(\mathfrak{t}^*, x^*)) = d_{KB}(T_0(\mathfrak{t}, x), T_1(\mathfrak{t}, x)). \tag{A.1}$$

In other words, d_{KB} may be considered as a distance on positive-semidefinite-frame-indifferent-tensor-valued measures.

To prove the claim it suffices to note that for any admissible path $(T_t, U_t, u_t)(\mathfrak{t}, x)$ in the old frame, the path

$$(T_t^*, U_t^*, u_t^*)(\mathfrak{t}^*, x^*) := \Big(Q_\mathfrak{t} T_t(\mathfrak{t}, x) Q_\mathfrak{t}^\top, Q_\mathfrak{t} U_t(\mathfrak{t}, x) Q_\mathfrak{t}^\top, Q_\mathfrak{t} u_t(\mathfrak{t}, x)\Big)$$

is admissible in the new frame, and has the same energy (2.1). These assertions can be verified by a straightforward computation: the only non-obvious issue for the validity of (2.2) in the new frame is that the spatial gradient is frame-indifferent:

$$\nabla_{x^*} w^* = Q_\mathfrak{t} (\nabla_x w) Q_\mathfrak{t}^\top$$

provided $w^* = Q_\mathfrak{t} w$, which is just a manifestation of the chain rule, cf., e.g., [35, 40].

Appendix B. Some technical facts

Proposition B.1 ([24]). *Let $(X, \|\cdot\|)$ be a separable normed vector space. Assume that there exists a sequence of seminorms $\{\|\cdot\|_k\}$ $(k = 0, 1, 2, \dots)$ on X such that for every $x \in X$ one has*

$$\|x\|_k \le C\|x\|$$

with a constant C independent of k, x, and

$$\|x\|_k \underset{k\to\infty}{\to} \|x\|_0.$$

Let φ_k $(k = 1, 2, \dots)$ be a uniformly bounded sequence of linear continuous functionals on $(X, \|\cdot\|_k)$, resp., in the sense that

$$c_k := \|\varphi_k\|_{(X, \|\cdot\|_k)^*} \le C.$$

Then the sequence $\{\varphi_k\}$ admits a converging subsequence $\varphi_{k_n} \to \varphi_0$ in the weak-$*$ topology of X^*, and

$$\|\varphi_0\|_{(X, \|\cdot\|_0)^*} \le c_0 := \liminf_k c_k. \tag{B.1}$$

Lemma B.2. *The matricial bounded-Lipschitz distance d_{BL} is sequentially lower semicontinuous with respect to the weak-$*$ topology.*

The proof is obvious since the supremum in the definition of d_{BL} can be restricted to smooth compactly supported functions, which are dense in $\mathcal{C}_0$.

Lemma B.3. *Let (X,ρ) be a metric space where every two points almost admit a midpoint. Assume that there exists a Hausdorff topology σ on X such that ρ-bounded sequences contain σ-converging subsequences, and ρ is sequentially lower semicontinuous with respect to σ. Then (X,ρ) is a geodesic space.*

Proof. Fix any two points $x_0, x_1 \in X$. It suffices to join them by a curve x_t such that

$$\rho(x_t, x_{\bar{t}}) \le |t-\bar{t}|\rho(x_0,x_1). \tag{B.2}$$

for all $t,\bar{t} \in [0,1]$ (which is a posteriori continuous).

Let us first observe that every two points $x,y \in X$ admit a midpoint, that is,

$$\rho(x,y) = 2\rho(x,z) = 2\rho(z,y).$$

for some $z \in X$. Indeed, take any sequence z_k of almost midpoints, i.e.,

$$|\rho(x,y) - 2\rho(x,z_k)| \le k^{-1}, \quad |\rho(x,y) - 2\rho(y,z_k)| \le k^{-1}.$$

The sequence $\{z_k\}$ is ρ-bounded, thus without loss of generality it σ-converges to some $z \in X$. Then

$$2\rho(x,z) \le \lim_{k\to\infty} 2\rho(x,z_k) = \rho(x,y),$$

$$2\rho(y,z) \le \lim_{k\to\infty} 2\rho(y,z_k) = \rho(x,y).$$

But its is clear from the triangle inequality that the latter inequalities must be equalities.

Let $Q = \{s \in [0,1] | \exists p \in \mathbb{N} : 2^p s \in \mathbb{N}\}$. With the existence of midpoints at hand, by a standard procedure [9, p. 43] one constructs points x_s ($s \in Q$) satisfying (B.2), that is, the function $s \mapsto x_s$ is $\rho(x_0,x_1)$-Lipschitz. Given any $t \in [0,1]$, we can approximate it by a sequence $\{s_n\} \in Q$. Since $s \mapsto x_s$ is Lipschitz on Q, x_{s_n} is a ρ-Cauchy sequence. Therefore it is ρ-bounded, and admits a subsequence which σ-converges to some $x_t \in X$. Due to the sequential lower semicontinuity of the distance ρ, we can pass to the limit in (B.2) for all $t,\bar{t} \in [0,1]$. □

Lemma B.4. *Let (X,ρ) be a metric space. Assume that there exists a Hausdorff topology σ on X such that ρ is sequentially lower semicontinuous with respect to σ. Let $(x^k)_t$, $t \in [0,1]$, be a sequence of curves lying in a common σ-sequentially compact set $K \subset X$. Let it be equicontinuous in the sense that there exists a symmetric continuous function $\omega : [0,1]\times[0,1] \to \mathbb{R}_+$, $\omega(t,t) = 0$, such that*

$$\rho((x^k)_t, (x^k)_{\bar{t}}) \le \omega(t,\bar{t}). \tag{B.3}$$

for all $t,\bar{t} \in [0,1]$. Then there exists a ρ-continuous curve x_t such that

$$\rho(x_t, x_{\bar{t}}) \le \omega(t,\bar{t}), \tag{B.4}$$

and (up to a not relabelled subsequence)

$$(x^k)_t \to x_t \tag{B.5}$$

for all $t \in [0,1]$ in the topology σ.

Proof. A standard Arzelà-Ascoli argument allows us to construct, for each rational number $t \in [0,1]$, some points x_t so that (B.5) holds up to a not relabelled subsequence. Due to the sequential lower semicontinuity of ρ, estimate (B.4) is true for all rational $t, \bar{t} \in [0,1]$. Approximating any point $t \in [0,1]$ by a sequence of rational numbers, by mimicking the reasoning from the proof of Lemma B.3 we can construct a ρ-continuous curve x_t satisfying (B.4) for every $t \in [0,1]$. To show that the convergence (B.5) takes place for all $t \in [0,1]$, one just repeats the argument from [1, last part of the proof of Proposition 3.3.1]. □

Remark B.5. Lemma B.3 (refined Hopf-Rinow) has been proved in [24] assuming that X is a complete length space, which is redundant. Similarly, Lemma B.4 (refined Arzelà-Ascoli) has been proved in [1] assuming that X is a complete metric space.

Appendix C. Matricial Otto calculus

We have seen in Remark 3.5 that some pieces of $(\mathbb{P}^+, d_{KB})$ are isometric to Riemannian manifolds. One can (at least formally) extend this geometry onto the whole $\mathbb{P}^+$ such that the corresponding geodesic distance coincides with d_{KB}. Namely, we can develop some kind of Otto calculus, cf. [32, 37, 38], on $(\mathbb{P}^+, d_{KB})$. Starting from this point, we are completely formal. As we observed in Section 2, the minimizing potentials in (2.1) can be chosen in the form $\mathfrak{U} = (U, \operatorname{div} U)$. This suggests to define the tangent spaces as

$$T_G\mathbb{P}^+ := \left\{ \exists U(x) \in \mathcal{S} : \Xi = (-\nabla(G \operatorname{div} U) + GU)^{Sym} \right\}$$

and

$$\|\Xi\|_{T_G\mathbb{P}^+} = \|U\|_{H^1_{\mathrm{div}}(\mathrm{d}G;\mathcal{S})} := \left(\int_{\mathbb{R}^d} \mathrm{d}G \operatorname{div} U \cdot \operatorname{div} U + \int_{\mathbb{R}^d} \mathrm{d}GU : U \right)^{1/2}.$$

Ignoring all smoothness issues, the operator

$$\Xi(U) = (-\nabla(G \operatorname{div} U) + GU)^{Sym} \tag{C.1}$$

is $H^1_{\mathrm{div}}(\mathrm{d}G;\mathcal{S})$-coercive, so the one-to-one correspondence between the tangent vectors Ξ and potentials $\mathfrak{U} = (U, \operatorname{div} U)$ is well defined. By polarization this defines a Riemannian metric on $T\mathbb{P}^+$, and

$$d^2_{KB}(G_0, G_1) = \inf \left\{ \int_0^1 \left\| \frac{dG_t}{dt} \right\|^2_{T_{G_t}\mathbb{P}^+} \mathrm{d}t \right\}.$$

The gradients of functionals $\mathcal{F} : \mathbb{P}^+ \to \mathbb{R}$ are given by

$$\operatorname{grad}_{KB} \mathcal{F}(G) = \left[-\nabla \left(G \operatorname{div} \frac{\delta \mathcal{F}}{\delta G} \right) + G \frac{\delta \mathcal{F}}{\delta G} \right]^{Sym}, \tag{C.2}$$

where $\frac{\delta \mathcal{F}}{\delta G}$ denotes the conventional first variation with respect to the Euclidean structure $\langle U_1, U_2 \rangle = \int_{\mathbb{R}^d} U_1 : U_2$. The gradient flows are matricial PDEs of the form

$$\partial_t G = \left[\nabla \left(G \operatorname{div} \frac{\delta \mathcal{F}}{\delta G} \right) - G \frac{\delta \mathcal{F}}{\delta G} \right]^{Sym}.$$

The interesting driving functionals include the von Neumann entropy

$$\mathcal{F}_N(G) = \int_{\mathbb{R}^d} G \log G - G$$

and the "volume"

$$\mathcal{F}_V(G) = \int_{\mathbb{R}^d} \sqrt{\det G}.$$

The gradient flow of $\mathcal{F}_N$ is a sort of matricial "heat flow" with logarithmic reaction. Indeed, for $d = 1$ it simply becomes $\partial_t G = \partial_{xx} G - G \log G$. The gradient flow of $\mathcal{F}_V$ has some similarities with the mean curvature flow (if we view G_t as an evolving Riemannian metric on $\mathbb{R}^d$). Unfortunately, it is not a genuinely geometric flow since the latter ones are expected to be invariant with respect to diffeomorphisms of $\mathbb{R}^d$ (for instance, the Ricci flow has this property), and our flow, in spite of the frame-indifference of the distance, does not behave in such a nice way.

The considerations above can be applied to the spherical space $(\mathbb{P}^1, d_{SKB})$. Remark 4.3 guides us to define

$$T_G\mathbb{P}^1 := \left\{ \exists U(x) \in \mathcal{S} : \Xi = [-\nabla(G \operatorname{div} U) + GU]^{Sym} - G \int_{\mathbb{R}^d} G : U \right\}$$

and

$$\|\Xi\|_{T_G\mathbb{P}^1} = \left(\int_{\mathbb{R}^d} \mathrm{d}G \operatorname{div} U \cdot \operatorname{div} U + \int_{\mathbb{R}^d} \mathrm{d}GU : U - \left(\int_{\mathbb{R}^d} \mathrm{d}G : U \right)^2 \right)^{1/2}.$$

The gradients of functionals $\mathcal{F} : \mathbb{P}^1 \to \mathbb{R}$ are

$$\operatorname{grad}_{SKB} \mathcal{F}(G) = \left[-\nabla \left(G \operatorname{div} \frac{\delta \mathcal{F}}{\delta G} \right) + G \frac{\delta \mathcal{F}}{\delta G} \right]^{Sym} - G \int_{\mathbb{R}^d} G : \frac{\delta \mathcal{F}}{\delta G}. \tag{C.3}$$

The second order calculus for both the cone and the sphere can be established by formally computing the geodesic equations, which leads to the definitions of Hessians and λ-convexity.

References

[1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows: in Metric Spaces and in the Space of Probability Measures*. Basel: Birkhäuser Basel, 2008.

[2] L. Ambrosio and P. Tilli. *Topics on analysis in metric spaces*, volume 25 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2004.

[3] N. Ay, J. Jost, H. Vân Lê, and L. Schwachhöfer. *Information geometry*. Springer, 2017.

[4] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge -Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[5] R. Bhatia. *Positive definite matrices*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2007.

[6] V. I. Bogachev. *Measure theory. Vol. II*. Springer-Verlag, Berlin, 2007.

[7] Y. Brenier. The initial value problem for the Euler equations of incompressible fluids viewed as a concave maximization problem. *Comm. Math. Phys.*, 364(2):579–605, 2018.

[8] M. R. Bridson and A. Haefliger. *Metric spaces of non-positive curvature*, volume 319 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999.

[9] D. Burago, Y. Burago, and S. Ivanov. *A course in metric geometry*. AMS, 2001.

[10] E. A. Carlen and J. Maas. An analog of the 2-Wasserstein metric in non-commutative probability under which the fermionic Fokker-Planck equation is gradient flow for the entropy. *Comm. Math. Phys.*, 331(3):887–926, 2014.

[11] E. A. Carlen and J. Maas. Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance. *J. Funct. Anal.*, 273(5):1810–1869, 2017.

[12] Y. Chen, W. Gangbo, T. T. Georgiou, and A. Tannenbaum. On the matrix Monge-Kantorovich problem. *arXiv preprint arXiv:1701.02826*, 2017.

[13] Y. Chen, T. T. Georgiou, and A. Tannenbaum. Matrix optimal mass transport: a quantum mechanical approach. *IEEE Trans. Automat. Control*, 63(8):2612–2619, 2018.

[14] Y. Chen, T. T. Georgiou, and A. Tannenbaum. Wasserstein geometry of quantum states and optimal transport of matrix-valued measures. In *Emerging applications of control and systems theory*, Lect. Notes Control Inf. Sci. Proc., pages 139–150. Springer, Cham, 2018.

[15] Y. Chen, T. T. Georgiou, and A. Tannenbaum. Interpolation of matrices and matrix-valued densities: the unbalanced case. *European J. Appl. Math.*, 30(3):458–480, 2019.

[16] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. An interpolating distance between optimal transport and fisher–rao metrics. *Foundations of Computational Mathematics*, 18(1):1–44, 2018.

[17] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. Unbalanced optimal transport: Dynamic and Kantorovich formulations. *Journal of Functional Analysis*, 274(11):3090–3123, 2018.

[18] J. Dittmann. Explicit formulae for the Bures metric. *Journal of Physics A: Mathematical and General*, 32(14):2663, 1999.

[19] R. M. Dudley. *Real analysis and probability*. Cambridge University Press, 2002.

[20] A. J. Duran and P. Lopez-Rodriguez. The L^p space of a positive definite matrix of measures and density of matrix polynomials in L^1. *J. Approx. Theory*, 90(2):299–318, 1997.

[21] F. Golse, C. Mouhot, and T. Paul. On the mean field and classical limits of quantum mechanics. *Comm. Math. Phys.*, 343(1):165–205, 2016.

[22] S. J. Gustafson and I. M. Sigal. *Mathematical concepts of quantum mechanics*. Universitext. Springer, Heidelberg, second edition, 2011.

[23] B. Khesin, J. Lenells, G. Misiołek, and S. C. Preston. Geometry of diffeomorphism groups, complete integrability and geometric statistics. *Geom. Funct. Anal.*, 23(1):334–366, 2013.

[24] S. Kondratyev, L. Monsaingeon, and D. Vorotnikov. A new optimal transport distance on the space of finite Radon measures. *Adv. Differential Equations*, 21(11-12):1117–1164, 2016.

[25] V. Laschos and A. Mielke. Geometric properties of cones with applications on the Hellinger-Kantorovich space, and a new distance on the space of probability measures. *J. Funct. Anal.*, 276(11):3529–3576, 2019.

[26] M. Liero, A. Mielke, and G. Savaré. Optimal transport in competition with reaction: the Hellinger-Kantorovich distance and geodesic curves. *SIAM J. Math. Anal.*, 48(4):2869–2911, 2016.

[27] M. Liero, A. Mielke, and G. Savaré. Optimal entropy-transport problems and a new hellinger–kantorovich distance between positive measures. *Inventiones mathematicae*, 211(3):969–1117, 2018.

[28] M. Mittnenzweig and A. Mielke. An entropic gradient structure for Lindblad equations and couplings of quantum systems to macroscopic models. *J. Stat. Phys.*, 167(2):205–233, 2017.

[29] K. Modin. Generalized Hunter-Saxton equations, optimal information transport, and factorization of diffeomorphisms. *J. Geom. Anal.*, 25(2):1306–1334, 2015.

[30] L. Ning and T. T. Georgiou. Metrics for matrix-valued measures via test functions. In *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, pages 2642–2647. IEEE, 2014.

[31] S.-i. Ohta. Totally geodesic maps into metric spaces. *Math. Z.*, 244(1):47–65, 2003.

[32] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.
[33] G. Peyré, L. Chizat, F.-X. Vialard, and J. Solomon. Quantum optimal transport for tensor field processing. *arXiv preprint arXiv:1612.08731*, 2016.
[34] F. Rezakhanlou. Optimal transport problem and contact structures. *preprint*, 2015.
[35] C. A. Truesdell. *A first course in rational continuum mechanics. Vol. 1*. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1977. General concepts, Pure and Applied Mathematics.
[36] A. Uhlmann. The metric of Bures and the geometric phase. In *Groups and related Topics, Gielerak et al. (eds.)*, pages 267–274. Springer, 1992.
[37] C. Villani. *Topics in optimal transportation*. American Mathematical Soc., 2003.
[38] C. Villani. *Optimal transport: old and new*. Springer Science & Business Media, 2008.
[39] D. Vorotnikov. Variational solutions to the abstract Euler equation. *arXiv e-prints*, page arXiv:1905.06059, May 2019.
[40] V. G. Zvyagin and D. A. Vorotnikov. *Topological approximation methods for evolutionary problems of nonlinear hydrodynamics*, volume 12 of *De Gruyter Series in Nonlinear Analysis and Applications*. Walter de Gruyter & Co., Berlin, 2008.

(Y. Brenier) Département de mathématiques et applications, CNRS DMA (UMR 8553), École Normale Supérieure, 45 rue d' Ulm, 75005 Paris, France
E-mail address: yann.brenier@ens.fr

(D. Vorotnikov) CMUC, Department of Mathematics, University of Coimbra, 3001-501 Coimbra, Portugal
E-mail address: mitvorot@mat.uc.pt