ON OPTIMAL TRANSPORT OF MATRIX-VALUED MEASURES

YANN BRENIER AND DMITRY VOROTNIKOV

ABSTRACT. The classical Monge-Kantorovich transport deals with probability densities, but it is tempting to try to transport more or less similar mathematical objects in a relevant way. In these preliminary notes, we will focus on optimal transport of positive-semidefinite matrix-valued measures. Our approach is not based on the Lindblad equation, and works exclusively in the unbalanced case.

1. INTRODUCTION

These raw notes introduce and study some optimal transport distances on the spaces of matrix-valued measures. This field is rather active, cf. [13, 8, 9, 12, 10, 11, 25, 21]. By contrast, our approach is not based on the Lindblad equation, but is somehow inspired by hydrodynamics [5] and rational mechanics [26]. It generalizes the unbalanced optimal transport of scalar Radon measures [17, 19, 20, 14]. Surprisingly, it does not work as a direct generalization of the Wasserstein distance, but operates exclusively in the unbalanced (Hellinger-Kantorovich aka Wasserstein-Fisher-Rao) case.

Notations and conventions.

- Basic spaces:

  \( \mathbb{R}^{d \times d} \)

  is the space of \( d \times d \) matrices, equipped with the Frobenius product

  \[ \Phi : \Psi = \text{Tr}(\Phi \Psi^\top) \]

  and the norm \( |\Phi| = \sqrt{\Phi : \Phi} \),

  \[ A^\text{Sym} = \frac{1}{2}(A + A^\top) \]

denotes 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,

$$\mathbb{P}^+$$

is the set of $\mathcal{P}^+$-valued Radon measures $P$ on $\mathbb{R}^d$ with finite $Tr(dP(\mathbb{R}^d))$,

$$\mathbb{P}^{++}$$

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 $Tr(dP(\mathbb{R}^d))$,

$$\mathbb{P}^1$$

is the set of $\mathcal{P}^+$-valued Radon measures $P$ on $\mathbb{R}^d$ with $Tr(dP(\mathbb{R}^d))=1$.

Remark 1.1. All the considerations of the paper are valid, mutatis mutandis, if the Radon measures above are defined on the torus $\mathbb{T}^d$ instead of $\mathbb{R}^d$.

- We will use the following simple inequalities

$$PA : A \leq Tr(P|A|^2), \quad Pq \cdot q \leq Tr(P|q|^2), \quad P \in \mathbb{P}^+, A \in \mathbb{R}^{d \times d}, q \in \mathbb{R}^d.$$  

$$(1.1)$$

$$A : B \geq 0, \ A, B \in \mathcal{P}^+.$$

$$(1.2)$$

- 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 $\|\phi\|_{\text{Lip}} = \|\nabla \phi\|_\infty + \|\phi\|_\infty$.

- 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

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

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

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

- 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

  $$\int_{\mathbb{R}^d} \phi d(F \# G_0) = \int_{\mathbb{R}^d} \phi \circ F \, dG_0$$
for all test functions $\phi \in C_b^b(\mathbb{R}^d)$.

- For curves $t \in [0, 1] \mapsto G_t \in \mathbb{P}^+$ we write $G \in C_w([0, 1]; \mathbb{P}^+)$ for the continuity with respect to the narrow topology.

- Given a non-identically-zero measure $G \in \mathbb{P}^+$ we will denote by $L^2(dG) = L^2(dG; \mathcal{S} \times \mathbb{R}^d)$ the Hilbert space obtained by completion of the quotient by the seminorm kernel of the space $C_b^1(\mathbb{R}^d; \mathcal{S} \times \mathbb{R}^d)$ equipped with the Hilbert seminorm

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

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

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

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

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

$$d_{BL}(G_0, G_1) = \sup_{\|\Phi\|_{\text{Lip}} \leq 1} \left| \int_{\mathbb{R}^d} \Phi : (dG_1 - dG_0) \right|.$$

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 [22]. In our situation the claim can still be shown by mimicking the proof strategy for the scalar-valued Radon measures [16,15]. The key observation [22] is that $S$-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 smoothly 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 metric space $(\mathbb{P}^+, d_{HK})$

**Definition 2.1.** Given two matrix measures $G_0, G_1 \in \mathbb{P}^+$ we define

$$d_{HK}^2(G_0, G_1) := \inf_{A(G_0, G_1)} \|U\|_{L^2(0, 1; L^2(dG_t))}^2,$$
where the admissible set $\mathcal{A}(G_0, G_1)$ consists of all couples $(G_t, U_t)_{t \in [0,1]}, \ U_t = (U_t, u_t)$, such that
\[
\begin{cases}
G \in C_w([0,1]; \mathbb{P}^+), \\
G|_{t=0} = G_0; \quad G|_{t=1} = G_1, \\
U \in L^2(0,T; L^2(dG_t)), \\
\partial_t G_t = \{-\nabla (G_t u_t) + G_t U_t\}^{\text{sym}} \quad \text{in the weak sense, i.e.,}
\end{cases}
\]

(2.2) \[\int_{\mathbb{R}^d} \Phi : (dG_t - dG_s) - \int_s^t \int_{\mathbb{R}^d} (dG_{\tau} : \partial_{\tau} \Phi_{\tau}) d\tau \]
\[= \int_s^t \int_{\mathbb{R}^d} (dG_{\tau} u_{\tau} \cdot \text{div} \Phi_{\tau} + dG_{\tau} U_{\tau} : \Phi_{\tau}) d\tau \]
for all test functions $\Phi \in C^1_b([0,1] \times \mathbb{R}^d; \mathcal{S})$ and $t, s \in [0,1]$.

**Remark 2.2.** We could have formally started from minimizing a more general Lagrangian, namely,

(2.3) \[d_{HK}^2(G_0, G_1) := \inf_{B(G_0, G_1)} \int_0^1 \left( \int_{\mathbb{R}^d} G_t^{-1}(x) q_t(x) \cdot q_t(x) + G_t^{-1}(x) R_t(x) : R_t(x) \, dx \right) dt, \]
where the admissible set $B(G_0, G_1)$ 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
\[
\begin{cases}
G|_{t=0} = G_0; \quad G|_{t=1} = G_1, \\
\partial_t G_t = \{-\nabla q_t + R_t\}^{\text{sym}}.
\end{cases}
\]

Perturbing an alleged minimizer by adding $(\delta q, \delta R)$ for which
\[L(\delta q, \delta R) := \{-\nabla (\delta q_t) + \delta R_t\}^{\text{sym}} = 0, (\delta q, \delta R)|_{t=0,1} = 0, \]
we see that the minimizer satisfies
\[\int_0^1 \left( \int_{\mathbb{R}^d} G_t^{-1}(x) q_t(x) \cdot \delta q_t(x) + G_t^{-1}(x) R_t(x) : \delta R_t(x) \, dx \right) dt = 0, \]
for all perturbations $(\delta q, \delta R)$ from $\text{Ker} \ L$. This implies that such a minimizer can be written in the form $q = G \text{div} \ U$, $R = GU$, for some $U_t(x) \in \mathcal{S}$, hence (2.3) yields (2.1) via setting $u := \text{div} \ U$.

We shall prove shortly that

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

We first need a preliminary technical bound:
Lemma 2.3. Let \( G \in C_b([0, 1]; \mathbb{P}^+) \) be a narrowly continuous curve, assume that the constraint (2.2) is satisfied for some potential \( \mathbf{u} \in L^2(0, T; L^2(dG_t)) \) with finite energy
\[
E = E[G; \mathbf{u}] = ||\mathbf{u}||^2_{L^2(0, T; L^2(dG_t))}
\]
and let \( M := 2(\max\{m_0, m_1\} + E) \) with \( m_i = Tr \, dG_t(\mathbb{R}^d) \). Then the masses are bounded uniformly in time, \( m_t = Tr \, dG_t(\mathbb{R}^d) \leq M \) and
\[
\forall \Phi \in C^1_b(\mathbb{R}^d; \mathcal{S}) : \quad \left| \int_{\mathbb{R}^d} \Phi : (dG_t - dG_s) \right| \leq (|| \Phi ||_\infty + || \Phi ||_\infty) \sqrt{ME} |t - s|^{1/2}
\]
for all \( 0 \leq s \leq t \leq 1 \).

Proof. By narrow continuity of \( t \mapsto G_t \) the masses \( m_t \) are uniformly bounded and \( m = \max_{t \in [0, 1]} m_t \) is finite. A Cauchy-Schwarz-like argument applied to the weak constraint (2.2), together with (1.1), imply
\[
\left| \int_{\mathbb{R}^d} \Phi : (dG_t - dG_s) \right| = \left| \int_s^t \left( \int_{\mathbb{R}^d} dG_\tau(x) u_\tau(x) \cdot \text{div} \, \Phi(x) + \int_{\mathbb{R}^d} dG_\tau(x) U_\tau(x) : \Phi(x) \right) d\tau \right|
\]
\[
\leq \left( \int_s^t \left( \int_{\mathbb{R}^d} dG_\tau(x) \text{div} \, \Phi(x) \cdot \text{div} \, \Phi(x) + \int_{\mathbb{R}^d} dG_\tau(x) \Phi(x) : \Phi(x) \right) d\tau \right)^{1/2}
\]
\[
\times \left( \int_s^t \left( \int_{\mathbb{R}^d} dG_\tau(x) u_\tau(x) \cdot u_\tau(x) + \int_{\mathbb{R}^d} dG_\tau(x) U_\tau(x) : U_\tau(x) \right) d\tau \right)^{1/2}
\]
\[
\leq (|| \text{div} \, \Phi ||_\infty + || \Phi ||_\infty) \sqrt{m} \cdot |t - s|^{1/2} E^{1/2},
\]
and it is enough to estimate \( m \leq M = 2(\max\{m_0, m_1\} + E) \) as in our statement. Choosing \( \Phi \equiv I \) we obtain from the previous estimate \( |m_t - m_s| \leq \sqrt{mE} |t - s|^{1/2} \).

Let \( t_0 \in [0, 1] \) be any time when \( m_{t_0} = m \): choosing \( t = t_0 \) and \( s = 0 \) we immediately get \( m \leq m_0 + \sqrt{mE} |t_0 - 0|^{1/2} \leq m_0 + \sqrt{mE} \), and some elementary algebra bounds \( m \leq 2(m_0 + E) \). Exchanging the roles of \( G_0, G_1 \) we get similarly \( m \leq 2(m_1 + E) \), and finally \( m \leq M \).

Proof of Theorem 4. Let us first show that \( d_{HK}(G_0, G_1) \) 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 \( \mathbf{u}_t = (-\frac{2}{1-t} I, 0) \) 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 \( E = 4 Tr(dP_t(\mathbb{R}^d)) < \infty \) (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
Let now 
\[ d = \text{definition of our distance and the explicit time scaling we get that} \]

\[ \text{Letting} \quad k \to \infty \text{we obtain for any fixed} \quad \Phi \in C^\infty(\mathbb{R}^d; \mathcal{S}) \text{the fundamental estimate} \quad \text{[2.4]} \]

\[ \left| \int_{\mathbb{R}^d} \Phi : (dG - dG_0) \right| \leq (\| \text{div} \Phi \|_\infty + \| \Phi \|_\infty) \sqrt{ME[G^k; \mathcal{U}^k]} \]

Since \( \lim_{k \to \infty} E[G^k; \mathcal{U}^k] = 0 \) we conclude that \( \int_{\mathbb{R}^d} \Phi : (dG - dG_0) \) for all \( \Phi \in C^\infty(\mathbb{R}^d; \mathcal{S}) \), 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 \( d_{HK}(G_0, G_1) \leq d_{HK}(G_0, P) + d_{HK}(P, G_1) \). We can assume that all three distances are nonzero otherwise the triangular inequality trivially holds by the previous point. Let now \( (G^k, \mathcal{U}^k) \) be a minimizing sequence in the definition of \( d_{HK}(G_0, P) = \lim_{k \to \infty} E[G^k; \mathcal{U}^k] \), and let similarly \( (G_t^k, \mathcal{U}_t^k) \) be such that \( d_{HK}(P, G_1) = \lim_{k \to \infty} E[G_t^k; \mathcal{U}_t^k] \).

For fixed \( \tau \in (0, 1) \) let \( (G_t, \mathcal{U}_t) \) be the continuous path obtained by first following \( (G_t^k, \mathcal{U}_t^k) \) from \( G_0 \) to \( P \) in time \( \tau \), and then following \( (G_t^k, \mathcal{U}_t^k) \) from \( P \) to \( G_1 \) in time \( 1 - \tau \). Then \( (G_t^k, \mathcal{U}_t^k) \) is an admissible path connecting \( G_0 \) to \( G_1 \), hence by definition of our distance and the explicit time scaling we get that

\[ d_{HK}^2(G_0, G_1) \leq E[G^k; \mathcal{U}^k] = \frac{1}{\tau}E[G_t^k; \mathcal{U}_t^k] + \frac{1}{1 - \tau}E[G_t^k; \mathcal{U}_t^k] \]

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

\[ d_{HK}^2(G_0, G_1) \leq \frac{1}{\tau}d_{HK}^2(G_0, P) + \frac{1}{1 - \tau}d_{HK}^2(P, G_1) \]

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

\[ d_{HK}^2(G_0, G_1) \leq \frac{1}{\tau}d_{HK}^2(G_0, P) + \frac{1}{1 - \tau}d_{HK}^2(P, G_1) = (d_{HK}(G_0, P) + d_{HK}(P, G_1))^2 \]

and the proof is complete. \( \qed \)

**Corollary 2.4.** The elements of a bounded set in \( (\mathbb{P}^+, d_{HK}) \) have uniformly bounded mass. Conversely, subsets of \( \mathbb{P}^+ \) with uniformly bounded mass are bounded in \( (\mathbb{P}^+, d_{HK}) \).
Proof. The first statement is an immediate consequence of Lemma 2.4. 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 2.5.** If $(G_t, \mathcal{U}_t)_{t \in [0,1]}$ is a narrowly continuous curve with total energy $E$ then $t \mapsto G_t$ is $1/2$-Hölder continuous w.r.t. $d_{HK}$, and more precisely

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

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

**Remark 2.6.** 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 $d_{HK}$, which might be less relevant in applications. For example, $d_{HK}(\delta x_k I, \delta_0 I) \to 0$ 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

**Theorem 2.** The convergence of matrix measures w.r.t. the distance $d_{HK}$ implies narrow convergence, and any Cauchy sequence in $(\mathbb{P}^+, d_{HK})$ 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

$$(3.1) \quad d_{BL}(G_0, G_1) \leq C_d \sqrt{m_0 + m_1}d_{HK}(G_0, G_1)$$

with some uniform $C_d$ depending only on the dimension.

**Proof.** Fix $G_0, G_1$, and let $(G_t, \mathcal{U}_t)$ be any admissible path from $G_0$ to $G_1$ with finite energy $E$. Taking the supremum over $\Phi$ with $\|\Phi\|_{\text{Lip}} \leq 1$ in (2.1) we get $d_{BL}(G_0, G_1) \leq C\sqrt{ME}$, where $M = 2(\max\{m_0, m_1\} + E)$ as in Lemma 2.3. Choosing now a minimizing sequence instead of an arbitrary path and taking the limit we essentially obtain the same estimate with $E = \lim E[G^k; \mathcal{U}^k] = d_{HK}^2(G_0, G_1)$, whence

$$d_{BL}(G_0, G_1) \leq C\sqrt{2(\max\{m_0, m_1\} + d_{HK}^2(G_0, G_1))d_{HK}(G_0, G_1)}.$$  

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

If $G^k$ is a Cauchy sequence in $(\mathbb{P}^+, d_{HK})$ with mass $m^k = Tr dG^k(\mathbb{R}^d)$. Since Cauchy sequences are bounded we control $4m^k = d_{HK}^2(G^k, 0) \leq C$ uniformly in $k$, as claimed.
thus from [3,3] we see that
\[ d_{BL}(G^p, G^q) \leq C d_{HK}(G^p, G^q). \]
Thus, \( G^k \) is \( d_{BL} \)-Cauchy. Similarly, if a sequence is \( d_{HK} \)-converging, it is \( d_{BL} \)- and hence narrowly converging (to the same limit).

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

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

**Proof.** Consider any two converging sequences
\[ G^k_0 \xrightarrow{k \to \infty} G_0, \quad G^k_1 \xrightarrow{k \to \infty} G_1 \]
of finite Radon measures from \( \mathbb{P}^+(\mathbb{R}^d) \). For each \( k \), the endpoints \( G^k_0 \) and \( G^k_1 \) can be joined by an admissible narrowly continuous path \((G^k_t, \Omega^k_t)_{t \in [0,1]}\) with energy
\[ E[G^k; \Omega^k] \leq d_{HK}^2(G^k_0, G^k_1) + k^{-1}. \]
Due to weak-* compactness, the masses \( m^k_0 = Tr dG^k_0(\mathbb{R}^d) \) and \( m^k_1 = Tr dG^k_1(\mathbb{R}^d) \) are bounded uniformly in \( k \in \mathbb{N} \). By Corollary [2,4] the set \( \bigcup_{k \in \mathbb{N}} \{G^k_0, G^k_1\} \) is bounded in \((\mathbb{P}^+, d_{HK})\), thus the energies \( E[G^k; \Omega^k] \) and the masses \( m^k_t = Tr dG^k_t(\mathbb{R}^d) \) are bounded uniformly in \( k \in \mathbb{N} \) and \( t \in [0,1] \)
\[ m^k_t \leq M \quad \text{and} \quad E[G^k; \Omega^k] \leq \overline{E}. \]
By the (classical) Banach-Alaoglu theorem with \( \mathbb{P}^+ \subset (C_0)^* \), all the curves \((G^k_t)_{t \in [0,1]}\) lie in a fixed weak-* sequentially relatively compact set \( \mathcal{K}_M = \{G \in \mathbb{P}^+: Tr dG(\mathbb{R}^d) \leq M\} \) uniformly in \( k, t \). By the fundamental estimate [2.4] we get
\[ \left| \int_{\mathbb{R}^d} \Phi : (dG^k_t - dG^k_s) \right| \leq \sqrt{ME} |t-s|^{1/2}(\| \div \Phi \|_\infty + \| \Phi \|_\infty) \leq C |t-s|^{1/2}(\| \nabla \Phi \|_\infty + \| \Phi \|_\infty) \]
for all \( \phi \in C^1_0 \), which implies
\[ \forall t, s \in [0,1], \forall k \in \mathbb{N} : \quad d_{BL}(G^k_s, G^k_t) \leq C |t-s|^{1/2}. \]
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 [3.2] in Appendix [3]), and the fact that \( G^k_t \in \mathcal{K}_M \), we conclude by a refined version of Arzelà-Ascoli
A NEW MATRIX DISTANCE

Theorem (Lemma \[B.4\] in 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

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

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

\[
\forall \phi \in \mathcal{C}_c(Q): \quad \int_Q \phi(t, x) d\mu^k(t, x) = \int_0^1 \left( \int_{\mathbb{R}^d} \phi(t, .) dG^k_t \right) dt.
\]

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

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

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

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

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

Namely, we set

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

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

and define the linear forms

\[
\varphi_k(\Phi, \phi) = \int_Q d\mu^k u^k \cdot \phi + d\mu^k U^k : \Phi, \quad k = 1, 2, \ldots.
\]

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

\[
c_k := \| \varphi_k \|_{(X, \| \cdot \|_k)^*} \leq \| \mu^k \|_{L^2(0,1; L^2(dG^k))} = \sqrt{E[G^k; U^k]} \leq \sqrt{d_{HK}^2(G^k_0, G^k_1) + k^{-1}}.
\]

Consequently, there exists a continuous functional \(\varphi_0\) on the space \((X, \| \cdot \|_0)\) such that up to a subsequence

\[
\forall (\Phi, \phi) \in \mathcal{C}^1_c(Q; \mathcal{S} \times \mathbb{R}^d): \quad \int_0^1 \left( \int_{\mathbb{R}^d} dG^k_t u^k_t \cdot \phi_t + dG^k_t U^k : \Phi_t \right) dt \to k \to \infty \varphi_0(\Phi)
\]

with moreover

\[
\| \varphi_0 \|_{(X, \| \cdot \|_0)^*} \leq \liminf_{k \to \infty} d_{HK}(G^k_0, G^k_1).
\]
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 $X/N_0$ of $X/N_0$ with respect to $\| \cdot \|_0$, which is exactly $L^2(0, 1; L^2(dG_t))$. As a consequence there exists $\mathcal{U} = (U, u) \in L^2(0, T; L^2(dG_t))$ such that

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

and

$$\| \mathcal{U} \|_{L^2(0, 1; L^2(dG_t))} = \| \varphi_0 \|_{(X, \| \cdot \|_0)^*},$$

and it is straightforward to check that $(G, \mathcal{U})$ 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

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

During the proof of Theorem 2 we observed the upper bound

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

Let us show that it can improved.

**Proposition 3.2.** For every pair $G_0, G_1 \in \mathbb{P}^+$ with masses $m_0, m_1$ one has

(3.4)

$$d^2_{HK}(G_0, G_1) \leq 4(m_0 + m_1).$$

**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

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

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

(3.5)

$$\langle \mathcal{U}, (\Phi, \phi) \rangle_{L^2(0, 1; L^2(dG))} = 2 \int_0^1 \int_{\mathbb{R}^d} \left( \sqrt{G_1} - \sqrt{G_0} \right) : \left( t\sqrt{G_1} + (1 - t)\sqrt{G_0} \right) \Phi_t d\mathcal{L}^d dt$$
for all \((\Phi, \phi) \in L^2(0,1; L^2(dG))\). It is not difficult to see that the constraint \(\mathcal{P}^2\) is satisfied. By definition, the energy of this path is \(\|\mathcal{U}\|_{L^2(0,1; L^2(dG))}^2\). By the Cauchy-Schwarz inequality and \((1.2)\),

\[
\langle \mathcal{U}, (\Phi, \phi) \rangle_{L^2(0,1; L^2(dG))}^2 \\
\leq 4 \int_0^1 \int_{\mathbb{R}^d} (\sqrt{G_1} - \sqrt{G_0}) \, d\mathcal{L}^d \, dt \\
\Phi_t \\
\left( t \sqrt{G_1} + (1-t) \sqrt{G_0} \right) \Phi_t \, d\mathcal{L}^d dt
\]

\[
\leq 4 \|\Phi, 0\|_{L^2(0,1; L^2(dG))}^2 \int_0^1 \int_{\mathbb{R}^d} (\sqrt{G_1} - \sqrt{G_0}) : (\sqrt{G_1} - \sqrt{G_0}) \, d\mathcal{L}^d dt
\]

\[
\leq 4 \|\Phi, 0\|_{L^2(0,1; L^2(dG))}^2 \int_0^1 \int_{\mathbb{R}^d} (Tr G_1 + Tr G_0) \, d\mathcal{L}^d dt.
\]

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

**Definition 3.3** (cf. \([7]\)). We say that two points \(x, y\) in a metric space \((X, \rho)\) almost admit a midpoint if there exists a sequence \(\{z_k\} \subset X\) such that

\[
|\rho(x, y) - 2\rho(x, z_k)| \leq k^{-1}, \quad |\rho(x, y) - 2\rho(y, z_k)| \leq k^{-1}.
\]

**Theorem 4.** \((\mathbb{P}^+, d_{HK})\) 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 \(d_{HK}\)-Lipschitz curve \(G\) such that \(d_{HK}(G_t, G_s) = |t - s|d_{HK}(G_0, G_1)\) and a potential \(\mathcal{U} \in L^2(0,1; L^2(dG_t))\) such that \(\|\mathcal{U}\|_{L^2(dG_t)} = cst = d_{HK}(G_0, G_1)\) 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 \(2.4\) and the (classical) Banach-Alaoglu theorem, \(d_{HK}\)-bounded sequences contain weakly-* converging subsequences. Now Lemma \(3.3\) (analogue of the Hopf-Rinow theorem for non-complete metric spaces) together with Theorem \(3\) imply that \((\mathbb{P}^+, d_{HK})\) 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 \([2, 7]\). \(\square\)

The next theorem gives some insight into the geometry of the space \((\mathbb{P}^+, d_{HK})\).

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

\[
g(A) = AG_* A.
\]

Then for any pair of commuting matrices \(A_0, A_1 \in \mathbb{P}^+\) one has

\[
d_{HK}^2(g(A_0), g(A_1)) = 4 \int_{\mathbb{R}^d} dG_*(A_1 - A_0) : (A_1 - A_0),
\]
and a geodesic between \( g(A_0) \) and \( g(A_1) \) is explicitly given by
\[(3.7) \quad \tilde{G}_t := g(tA_1 + (1-t)A_0).\]

**Proof. Step 1.** Define a potential \( \tilde{\Phi}_t \in L^2(0,1; L^2(d\tilde{G})) \) by Riesz duality as
\[(3.8) \quad \langle \tilde{\Phi}, (\Phi, \phi) \rangle_{L^2(0,1; L^2(d\tilde{G}))} = 2 \int_0^1 \int_{\mathbb{R}^d} dG_s(A_1 - A_0) : (tA_1 - (1-t)A_0) \Phi_t \, dt \]
for all \((\Phi, \phi) \in L^2(0,1; L^2(d\tilde{G}))\). A straightforward computation shows that \((\tilde{G}_t, \tilde{\Phi}_t)\) satisfies the constraint \((2.2)\). The energy of this path coincides with \( \|\tilde{\Phi}\|^2_{L^2(0,1; L^2(d\tilde{G}))} \).

By the Cauchy-Schwarz inequality,
\[
\langle \tilde{\Phi}, (\Phi, \phi) \rangle_{L^2(0,1; L^2(d\tilde{G}))}^2 \\
\leq 4 \int_0^1 \int_{\mathbb{R}^d} dG_s(A_1 - A_0) : (A_1 - A_0) \int_{\mathbb{R}^d} dG_s(tA_1 + (1-t)A_0) \Phi_t : (tA_1 + (1-t)A_0) \Phi_t \, dt \\
\leq 4 \|\Phi, \phi\|^2_{L^2(0,1; L^2(d\tilde{G}))} \int_0^1 \int_{\mathbb{R}^d} dG_s(A_1 - A_0) : (A_1 - A_0) \, dt.
\]

Thus, the energy \( E[\tilde{G}_t, \tilde{\Omega}_t] \) 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 \((A_0, A_1) \in \mathcal{P}^+ \) would immediately follow by letting \( \epsilon \to 0_+ \) in the triangle inequality
\[d_{HK}(g(A_0), g(A_1)) \geq d_{HK}(g(A_0 + \epsilon I), g(A_1)) - d_{HK}(g(A_0 + \epsilon I), g(A_0)).\]

**Step 3.** Consider any admissible path \((G_t, \Omega_t)_{t \in [0,1]}\) connecting \( G_0 := g(A_0) \) to \( G_1 := g(A_1) \). Let \( \lambda \) be any scalar probability measure on \( \mathbb{R}^d \). Set \( \tilde{G}_t := \lambda \int_{\mathbb{R}^d} dG_t \), and define \( \tilde{\Omega}_t \in L^2(0,1; L^2(d\tilde{G})) \) by duality as
\[(3.9) \quad \langle \tilde{\Omega}, (\Phi, \phi) \rangle_{L^2(0,1; L^2(d\tilde{G}))} = \int_0^1 \int_{\mathbb{R}^d} dG_t U_t \int_{\mathbb{R}^d} \Phi_t d\lambda \, dt
\]
for all \((\Phi, \phi) \in L^2(0,1; L^2(d\tilde{G}))\). Then \((\tilde{G}, \tilde{\Omega})\) is an admissible path (joining \( A_0 \lambda \int_{\mathbb{R}^d} dG_s A_0 \) and \( A_1 \lambda \int_{\mathbb{R}^d} dG_s A_1 \)). We claim that it has lesser energy than \((G, \Omega)\). To prove the claim, we approximate this path with the sequence \( \tilde{G}^k_t := \lambda \int_{\mathbb{R}^d} (k^{-1}I + dG_t) \); the corresponding potentials are
\[(3.10) \quad \langle \tilde{\Phi}^k, (\Phi, \phi) \rangle_{L^2(0,1; L^2(d\tilde{G}^k))} = \int_0^1 \int_{\mathbb{R}^d} dG_t U_t \int_{\mathbb{R}^d} \Phi_t d\lambda \, dt.
\]
Let us equip the linear space \( \mathbb{R}^{d \times d} \) with the scalar product
\[
(B, B)_{k,t} = k^{-1}B : B + \int_{\mathbb{R}^d} dG_t B : B,
\]
and let $\Pi_{k,t}$ be the orthogonal projection onto the subspace $\mathcal{S}$. One explicitly computes that

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

Then

$$E[\tilde{G}^k; \tilde{U}^k] = \int_0^1 \int_{\mathbb{R}^d} d\lambda \left( k^{-1} I + \int_{\mathbb{R}^d} dG_t \right) \tilde{U}_t^k : \tilde{U}_t^k \ dt = \int_0^1 (\tilde{U}_t^k, \tilde{U}_t^k)_{k,t} \ dt$$

$$\leq \int_0^1 \left( \left( k^{-1} I + \int_{\mathbb{R}^d} dG_t \right)^{-1} \int_{\mathbb{R}^d} dG_t U_t, \left( k^{-1} I + \int_{\mathbb{R}^d} dG_t \right)^{-1} \int_{\mathbb{R}^d} dG_t U_t \right)_{k,t} dt$$

$$\leq \int_0^1 \left( k^{-1} I + \int_{\mathbb{R}^d} dG_t \right) \left( \left( k^{-1} I + \int_{\mathbb{R}^d} dG_t \right)^{-1} \int_{\mathbb{R}^d} dG_t U_t \right) dt$$

$$+ \int_0^1 dG_t \left( U_t - \left( k^{-1} I + \int_{\mathbb{R}^d} dG_t \right)^{-1} \int_{\mathbb{R}^d} dG_t U_t \right) dt$$

$$= \int_0^1 \int_{\mathbb{R}^d} dG_t U_t : U_t \ dt$$

$$- k^{-1} \int_0^1 \left( \left( k^{-1} I + \int_{\mathbb{R}^d} dG_t \right)^{-1} \int_{\mathbb{R}^d} dG_t U_t \right) : \left( \left( k^{-1} I + \int_{\mathbb{R}^d} dG_t \right)^{-1} \int_{\mathbb{R}^d} dG_t U_t \right) dt 
\leq E[G; \mathcal{U}].$$

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

**Step 4.** Obviously, the right-hand side of (3.3) does not change if we replace $G_*$ by $\lambda D$, where $D := \int_{\mathbb{R}^d} dG_* \in \mathcal{P}^+$, and $\lambda$ is as above. Thus, by the previous steps it is enough to check that the energies of the admissible paths of the form $G_t = \lambda F_i$ with $F_i \in \mathcal{P}^+$, $F_i = A_i D A_i$, $i = 0, 1$, $A_0 \in \mathcal{P}^{++}$, with constant-in-space potentials $\mathcal{U}_t = (U_t, 0) \in L^2(0, 1; L^2(dG))$ 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
$$U_t = 2(tA_1 + (1-t)A_0)^{-1}(A_1 - A_0), \ t \neq 1.$$ The corresponding energy is exactly 
$$4D(A_1 - A_0) : (A_1 - A_0). \ \Box$$

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

**Remark 3.5.** 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 [3]

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

Then

$$(3.11) \langle \Xi_1, \Xi_2 \rangle_P := PU_{\Xi_1} : U_{\Xi_2}$$

is a Riemannian metric on \( \mathcal{P}^{++} \). 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 [3]

$$XP_0X = P_1.$$

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 \( \lambda \) 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 [23]). Moreover, in view of Remark [24] \( d_R(P_0, P_1) = d_{HK}(P_0\lambda, P_1\lambda) = d_H(P_0\lambda, P_1\lambda). \)

The midpoint \( \frac{1}{2}(I + X)P_0(I + X) \) can serve to define an allegedly new matrix mean \( P_0P_1 \), which may be dubbed the Hellinger mean. More conventional matrix means are discussed in [3].

### 4. The spherical distance and the conic structure

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

**Definition 4.1.** Given two matrix measures \( G_0, G_1 \in \mathbb{P}^1 \) we define

$$(4.1) \ d_{SHK}^2(G_0, G_1) := \inf_{A_t(G_0,G_1)} \int_0^1 \left( \int_{\mathbb{R}^d} dG_t(x)u_t(x) \cdot u_t(x) + \int_{\mathbb{R}^d} dG_t(x)U_t(x) : U_t(x) \right) dt.$$
where the admissible set $\mathcal{A}_1(G_0, G_1)$ consists of all couples $(G_t, \mathcal{U}_t)_{t \in [0, 1]}$ such that

$$
\begin{align*}
G & \in C_w([0, 1]; \mathbb{P}^1), \\
G|_{t=0} = G_0, & G|_{t=1} = G_1, \\
\mathcal{U} & \in L^2(0, T; L^2(dG_t)), \\
\partial_t G_t & = \{-\nabla(G_t u_t) + G_t U_t\}^{\text{sym}} \quad \text{in the weak sense.}
\end{align*}
$$

**Proposition 4.2.** $d_{SHK}$ 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 $d_{SHK} \geq d_{HK}$ on $\mathbb{P}^1$.

**Remark 4.3.** 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

$$
(4.2) \quad d_{SHK}^2(G_0, G_1) := \inf_{\mathcal{A}_2(G_0, G_1)} \int_0^1 \left( \int_{\mathbb{R}^d} dG_t(x) u_t(x) \cdot u_t(x) + \int_{\mathbb{R}^d} dG_t(x) U_t(x) : U_t(x) - \left( \int_{\mathbb{R}^d} dG_t(x) : U_t(x) \right)^2 \right) dt.
$$

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

$$
\begin{align*}
G & \in C_w([0, 1]; \mathbb{P}^1), \\
G|_{t=0} = G_0, & G|_{t=1} = G_1, \\
\mathcal{U} & \in L^2(0, T; L^2(dG_t)), \\
\partial_t G_t + G_t \int_{\mathbb{R}^d} dG_t : U_t = \{-\nabla(G_t u_t) + G_t U_t\}^{\text{sym}} \quad \text{in the weak sense.}
\end{align*}
$$

Indeed, $\mathcal{A}_1(G_0, G_1) = \mathcal{A}_2(G_0, G_1) \cap \left[ \int_{\mathbb{R}^d} dG_t : U_t \equiv 0 \right]$, 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 $(G_t, \mathcal{U}_t)$, where $\mathcal{U}_t \in L^2(0, T; L^2(dG_t))$ is defined by duality via

$$
(4.3) \quad \left< \mathcal{U}, (\Phi, \phi) \right>_{L^2([0, 1]; L^2(dG_t))} = \int_0^1 \left( \int_{\mathbb{R}^d} dG_t(x) u_t(x) \cdot \phi_t(x) + \int_{\mathbb{R}^d} dG_t(x) \left[ U_t(x) - \int_{\mathbb{R}^d} dG_t(y) : U_t(y) \right] : \Phi_t(x) \right) dt.
$$

We recall [6, 7] that, given a metric space $(X, d_X)$ of diameter $\leq \pi$, one can define another metric space $(\mathcal{C}(X), d_{\mathcal{C}(X)}))$, called a cone over $X$, in the following manner. Consider the quotient $\mathcal{C}(X) := X \times [0, \infty)/X \times \{0\}$, that is, all points of the fiber $X \times \{0\}$ constitute a single point of the cone which is called the apex. Now set

$$
(4.4) \quad d_{\mathcal{C}(X)}([x_0, r_0], [x_1, r_1]) := r_0^2 + r_1^2 - 2r_0r_1 \cos(d_X(x_0, x_1)).
$$
The cones enjoy neat scaling and other nice geometric properties \cite{18}. A particularly regular situation appears when the diameter of $X$ is strictly less than $\pi$, since in this case there is a one-to-one correspondence between the geodesics in $X$ and $\mathcal{C}(X)$. Given a cone $Y = \mathcal{C}(X)$, $X$ may be referred to as the sphere in $Y$.

**Lemma 4.4.** If $X$ is a length space, and $Y = \mathcal{C}(X)$, 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 $d_X(x_+, x_-) \geq d_X([x_1], [x_-, 1])$ 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

$$L_Y([x_t, 1]) \geq qL_X(x_t)$$

for any $q < 1$. By linearity of (4.5), it is enough to prove it for curves of sufficiently small length. From \cite[Ex. 3.6.14]{7} we infer that

$$L_Y([x_t, 1]) \geq 2 \sin \left(\frac{L_X(x_t)}{2}\right),$$

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

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

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

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

**Theorem 6.** i) The space $(\mathbb{P}^1, d_{SHK})$ is a geodesic space of diameter $\leq \pi$.

ii) $(\mathbb{P}^+, d_{HK}/2)$ is a metric cone over $(\mathbb{P}^1, d_{SHK}/2)$, where $\mathbb{P}^+$ is identified with $\mathcal{C}(\mathbb{P}^1)$ via $G \leftrightarrow [G/r^2, r]$.

**Proof.** Step 1. We first observe that it suffices to show that $(\mathbb{P}^+, d_{HK}/2)$ 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

$$d_{HK}(G_0, G_1)/2 \leq \sqrt{2}.$$  

Since $(\mathbb{P}^+, d_{HK}/2)$ 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 \cite[Corollary 5.11]{6}, $(\mathbb{P}^1, d)$ is a geodesic space. Evoking Lemma 4.4 and Definition 4.1 we see that $d$ actually coincides with $d_{SHK}/2$. 


Step 2. In view of (4.6) and [18, Theorem 2.2], it suffices to show the following scaling property which characterizes the cones:

\[(4.7)\]

\[d_{HK}^2(r_0^2 G_0, r_1^2 G_1) = r_0 r_1 d_{HK}^2(G_0, G_1) + 4(r_0 - r_1)^2,\]

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 $\Phi = \Phi_a = (r_0 + (r_1 - r_0)t(a))I$:

$$E[\tilde{G}_t; \tilde{U}_t, \tilde{u}_t] = \int_0^1 \left( \int_{\mathbb{R}^d} d\tilde{G}_t \tilde{u}_t \cdot \tilde{u}_t + \int_{\mathbb{R}^d} d\tilde{G}_t \tilde{U}_t : \tilde{U}_t \right) dt$$

$$= r_0 r_1 \int_0^1 a'(t) \left( \int_{\mathbb{R}^d} dG_{a(t)} u_{a(t)} \cdot u_{a(t)} + \int_{\mathbb{R}^d} dG_{a(t)} U_{a(t)} : U_{a(t)} \right) dt$$

$$+ 4(r_1 - r_0) \int_0^1 a'(t)(r_0 + (r_1 - r_0)t) \int_{\mathbb{R}^d} dG_{a(t)} : U_{a(t)} dt$$

$$+ 4(r_1 - r_0)^2 \int_0^1 \int_{\mathbb{R}^d} dG_{a(t)} : I \, dt$$

$$= r_0 r_1 \int_0^1 \left( \int_{\mathbb{R}^d} dG_{a} u_{a} \cdot u_{a} + \int_{\mathbb{R}^d} dG_{a} U_{a} : U_{a} \right) \, da$$

$$+ 4(r_1 - r_0) \int_0^1 (r_0 + (r_1 - r_0)t(a)) \int_{\mathbb{R}^d} dG_{a} : U_{a} \, da$$

$$+ 4(r_1 - r_0)^2 \int_0^1 t'(a) \int_{\mathbb{R}^d} dG_{a} : I \, da$$

$$= r_0 r_1 E[G_t; U_t, U_t]$$

$$+ 4(r_1 - r_0)(r_0 + (r_1 - r_0)t(1)) \int_{\mathbb{R}^d} dG_1 : I$$

$$- 4(r_1 - r_0)(r_0 + (r_1 - r_0)t(0)) \int_{\mathbb{R}^d} dG_0 : I$$

$$= r_0 r_1 E[G_t; U_t, U_t] + 4(r_0 - r_1)^2.$$

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

**Appendix A. Frame-indifference**

The principle of material frame-indifference [26] 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

$$t^* = t - t_0,$$

$$x^* = c^*(t) + Q_t x,$$
where \( t_0 \in \mathbb{R}, c^* : \mathbb{R} \to \mathbb{R}^d \), \( Q_t \) 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 \( w^* = Q_tw \). 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

\[
T^*(t^*, x^*) = Q_t T(t, x) Q_t^T.
\]

We claim that our distance \( d_{HK} \) complies with the frame indifference:

\[
(A.1) \quad d_{HK}(T_0^*(t^*, x^*), T_1^*(t^*, x^*)) = d_{HK}(T_0(t, x), T_1(t, x)).
\]

In other words, \( d_{HK} \) 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 \((T_t, U_t, u_t)(t, x)\) in the old frame, the path

\[
(T_t^*, U_t^*, u_t^*)(t^*, x^*) := (Q_t T_t(t, x) Q_t^T, Q_t U_t(t, x) Q_t^T, Q_t u_t(t, x))
\]

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:

\[
\nabla x^* w^* = Q_t (\nabla x w) Q_t^T
\]

provided \( w^* = Q_tw \), which is just a manifestation of the chain rule, cf., e.g., [26, 29].

Appendix B. Some technical facts

**Proposition B.1** ([17]). Let \((X, \| \cdot \|)\) be a separable normed vector space. Assume that there exists a sequence of seminorms \( \{\| \cdot \|_k\} (k = 0, 1, 2, \ldots) \) on \( X \) such that for every \( x \in X \) one has

\[
\|x\|_k \leq C\|x\|
\]

with a constant \( C \) independent of \( k, x \), and

\[
\|x\|_k \to \|x\|_0.
\]

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

\[
c_k := \|\varphi_k\|_{(X, \| \cdot \|_k)^*} \leq C.
\]

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

\[
(B.1) \quad \|\varphi_0\|_{(X, \| \cdot \|_0)^*} \leq c_0 := \lim \inf_k c_k.
\]
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 $C_0$.

Lemma B.3. Let $(X, \varrho)$ be a metric space where every two points almost admit a midpoint. Assume that there exists a Hausdorff topology $\sigma$ on $X$ such that $\varrho$-bounded sequences contain $\sigma$-converging subsequences, and $\varrho$ is sequentially lower semicontinuous with respect to $\sigma$. Then $(X, \varrho)$ 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
\[ \varrho(x_t, x_t) \leq |t - \bar{t}|\varrho(x_0, x_1). \]
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,
\[ \varrho(x, y) = 2\varrho(x, z) = 2\varrho(z, y). \]
for some $z \in X$. Indeed, take any sequence $z_k$ of almost midpoints, i.e.,
\[ |\varrho(x, y) - 2\varrho(x, z_k)| \leq k^{-1}, \quad |\varrho(x, y) - 2\varrho(y, z_k)| \leq k^{-1}. \]
The sequence $\{z_k\}$ is $\varrho$-bounded, thus without loss of generality it $\sigma$-converges to some $z \in X$. Then
\[ 2\varrho(x, z) \leq \lim_{k \to \infty} 2\varrho(x, z_k) = \varrho(x, y), \]
\[ 2\varrho(y, z) \leq \lim_{k \to \infty} 2\varrho(y, z_k) = \varrho(x, y). \]
But it 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 [7, p. 43] one constructs points $x_s (s \in Q)$ satisfying (B.2), that is, the function $s \mapsto x_s$ is $\varrho(x_0, x_1)$-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 $\varrho$-Cauchy sequence. Therefore it is $\varrho$-bounded, and admits a subsequence which $\sigma$-converges to some $x_t \in X$. Due to the sequential lower semicontinuity of the distance $\varrho$, we can pass to the limit in (B.2) for all $t, \bar{t} \in [0, 1]$.

Lemma B.4. Let $(X, \varrho)$ be a metric space. Assume that there exists a Hausdorff topology $\sigma$ on $X$ such that $\varrho$ is sequentially lower semicontinuous with respect to $\sigma$. Let $(x^t), t \in [0, 1]$, be a sequence of curves lying in a common $\sigma$-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
\[ \varrho((x^t), (x^\bar{t})) \leq \omega(t, \bar{t}). \]
for all \( t, \bar{t} \in [0, 1] \). Then there exists a \( \rho \)-continuous curve \( x_t \) such that
\[
\rho(x_t, x_{\bar{t}}) \leq \omega(t, \bar{t}),
\]
and (up to a not relabelled subsequence)
\[
(x^k)_t \rightarrow x_t
\]
for all \( t \in [0, 1] \) in the topology \( \sigma \).

**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 \( \rho \), 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 \( \rho \)-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 [17] 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. MATRICAL OTTO CALCULUS**

We have seen in Remark [B.5] that some pieces of \((\mathbb{P}^+, d_{HK})\) 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 \( d_{HK} \). Namely, we can develop some kind of Otto calculus, cf. [24, 27, 28], on \((\mathbb{P}^+, d_{HK})\). Starting from this point, we are completely formal. As we observed in Remark 2.2, the minimizing potentials in (2.1) can be chosen in the form \( \mathfrak{U} = (U, \text{div } U) \). This suggests to define the tangent spaces as
\[
T_{G\mathbb{P}^+} := \left\{ \exists U(x) \in \mathcal{S} : \Xi = (-\nabla (G \text{ div } U) + GU)^{\text{Sym}} \right\}
\]
and
\[
\|\Xi\|_{T_{G\mathbb{P}^+}} = \|U\|_{H^1_{\text{div}}(dG; \mathcal{S})} := \left( \int_{\mathbb{R}^d} dG \text{ div } U \cdot \text{ div } U + \int_{\mathbb{R}^d} dGU : U \right)^{1/2}.
\]
Ignoring all smoothness issues, the operator
\[
\Xi(U) = (-\nabla (G \text{ div } U) + GU)^{\text{Sym}}
\]
is \( H^1_{\text{div}}(dG; \mathcal{S}) \)-coercive, so the one-to-one correspondence between the tangent vectors \( \Xi \) and potentials \( \mathfrak{U} = (U, \text{div } U) \) is well defined. By polarization this defines a
Riemannian metric on $T\mathbb{P}^+$, and
\[ d_{HK}^2(G_0, G_1) = \inf \left\{ \int_0^1 \left\| \frac{dG_t}{dt} \right\|_{T_{G_t}\mathbb{P}^+}^2 \, dt \right\}. \]

The gradients of functionals $F : \mathbb{P}^+ \to \mathbb{R}$ are given by
\[ \text{grad}_{HK} F(G) = \left[ -\nabla \left( G \text{ div } \frac{\delta F}{\delta G} \right) + G^2 \left( \frac{\delta F}{\delta G} \right)^{\text{sym}} \right], \tag{C.2} \]
where $\frac{\delta F}{\delta G}$ 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
\[ \partial_t G = \left[ \nabla \left( G \text{ div } \frac{\delta F}{\delta G} \right) - G^2 \left( \frac{\delta F}{\delta G} \right)^{\text{sym}} \right]. \]
The interesting driving functionals include the von Neumann entropy
\[ F_N(G) = \int_{\mathbb{R}^d} G \log G - G \]
and the “volume”
\[ F_V(G) = \int_{\mathbb{R}^d} \sqrt{\det G}. \]
The gradient flow of $F_N$ is a sort of matricial “heat flow” with logarithmic reaction. Indeed, for $d = 1$ it simply becomes $\partial_t G = G_{xx} - G \log G$. The gradient flow of $F_V$ 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 $(\mathbb{P}^1, d_{SHK})$. Remark 4.3 guides us to define
\[ T_G\mathbb{P}^1 := \left\{ \exists U(x) \in \mathcal{S} : \Xi = \left[ -\nabla (G \text{ div } U) + GU \right]^{\text{sym}} - G \int_{\mathbb{R}^d} G : U \right\} \]
and
\[ \|\Xi\|_{T_G\mathbb{P}^1} = \left( \int_{\mathbb{R}^d} dG \text{ div } U \cdot \text{ div } U + \int_{\mathbb{R}^d} dGU : U - \left( \int_{\mathbb{R}^d} dG : U \right)^2 \right)^{1/2}. \]
The gradients of functionals $F : \mathbb{P}^1 \to \mathbb{R}$ are
\[ \text{grad}_{SHK} F(G) = \left[ -\nabla \left( G \text{ div } \frac{\delta F}{\delta G} \right) + G^2 \left( \frac{\delta F}{\delta G} \right)^{\text{sym}} \right] - G \int_{\mathbb{R}^d} G : \frac{\delta F}{\delta G}. \tag{C.3} \]
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 $\lambda$-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] R. Bhatia. Positive definite matrices. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2007.
[4] V. I. Bogachev. Measure theory. Vol. II. Springer-Verlag, Berlin, 2007.
[5] Y. Brenier. The initial value problem for the Euler equations of incompressible fluids viewed as a concave maximization problem. ArXiv e-prints, June 2017.
[6] 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.
[7] D. Burago, Y. Burago, and S. Ivanov. A course in metric geometry. AMS, 2001.
[8] 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.
[9] E. A. Carlen and J. Maas. Gradient flow and entropy inequalities for quantum Markov semi-groups with detailed balance. J. Funct. Anal., 273(5):1810–1869, 2017.
[10] Y. Chen, W. Gangbo, T. T. Georgiou, and A. Tannenbaum. On the matrix monge-kantorovich problem. arXiv preprint arXiv:1701.02826, 2017.
[11] Y. Chen, T. Georgiou, and A. Tannenbaum. Interpolation of matrices and matrix-valued densities: The unbalanced case. European Journal of Applied Mathematics, 2018.
[12] Y. Chen, T. T. Georgiou, and A. Tannenbaum. Matrix optimal mass transport: a quantum mechanical approach. IEEE Transactions on Automatic Control, 2017.
[13] 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.
[14] 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.
[15] R. M. Dudley. Real analysis and probability. Cambridge University Press, 2002.
[16] 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.
[17] 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.
[18] 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. ArXiv e-prints, Dec. 2017.
[19] 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.

[20] 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.

[21] 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.

[22] 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.

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

[24] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.

[25] G. Peyré, L. Chizat, F.-X. Vialard, and J. Solomon. Quantum optimal transport for tensor field processing. *arXiv preprint arXiv:1612.08731*, 2016.

[26] 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.

[27] C. Villani. *Topics in optimal transportation*. American Mathematical Soc., 2003.

[28] C. Villani. *Optimal transport: old and new*. Springer Science & Business Media, 2008.

[29] 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.