Euro. Jnl of Applied Mathematics: page 1 of 27 © Cambridge University Press 2019.
doi:10.1017/S0956792519000172

On the matrix Monge–Kantorovich problem†

YONGXIN CHEN[1], WILFRID GANGBO[2], TRYPHON T. GEORGIOU[3]
and ALLEN TANNENBAUM[2]

[1]*School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, USA*
email: yongchen@gatech.edu
[2]*Department of Mathematics, UCLA, Los Angeles, CA, USA*
email: wgangbo@math.ucla.edu
[3]*Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA, USA*
email: tryphon@uci.edu
[4]*Departments of Computer Science and Applied Mathematics & Statistics, Stony Brook University, NY, USA*
email: allen.tannenbaum@stonybrook.edu

(*Received 30 October 2018*; *revised 1 June 2019*; *accepted 15 July 2019*)

The classical Monge–Kantorovich (MK) problem as originally posed is concerned with how best to move a pile of soil or rubble to an excavation or fill with the least amount of work relative to some cost function. When the cost is given by the square of the Euclidean distance, one can define a metric on densities called the *Wasserstein distance*. In this note, we formulate a natural matrix counterpart of the MK problem for positive-definite density matrices. We prove a number of results about this metric including showing that it can be formulated as a convex optimisation problem, strong duality, an analogue of the Poincaré–Wirtinger inequality and a Lax–Hopf–Oleinik-type result.

Key words: Optimal mass transport, quantum mechanics, Wasserstein distance

2010 Mathematics Subject Classification: Primary: 46A20; Secondary: 49N15; 81Q35

1 Introduction

The mass transport problem was first formulated by Monge in 1781 and concerned finding the optimal way, in the sense of minimal transportation cost, of moving a pile of soil from one site to another. This problem of *optimal mass transport* (OMT) was given a modern formulation in the work of Kantorovich, and so is now known as the Monge–Kantorovich (MK) problem; see [21, 23] and the many references therein. As originally formulated, the problem is *static*. Namely, given two probability densities, one can define a metric, now known as the *Wasserstein distance*, that quantifies the cost of transport and enjoys a number of remarkable properties as described in [21, 23]. Optimal mass transport is a very active area of research with applications to numerous disciplines including probability, econometrics, fluid dynamics, automatic control, transportation, statistical physics, shape optimisation, expert systems and meteorology.

† This project was supported by AFOSR grants (FA9550-17-1-0435 and FA9550-18-1-0502), grants from the National Center for Research Resources (P41-RR-013218) and the National Institute of Biomedical Imaging and Bioengineering (P41-EB-015902), National Science Foundation grants (DMS-1160939, 1665031, 1807664, 1839441 and 1901599), grants from the National Institutes of Health (1U24CA18092401A1 and R01-AG048769) and a postdoctoral fellowship through Memorial Sloan Kettering Cancer Center.

A major development in optimal mass transport theory was realised in the seminal dynamic approach by Benamou and Brenier [7]. These authors base their approach on ideas from fluid mechanics via the minimisation of a kinetic energy functional subject to a continuity constraint. Indeed, the work of Benamou and Brenier [7], together with a number of groundbreaking contributions that took place in the late 1990s and shortly thereafter, helped launch the modern fast developing phase of the theory of optimal mass transport. See [15, 10, 13, 20, 23] and the references therein.

The purpose of the present work is to develop a certain non-commutative counterpart of the Benamou–Brenier theory that gives rise to a corresponding metric structure on the space of positive-definite matrices. Indeed, our work is based on a continuity equation that imposes a 'mass preservation' constraint on a quadratic 'kinetic energy'. We then study a convex optimisation problem that leads to a certain Riemannian structure on densities matrices and generalises the work of Otto [20] to the current non-commutative setting.

Early motivation for studying such OMT-like geometries for matrices and matrix-valued distributions has its roots in the theory of vector-valued stochastic processes [19]. However, soon after this early attempt, it became apparent that deeper connections can be sought between optimal mass transport, non-commutative geometry and quantum mechanics [18, 2]. The first work proposing a non-commutative version of OMT based on a continuity equation based on [7] was by Carlen and Maas [2].

Building on these earlier works, Chen et al. [4] developed non-commutative optimal transport where density matrices ρ (i.e., Hermitian matrices that are positive definite and have unit trace) replace probability distributions, and where 'transport' corresponds to a flow on the space of such matrices that minimises a corresponding action integral, thereby giving a fluid dynamical flavour to matrix OMT in the style of [7]. At the same time and in parallel with [4], similar approaches for matrix OMT were carried out independently by Carlen and Maas [3] and Mittnenzweig and Mielke [17]. However, these works aimed at entropic gradient flows with respect to the so-constructed Wasserstein metric and are quite distinct from the theory presented herein. Indeed, our theory relies on quadratic functionals that are more amenable to computation and more suitable for signal and image processing applications (e.g., smoothing of matrix-valued power spectra and Diffusion Tensor Imagery in radiology). For numerical implementations and image processing examples, we refer the reader to [6].

The present paper is a continuation of [4] in which a number of the results are given rigorous mathematical proofs. Since much of paper involves rather technical arguments, we will summarise the main results here. First of all, in Section 2, we set up the necessary background for defining our version of the Wasserstein 2-metric on positive density matrices. The key is a functional inspired by the kinetic energy formulation of Benamou–Brenier [7], and a corresponding continuity equation. Explicitly, the functional in [7] involves $\rho|v|^2$. Benamou and Brenier make a change in coordinates via taking the momentum $m = \rho v$ and getting the convex function $\frac{|m|^2}{\rho}$, to which they can apply methods of convex optimisation and duality. In our formulation, the non-convex function trace$(\rho v^* v)$ (see Section 2 for the precise definitions) plays the role of kinetic energy. Accordingly, we define in Sections 5 and 6, a convex function $F(\rho, m)$ that *exactly plays the role* of $\frac{|m|^2}{\rho}$ in [7]. We are now in the realm of convex optimisation for our non-commutative Wasserstein framework, and we can derive a strong duality result (see (7.2), (7.3) and Theorem 22 in Section 7). Some of the other key contributions include an analogue of the Poincaré–Wirtinger inequality in Section 3 that is needed in some of the proofs. Finally, in

Section 8, we prove the constancy of F along optimal trajectories, give a rigorous proof of the Riemannian structure on the space of positive matrix densities using our matrix Wassterstein metric and prove an Lax–Oleinik–Hopf result (Theorem 23). These results were originally reported in January 2017 in an arxiv preprint [5].

2 Continuity equation and Wasserstein distance

In this section, we set up the continuity equation that is the basis for our formulation of the Wasserstein distance for density matrices. We follow closely the recent paper [4]. In that work, an approach is developed based on the *Lindblad equation* which describes the evolution of open quantum systems. Open quantum systems are thought of as being coupled to a larger system (heat bath), and thus cannot in general be described by a wave function and a unitary evolution. The proper description is in terms of a density operator ρ [12] which in turn obeys the Lindblad equation where we assume $\hbar = 1$:

$$\dot{\rho} = -i\,[H, \rho] + \sum_{k=1}^{N} \left(L_k \rho L_k^* - \frac{1}{2}\rho L_k^* L_k - \frac{1}{2} L_k^* L_k \rho \right). \tag{2.1}$$

Here, * as superscript denotes conjugate transpose and $[H, \rho] := H\rho - \rho H$ denotes the commutator. The first term on the right-hand side describes the evolution of the state under the effect of the Hamiltonian H, and it is unitary (energy preserving), while the other terms on the right-hand side model diffusion and, thereby, capture the dissipation of energy; these dissipative terms together represent the quantum analogue of the Laplacian operator Δ (as it will become clear shortly). The Lindblad equation defines a non-unitary evolution of the density and the calculus we develop next actually underscores parallels with classical diffusion and the Fokker–Planck equation.

Denote by $\mathcal{H}$ and $\mathcal{S}$ the set of $n \times n$ Hermitian and skew-Hermitian matrices, respectively. We will assume that all of our matrices are fixed to be $n \times n$. Next, we denote the space of block-column vectors consisting of N elements in $\mathcal{S}$ and $\mathcal{H}$ as $\mathcal{S}^N$ and $\mathcal{H}^N$, respectively. We let $\mathcal{H}_+$ and $\mathcal{H}_{++}$ denote the cones of non-negative and positive-definite matrices, respectively, and

$$\mathcal{D} := \{\rho \in \mathcal{H}_+ \mid \mathrm{tr}\,(\rho) = 1\}, \tag{2.2}$$

$$\mathcal{D}_+ := \{\rho \in \mathcal{H}_{++} \mid \mathrm{tr}\,(\rho) = 1\}. \tag{2.3}$$

Note that the elements of $\mathcal{D}_+$ are strictly positive. Next, the tangent space of $\mathcal{D}_+$ at any $\rho \in \mathcal{D}_+$ is given by

$$T_\rho = \{\delta \in \mathcal{H} \mid \mathrm{tr}(\delta) = 0\}, \tag{2.4}$$

and we use the standard notion of inner product, namely

$$\langle X; Y \rangle = \mathrm{tr}(X^* Y),$$

for both $\mathcal{H}$ and $\mathcal{S}$. For $X, Y \in \mathcal{H}^N$ ($\mathcal{S}^N$), respectively,

$$\langle X; Y \rangle = \sum_{k=1}^{N} \mathrm{tr}(X_k^* Y_k).$$

Given $X = [X_1^*, \cdots, X_N^*]^* \in \mathcal{H}^N$ ($\mathcal{S}^N$), $Y \in \mathcal{H}$ ($\mathcal{S}$), set

$$XY = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix} Y := \begin{bmatrix} X_1 Y \\ \vdots \\ X_N Y \end{bmatrix},$$

and

$$YX = Y \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix} := \begin{bmatrix} YX_1 \\ \vdots \\ YX_N \end{bmatrix}.$$

If we assume that $L_k = L_k^*$, i.e., $L_k \in \mathcal{H}$ for all $k \in 1, \ldots, N$, then we can define

$$\nabla_L : \mathcal{H} \to \mathcal{S}^N, \quad X \mapsto \begin{bmatrix} L_1 X - X L_1 \\ \vdots \\ L_N X - X L_N \end{bmatrix} \tag{2.5}$$

as the *gradient operator*. Note that ∇_L acts just like the standard gradient operator (or discrete derivation), and in particular, satisfies

$$\nabla_L(XY + YX) = (\nabla_L X)Y + X(\nabla_L Y) + (\nabla_L Y)X + Y(\nabla_L X), \quad \forall X, Y \in \mathcal{H}. \tag{2.6}$$

The dual of ∇_L with respect to $\langle ; \rangle$, which is an analogue of the (negative) *divergence operator*, is defined by

$$\nabla_L^* : \mathcal{S}^N \to \mathcal{H}, \quad Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix} \mapsto \sum_k^N L_k Y_k - Y_k L_k. \tag{2.7}$$

The duality

$$\langle \nabla_L X; Y \rangle = \langle X; \nabla_L^* Y \rangle$$

follows by definition.

With these definitions, we define the (matricial) *Laplacian* as

$$\Delta_L X := -\nabla_L^* \nabla_L X = \sum_{k=1}^N (2 L_k X L_k^* - X L_k^* L_k - L_k^* L_k X), \quad X \in \mathcal{H},$$

which is exactly (after scaling by $1/2$) the diffusion term in the Lindblad equation (2.1). Therefore, Lindblad's equation (under the assumption that $L_k = L_k^*$) can be rewritten as

$$\dot{\rho} = -i[H, \rho] + \frac{1}{2} \Delta_L \rho,$$

i.e., as a continuity equation expressing flow under the influence of a suitable vector field.

In our case, we will consider a *continuity equation* of the form

$$\dot{\rho} = \nabla_L^* m,$$

for $m \in \mathcal{S}^N$ a suitable 'momentum field'. In particular, we are interested in the following family of continuity equations:

$$\dot{\rho} = \nabla_L^* M_\rho(v), \tag{2.8}$$

where the momentum field is expressed as a *non-commutative product* $M_\rho(v) \in \mathcal{S}^N$ between a 'velocity field' $v \in \mathcal{S}^N$ and the density matrix ρ.

Several such 'non-commutative products' have been considered (see [4]), however, in the present work, we consider the following case:

$$M_\rho(v) := \frac{1}{2}(v\rho + \rho v), \tag{2.9a}$$

which gives

$$\dot{\rho} = \frac{1}{2}\nabla_L^*(v\rho + \rho v) \tag{2.9b}$$

and $v = [v_1^*, \ldots, v_N^*]^* \in \mathcal{S}^N$. Clearly $v\rho + \rho v \in \mathcal{S}^N$, which is consistent with the definition of ∇_L^*. In [4], we call this the *anti-commutator* case, since

$$v\rho + \rho v =: \{v, \rho\}$$

is the anti-commutator when applied to elements of an associative algebra. In [4], another possibility is considered for the multiplication operator $M_\rho(v)$.

Given two density matrices $\rho_0, \rho_1 \in \mathcal{D}_+$, we formulate the optimisation problem (following [4])

$$W_2(\rho_0, \rho_1)^2 := \inf_{\rho \in \mathcal{D}_+, v \in \mathcal{S}^N} \int_0^1 \operatorname{tr}(\rho v^* v) dt, \tag{2.10a}$$

$$\dot{\rho} = \frac{1}{2}\nabla_L^*(v\rho + \rho v), \tag{2.10b}$$

$$\rho(0) = \rho_0, \quad \rho(1) = \rho_1, \tag{2.10c}$$

and define the *Wasserstein distance* $W_2(\rho_0, \rho_1)$ between ρ_0 and ρ_1 to be the square root of the infimum of the cost (2.10a). Other choices for $M_\rho(v)$ in (2.9a) give alternative Wasserstein metrics, as noted in [4]. In order for the metric W_2 to be well-defined for all $\rho_0, \rho_1 \in \mathcal{D}_+$, we need to assume that $\ker(\nabla_L)$ is spanned by the identity matrix. For example, if we choose the $L_1, \ldots, L_N$ to be a basis of the Hermitian matrices, then this property holds with $N = n(n+1)/2$. Actually, one can take $N = 2$ and still have this condition satisfied, namely with

$$L_1 = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \ldots & 0 \end{bmatrix}, \quad L_2 = \operatorname{diag}[1,\ 2, \ldots,\ n-1,\ 0].$$

How to choose proper L_k for real applications remains an open problem. On the other hand, the results in the present paper, however, carry through without this assumption.

Finally, we note that the triangle inequality for W_2 follows a standard argument by concatenating two optimal solutions into one feasible solution with proper reparameterisation.

3 Quadratic forms and Poincaré–Wirtinger inequality

In this section, we prove some initial convexity results as well as a Poincaré–Wirtinger-type inequality that we will need in the sequel. Recall that classically, the Poincaré–Wirtinger theorem gives bounds on a function in terms of its derivatives; see [23] for an exact statement. We begin with some notation. If $m_1, \ldots, m_N \in \mathbb{C}^{n\times n}$, we define the column vector $m \in \mathbb{C}^{nN\times n}$ with matrix entries the m_i's by

$$m = (m_1^*, \cdots, m_N^*)^*,$$

the column vector $m_* \in \mathbb{C}^{nN\times n}$ with entries m_i^*'s, by

$$m_* = (m_1, \cdots, m_N)^*.$$

For $m, b \in \mathbb{C}^{nN\times n}$, i.e., with $b = (b_1^*, \ldots, b_N^*)^*$ for $b_1, \ldots, b_N \in \mathbb{C}^{n\times n}$, we define the inner products

$$\langle m_i; b_i\rangle = \mathrm{tr}(m_i^* b_i), \quad \langle m; b\rangle = \mathrm{tr}(m^* b) = \sum_{i=1}^{N} \langle m_i; b_i\rangle$$

and introduce

$$m \cdot b = \frac{1}{2}\big(\langle m; b\rangle + \langle b; m\rangle\big) \in \mathbb{R}.$$

Then, for $v \in \mathbb{C}^{nN\times n}$ and $\rho \in \mathcal{H}_+$, we define the quadratic form

$$Q_\rho(v) := \mathrm{tr}(\rho v^* v) = \langle v\rho; v\rangle. \tag{3.1}$$

The following lemma is an easy consequence of the definition and can be readily verified.

Lemma 1 *Let $\rho \in \mathcal{H}_+$ and $v, w \in \mathbb{C}^{nN\times n}$. The following hold:*

(i) $Q_\rho(v) \geq 0$ *and* $Q_\rho(v) = 0$ *if* $v = 0$*; when* $\rho \in \mathcal{H}_{++}$*, it becomes if and only if,*
(ii) $Q_\rho(v + w) = Q_\rho(v) + Q_\rho(w) + \langle v\rho; w\rangle + \langle w\rho; v\rangle$*,*
(iii) $Q_\rho((1 - t)v + tw) = (1 - t)Q_\rho(v) + tQ_\rho(w) - t(1 - t)Q_\rho(v - w)$*,*
(iv) *if we further assume that* $v, w \in \mathcal{S}^N$*, then* $\langle w\rho; v\rangle = \langle \rho v; w\rangle$*.*

Since $||\cdot||$ (the standard norm on $\mathcal{H}$) is uniformly convex and $\ker(\nabla_L)$ is a finite dimensional space, there exists a unique $\mathrm{proj}(X) \in \ker(\nabla_L)$, the orthogonal projection of X onto $\ker(\nabla_L)$, such that

$$\min_{Z\in\ker(\nabla_L)} ||X - Z|| = ||X - \mathrm{proj}(X)||.$$

If we denote by $\ker(\nabla_L)^\perp$ the orthogonal complement of $\ker(\nabla_L)$ in $\mathcal{H}$, then

$$\mathcal{H} = \ker(\nabla_L) \oplus \ker(\nabla_L)^\perp.$$

Lemma 2 *For any $\rho \in \mathcal{H}_+$, the map $X \to Q_\rho(\nabla_L X)$ is convex on $\mathcal{H}$. If in addition $\rho > 0$, then the map is strictly convex on $\ker(\nabla_L)^\perp$.*

Proof From Lemma 1, we obtain that, for $t \in (0, 1)$ and $X, Y \in \mathcal{H}$,

$$Q_\rho\big((1 - t)\nabla_L X + t\nabla_L Y\big) = (1 - t)Q_\rho(\nabla_L X) + tQ_\rho(\nabla_L Y) - t(1 - t)Q_\rho(\nabla_L X - \nabla_L Y).$$

The convexity follows from the fact that $Q_\rho(\cdot) \geq 0$. Furthermore, if $\rho > 0$, then $Q_\rho(\nabla_L X - \nabla_L Y) > 0$ unless $\nabla_L X - \nabla_L Y = 0$. Hence, $X \to Q_\rho(\nabla_L X)$ is strictly convex on $\ker(\nabla_L)^\perp$. □

Theorem 3 (Poincaré–Wirtinger inequality) *Let $\mathbb{K} \subset \mathcal{D}_+$ be a compact set. Then, there exists a constant $c_\mathbb{K} > 0$ such that for all $X \in \mathcal{H}$ and $\rho \in \mathbb{K}$,*

$$Q_\rho\big(\nabla_L(X - \mathrm{proj}(X))\big) \geq c_\mathbb{K} \big\|X - \mathrm{proj}(X)\big\|^2.$$

Proof Define

$$c_\mathbb{K} := \inf_{\rho, X}\Big\{\mathrm{tr}\big(\rho(\nabla_L X)^* \nabla_L X\big) \mid \rho \in \mathbb{K}, X \in \ker(\nabla_L)^\perp, ||X|| = 1\Big\} \tag{3.2}$$

and let $(\rho_k, X_k)_k$ be a minimising sequence in (3.2). This infimum of a continuous function over a compact set is a minimum, attained at a certain (ρ, X). Since $X \in \ker(\nabla_L)^\perp$, we cannot have $X \in \ker(\nabla_L)$ because, we would otherwise have $X = 0$ which will contradict the fact that $||X|| = 1$. Since $\rho > 0$, we conclude that

$$\mathrm{tr}\Big(\rho(\nabla_L X)^* \nabla_L X\Big) > 0. \tag{3.3}$$

In conclusion,

$$\mathrm{tr}\big(\rho(\nabla_L Y)^* \nabla_L Y\big) \geq c_\mathbb{K} > 0,$$

for any $\rho \in \mathbb{K}$ and any $Y \in \ker(\nabla_L)^\perp$ such that $||Y|| = 1$. By homogeneity, this completes the proof of the theorem. □

Lemma 4 *Given $\rho \in \mathcal{H}_{++}$ and $v \in \mathcal{S}^N$, there is a unique element $\nabla_L X, X \in \mathcal{H}$, such that*

$$Q_\rho(v - \nabla_L X) \leq Q_\rho(v - \nabla_L Y),$$

for all $Y \in \mathcal{H}$. Furthermore, the minimiser is characterised by the Euler–Lagrange equations

$$(v - \nabla_L X)\rho + \rho(v - \nabla_L X) \in \ker(\nabla_L^*). \tag{3.4}$$

Proof Let $(X_\ell)_\ell \subset \mathcal{H}$ be a sequence such that

$$\lim_{\ell \to \infty} Q_\rho(v - \nabla_L X_\ell) = \inf_{Y \in \mathcal{H}} Q_\rho(v - \nabla_L Y).$$

Note that $\big(Q_\rho(\nabla_L X_\ell)\big)_\ell$ is bounded by definition and the convexity of $Q_\rho(\cdot)$. Replacing X_ℓ by $X_\ell - \mathrm{proj}(X_\ell)$ if necessary, we use the Poincaré–Wirtinger inequality to conclude that $(X_\ell)_\ell$ is a bounded sequence. Passing to a subsequence if necessary, we may assume that $(X_\ell)_\ell$ converges to some X which minimises $Q_\rho(v - \nabla_L X)$ over $\mathcal{H}$. The uniqueness follows from the strict convexity of $Q_\rho(\cdot)$ and condition (3.4) expresses stationarity. □

4 Flow rates in the space of densities

We now return to the continuity equation

$$\dot{\rho} = f$$

with the *flow rate* f being the divergence of a momentum field p, i.e.,

$$f = \nabla_L^* p, \tag{4.1}$$

with $p \in \mathcal{S}^N$, so that $f \in \mathcal{H}$ as well as $\mathrm{tr}(f) = 0$. In particular, we are interested in the case where the momentum is a linear function of ρ of the form $p = M_\rho(v)$ (see (2.9a)); then $p = \frac{1}{2}(m - m_*)$ with $m = v\rho \in \mathbb{C}^{Nn\times n}$, $\rho \in \mathcal{H}_{++}$ and $v \in \mathcal{S}^N$, and

$$f = \frac{1}{2}\nabla_L^*(v\rho + \rho v).$$

Since the range of ∇_L^* coincides with $\ker(\nabla_L)^\perp$, any f belongs to $\ker(\nabla_L)^\perp$. The next theorem states that, in this case, the converse holds, namely, not only that given $\rho \in \mathcal{H}_{++}$, any $f \in \ker(\nabla_L)^\perp$ can be written as above with $m = v\rho$, but also that $v \in \mathcal{S}^N$ can be selected in the range of ∇_L and that this choice is unique.

Theorem 5 *For any $\rho \in \mathcal{H}_{++}$ and $f \in \ker(\nabla_L)^\perp$, there exists a unique $X \in \ker(\nabla_L)^\perp$ such that*

$$f = \frac{1}{2}\nabla_L^*\big(\nabla_L X\rho + \rho\nabla_L X\big). \tag{4.2}$$

Furthermore, if $\mathbb{K}$ is a compact subset of $\mathcal{D}_+$ and $\rho/\mathrm{tr}(\rho) \in \mathbb{K}$, then there exists $c_\mathbb{K} > 0$ such that

$$||f|| \geq c_\mathbb{K}\ \mathrm{tr}(\rho)\,||X||. \tag{4.3}$$

Proof Define the functional

$$I(Y) = \frac{1}{2}Q_\rho(\nabla_L Y) - \langle f; Y\rangle, \quad \forall\, Y \in \mathcal{H}.$$

To avoid trivialities, we assume that $f \neq 0$. Observe that $I(Y) \equiv 0$ on $\ker(\nabla_L)$ and for $0 < \epsilon << 1$, we have that

$$I(\epsilon f) = \frac{\epsilon^2}{2}Q_\rho(\nabla_L f) - \epsilon||f||^2 < 0.$$

Thus, λ_0, the infimum of I over $\mathcal{H}$ is negative. Let $(Y_\ell)_\ell$ be a minimising sequence. Since $\lambda_0 < 0$, for ℓ large enough, $I(Y_\ell) < 0$ and so, $Y_\ell \in \mathcal{H}\setminus\ker(\nabla_L)$. Replacing Y_ℓ by $Y_\ell - \mathrm{proj}(Y_\ell)$ if necessary, we may assume that $Y_\ell \in \ker(\nabla_L)^\perp$. By the Poincaré–Wirtinger inequality

$$0 > I(Y_\ell) \geq c_\mathbb{K}\,\mathrm{tr}\,(\rho)||Y_\ell||^2 - ||f||\,||Y_\ell||.$$

Consequently, $(Y_\ell)_\ell$ is a bounded sequence and so, passing to a subsequence if necessary, we may assume that $(Y_\ell)_\ell$ converges to some $X \in \ker(\nabla_L)^\perp$. We have $0 \geq c_\mathbb{K}\ \mathrm{tr}(\rho)||X||^2 - ||f||\,||X||$ and so, (4.3) holds.

If $Y \in \mathcal{H}$ is arbitrary, then for any real number ϵ, we use Lemma 1 to conclude that

$$I(X + \epsilon Y) = I(X) + \epsilon\left\langle \frac{1}{2}\nabla_L^*\big(\nabla_L X\rho + \rho\nabla_L X\big) - f; Y\right\rangle + o(\epsilon).$$

By Lemma 2, I is convex on $\mathcal{H}$ and so, X is a critical point of I if and only if X minimises I. Thus (4.2) holds if and only if X minimises I. Since, the same lemma gives that I is strictly convex on $\ker(\nabla_L)^\perp$, I admits a unique minimiser on $\ker(\nabla_L)^\perp$ which means that there exists a unique $X \in \ker(\nabla_L)^\perp$ such that (4.2) holds. □

Remark 6 *The uniqueness and existence of the representation may also be proven as follows. We first note that provided $\rho \in \mathcal{H}_{++}$, the non-commutative multiplication in (2.9a) defines a positive-definite Hermitian operator*

$$M_\rho : \mathcal{S}^N \to \mathcal{S}^N \; : \; v \mapsto \frac{1}{2}(v\rho + \rho v).$$

It follows that $\nabla_L^* M_\rho \nabla_L$, when restricted to $\ker(\nabla_L)^\perp = \mathrm{range}(\nabla_L^*)$, is positive and therefore invertible. Thus, for all $f \in \ker(\nabla_L)^\perp$, (4.2) has a unique solution $X \in \ker(\nabla_L)^\perp$. One also gets a lower bound on the norm of f as follows. If $\lambda_{\min} > 0$ denotes the smallest eigenvalue of

$$\nabla_L^* M_\rho \nabla_L|_{\ker(\nabla_L)^\perp} : \ker(\nabla_L)^\perp \to \ker(\nabla_L)^\perp,$$

then

$$\|f\| \geq \lambda_{\min} \|X\|.$$

5 Flows in the space of densities

We begin with establishing a canonical representation of flow rates that minimise a certain analogue of *kinetic energy* of our matrix-valued flows.

Proposition 1 *Suppose $\rho \in \mathcal{H}_+$, $f \in \ker(\nabla_L)^\perp$, $X \in \mathcal{H}$ satisfy (4.2), and that $v \in \mathcal{S}^N$ is such that*

$$f = \frac{1}{2}\nabla_L^*\big(v\rho + \rho v\big).$$

The following hold:

(i) *For all $Y \in \mathcal{H}$,*

$$\frac{1}{2}\,\mathrm{tr}(\rho v^* v) \geq \langle f; Y\rangle - \frac{1}{2} Q_\rho(\nabla_L Y)$$

and equality holds if and only if $v = \nabla_L X$.

(ii) *Further assume that $\rho > 0$ (which by Theorem 5 is a sufficient condition for (4.2) to hold). Then*

$$\min_{m \in \mathbb{C}^{nN\times n}} \left\{ \frac{1}{2}\langle m; m\rho^{-1}\rangle \mid f = \frac{1}{2}\nabla_L^*(m - m_*) \right\} = \max_{Y\in\mathcal{H}} \left\{ \langle f; Y\rangle - \frac{1}{2} Q_\rho(\nabla_L Y) \right\}.$$

Besides, the maximum is uniquely attained by the X which satisfies (4.2) and the minimum is uniquely attained by $m = \nabla_L X \rho$ for the same X.

Proof (i) We have

$$\frac{1}{2}\,\mathrm{tr}(\rho v^* v) = \frac{1}{2}\langle v\rho^{\frac{1}{2}}; v\rho^{\frac{1}{2}}\rangle + \langle f - \frac{1}{2}\nabla_L^*\big(v\rho + \rho v\big); Y\rangle = \frac{1}{2}\langle v\rho^{\frac{1}{2}}; v\rho^{\frac{1}{2}}\rangle + \langle f; Y\rangle - \frac{1}{2}\langle v\rho + \rho v; \nabla_L Y\rangle.$$

Since both v as well as $\nabla_L Y$ belong to $\mathcal{S}^N$ (cf. Lemma 1 (iv)),

$$\langle v\rho + \rho v; \nabla_L Y\rangle = \langle v\rho^{\frac{1}{2}}; \nabla_L Y \rho^{\frac{1}{2}}\rangle + \langle \nabla_L Y \rho^{\frac{1}{2}}; v\rho^{\frac{1}{2}}\rangle.$$

We conclude that

$$\frac{1}{2}\operatorname{tr}(\rho v^* v) = \frac{1}{2}\|v\rho^{\frac{1}{2}} - \nabla_L Y \rho^{\frac{1}{2}}\|^2 + \langle f; Y\rangle - \frac{1}{2}\|\nabla_L Y\rho^{\frac{1}{2}}\|^2 \geq \langle f; Y\rangle - \frac{1}{2}Q_\rho(\nabla_L Y).$$

(ii) Computations similar to the ones in (i) reveal that

$$\frac{1}{2}\langle m; m\rho^{-1}\rangle = \frac{1}{2}\|m\rho^{-\frac{1}{2}} - \nabla_L Y\rho^{\frac{1}{2}}\|^2 + \langle f; Y\rangle - \frac{1}{2}\|\nabla_L Y \rho^{\frac{1}{2}}\|^2 \tag{5.1}$$

and so, for $Y \in \mathcal{H}$,

$$\frac{1}{2}\langle m; m\rho^{-1}\rangle \geq \langle f; Y\rangle - \frac{1}{2}Q_\rho(\nabla_L Y).$$

□

We proceed to consider paths $\rho(t) \in \mathcal{H}_+$ for $t \in [0, 1]$ along with corresponding flow rates and action integrals. A corollary of the above proposition ascertains the measurability of the canonical representation of the velocity field v.

Corollary 7 *Let $\mathbb{L} \subset \mathcal{H}_{++}$ and denote by $A : \ker(\nabla_L)^\perp \times \mathbb{L} \to \mathcal{H}$ the map which to (f, ρ) associates $X \in \ker(\nabla_L)^\perp$ such that (4.2) holds.*

- (i) *If $\mathbb{L}$ is a compact subset of $\mathcal{H}_{++}$, then A is continuous.*
- (ii) *If $\rho : [0, 1] \to \mathcal{H}_{++}$ and $f : [0, 1] \to \ker(\nabla_L)^\perp$ are continuous at $t_0 \in [0, 1]$, then $A(f, \rho)$ is continuous at t_0.*
- (iii) *If $\rho \in L^1(0, 1; \mathcal{H}_{++})$ and $f \in L^1(0, 1; \ker(\nabla_L)^\perp)$ are measurable, then $A(f, \rho)$ is measurable.*

Proof (i) Let $\mathbb{K}$ be the set of $\rho/\operatorname{tr}(\rho)$ such that $\rho \in \mathbb{L}$. Let $(f_\ell, \rho_\ell)_\ell$ be a sequence in $\ker(\nabla_L)^\perp \times \mathbb{L}$ converging to (f, ρ). By Theorem 5, $(X_\ell)_\ell := \big(A(f_\ell, \rho_\ell)\big)_l$ is a bounded sequence in $\ker(\nabla_L)^\perp$ and so, has all its points of accumulation in $\ker(\nabla_L)^\perp$. If X is any such point of accumulation, then clearly

$$f = \frac{1}{2}\nabla_L^*(\nabla_L X\rho + \rho\nabla_L X).$$

Since ρ is invertible, X is unique and so, $A(f, \rho) = X$. This establishes (i).

(ii) Condition (ii) is a direct consequence of (i).

(iii) Approximate ρ in the L^1-norm by a sequence $(\rho_\ell)_\ell \subset C([0, 1]; \mathcal{H}_{++})$ which converges pointwise almost everywhere to ρ. Similarly, approximate f in the L^1-norm by a sequence $(f_\ell)_\ell \subset C([0, 1]; \ker(\nabla_L)^\perp)$ which converges pointwise almost everywhere to f. By (i), $\big(A(f_\ell, \rho_\ell)\big)_\ell$ converges pointwise almost everywhere to $A(f, \rho)$ and so, $A(f, \rho)$ is measurable. This establishes (iii). □

Lemma 8 *If $\rho_0, \rho_1 \in \mathcal{D}_+$, then the following hold:*

(i) *If $\rho_1 - \rho_0 \in \ker(\nabla_L)^\perp$, then there exists a Borel map $t \to X(t) \in \ker(\nabla_L)^\perp$ and a Borel map $t \to \rho(t)$ starting at $\rho(0) = \rho_0$ and ending at $\rho(1) = \rho_1$ such that*

$$\dot\rho(t) = \frac{1}{2}\nabla_L^*\big(\nabla_L X(t)\rho(t) + \rho(t)\nabla_L X(t)\big) \tag{5.2}$$

in the sense of distributions, and

$$\int_0^1 Q_{\rho(t)}(\nabla_L X(t))dt < \infty. \tag{5.3}$$

(ii) *Conversely, assume that there exist a Borel map $t \to v(t) \in \mathcal{S}^N$ and a Borel map $t \to \rho(t)$ starting at ρ_0 and ending at ρ_1 such that*

$$\dot{\rho}(t) = \frac{1}{2}\nabla_L^*\big(v(t)\rho(t) + \rho(t)v(t)\big) \tag{5.4}$$

in the sense of distributions, and

$$\int_0^1 \mathrm{tr}(\rho(t)v(t)^* v(t))dt < \infty. \tag{5.5}$$

Then $\rho_1 - \rho_0 \in \ker(\nabla_L)^\perp$.

Proof (i) Assume $\rho_1 - \rho_0 \in \ker(\nabla_L)^\perp$, set $\rho(t) = (1-t)\rho_0 + t\rho_1$. Then, $\mathbb{K} := \{\rho(t) \mid t \in [0, 1]\}$ is a compact subset of $\mathcal{D}_+$. For each $t \in [0, 1]$, we use Theorem 5 to find a unique $X(t) \in \ker(\nabla_L)^\perp$ such that

$$\rho_1 - \rho_0 = \frac{1}{2}\nabla_L^*\big(\nabla_L X(t)\rho(t) + \rho(t)\nabla_L X(t)\big)$$

and

$$||X(t)|| \leq c_{\mathbb{K}}.$$

By Corollary 7, $t \to X(t) \in \ker(\nabla_L)^\perp$ is continuous. Hence, (5.2) and (5.3) hold.

(ii) Conversely, assume (5.4) and (5.5) hold. Let $Y \in \ker(\nabla_L)$, then

$$\langle \rho_1 - \rho_0; Y\rangle = \frac{1}{2}\int_0^1 \langle \nabla_L^*\big(v(t)\rho(t) + \rho(t)v(t)\big); Y\rangle dt = \frac{1}{2}\int_0^1 \langle v(t)\rho(t) + \rho(t)v(t); \nabla_L Y\rangle dt = 0.$$

Since $Y \in \ker(\nabla_L)$ is arbitrary, we conclude that $\rho_1 - \rho_0 \in \ker(\nabla_L)^\perp$. □

Remark 9 *Observe that in Lemma 8 (ii), if we relax the assumptions on ρ_0 and ρ_1 by merely imposing that $\rho_0, \rho_1 \in \mathcal{H}_+$, then (5.4) and (5.5) still imply $\rho_1 - \rho_0 \in \ker(\nabla_L)^\perp$.*

For $\rho \in \mathcal{H}_{++}$ and $m \in \mathbb{C}^{nN\times n}$, we set

$$F(\rho, m) := \frac{1}{2}\langle m, m\rho^{-1}\rangle.$$

Given $\rho_0, \rho_1 \in \mathcal{D}_+$, denote by $\mathcal{C}(\rho_0, \rho_1)$ the set of paths (ρ, v), such that $\rho \in C^1([0, 1], \mathcal{D}_+)$, start at ρ_0 and end at ρ_1, $v : (0, 1) \to \mathcal{S}^N$ is Borel, $Q_\rho(v) \in L^1(0, 1)$ and

$$\dot{\rho} = \frac{1}{2}\nabla_L^*(v\rho + \rho v)$$

in the sense of distributions on $(0, 1)$. Similarly, we define $\tilde{\mathcal{C}}(\rho_0, \rho_1)$ to be the set of paths (ρ, m) such that $\rho \in C^1([0, 1], \mathcal{D}_+)$ start at ρ_0 and end at ρ_1, $m : (0, 1) \to \mathbb{C}^{nN\times n}$ is Borel, $F(\rho, m) \in L^1(0, 1)$ and

$$\dot{\rho} = \frac{1}{2}\nabla_L^*(m - m_*)$$

in the sense of distributions on $(0, 1)$.

Observe that if $v \in \mathcal{S}^N$ and we set $m = v\rho$, then

$$F(\rho, m) = \frac{1}{2}\,\mathrm{tr}(\rho v^* v)$$

and so, the embedding $(\rho, v) \to (\rho, v\rho)$ of $\mathcal{C}(\rho_0, \rho_1)$ into $\tilde{\mathcal{C}}(\rho_0, \rho_1)$, extends $\frac{1}{2}\,\mathrm{tr}(\rho v^* v)$ to $F(\rho, m)$. Consequently,

$$\inf_{(\rho,v)}\left\{\int_0^1 \frac{1}{2}\,\mathrm{tr}(\rho v^* v)dt \mid (\rho, v) \in \mathcal{C}(\rho_0, \rho_1)\right\} \geq \inf_{(\rho,m)}\left\{\int_0^1 F(\rho, m)dt \mid (\rho, m) \in \tilde{\mathcal{C}}(\rho_0, \rho_1)\right\}.$$

We next show that the inequality can be turned into an equality.

Lemma 10 *If $\rho_0, \rho_1 \in \mathcal{D}_+$, then*

$$\inf_{(\rho,v)}\left\{\int_0^1 \frac{1}{2}\,\mathrm{tr}(\rho v^* v)dt \mid (\rho, v) \in \mathcal{C}(\rho_0, \rho_1)\right\} = \inf_{(\rho,m)}\left\{\int_0^1 F(\rho, m)dt \mid (\rho, m) \in \tilde{\mathcal{C}}(\rho_0, \rho_1)\right\}.$$

Proof It suffices to show that for any $(\rho, m) \in \tilde{\mathcal{C}}(\rho_0, \rho_1)$, there exists $(\rho, v) \in \mathcal{C}(\rho_0, \rho_1)$ such that

$$\int_0^1 \frac{1}{2}\,\mathrm{tr}(\rho v^* v)dt \leq \int_0^1 F(\rho, m)dt. \tag{5.6}$$

Observe that for almost every $t \in (0, 1)$, we have

$$\dot{\rho}(t) = \frac{1}{2}\nabla_L^*(m(t) - m_*(t)). \tag{5.7}$$

Since both $t \to \rho(t)$ and $t \to \dot{\rho}(t)$ are continuous, by Corollary 7, there exists a continuous map $X : [0, 1] \to \mathcal{H}$ such that

$$\dot{\rho}(t) = \frac{1}{2}\nabla_L^*\Big(\nabla_L X(t)\rho(t) + \rho(t)\nabla_L X(t)\Big). \tag{5.8}$$

Thus, for almost every $t \in (0, 1)$, both (5.7) and (5.8) hold and so, by Proposition 1

$$\frac{1}{2}\,\mathrm{tr}\big(\rho(t)v^*(t)v(t)\big) \leq F\big(\rho(t), m(t)\big)$$

for almost every $t \in (0, 1)$ with $v(t) = \nabla_L X(t)$. Thus, (5.6) holds, which concludes the proof. □

6 Relaxation of velocity–momentum fields

Given $\rho_0, \rho_1 \in \mathcal{D}_+$, we are interested in characterising the paths (ρ, v) in $\mathcal{H}_+ \times \mathcal{S}^N$ that minimise the ‘action integral’, i.e., paths that possibly attain

$$\inf_{\rho\in\mathcal{H}_+, v\in\mathcal{S}^N}\left\{\int_0^1 \mathrm{tr}(\rho v^* v)dt \,\middle|\, \dot{\rho} = \frac{1}{2}\nabla_L^*(v\rho + \rho v),\;\; \rho(0) = \rho_0,\;\; \rho(1) = \rho_1\right\}. \tag{6.1}$$

When $\rho > 0$, Lemma 10 replaced $\mathrm{tr}(\rho v^* v)$ by a new expression $F(\rho, v\rho)$, introducing a new problem which, under appropriate conditions, is a relaxation of (6.1). It then becomes necessary to extend F to the whole set $\mathcal{H} \times \mathbb{C}^{nN\times n}$ and study the convexity properties of the extended functional. We start by introducing the open sets

$$\mathcal{O}_0 := \mathcal{H}_{++} \times \mathbb{C}^{nN\times n}, \quad \mathcal{O}_\infty := \{\rho \in \mathcal{H}\backslash\mathcal{H}_+\} \times \mathbb{C}^{nN\times n}.$$

We define the functions $F, F_0, G : \mathcal{H} \times \mathbb{C}^{nN\times n} \to [0, \infty]$ given by

$$G(\rho, m) := \inf_{(\rho_\ell, m_\ell)} \liminf_{\ell\to\infty} \left\{ \frac{1}{2}\langle m_\ell ; m_\ell \rho_\ell^{-1}\rangle \mid (\rho_\ell, m_\ell)_\ell \subset \mathcal{O}_0 \text{ converges to } (\rho, m)\right\} \tag{6.2}$$

$$F(\rho, m) = \begin{cases} \frac{1}{2}\langle m; m\rho^{-1}\rangle \text{ if } & (\rho, m) \in \mathcal{O}_0 \\ G(\rho, m) \text{ if } & (\rho, m) \in \mathcal{H} \times \mathbb{C}^{nN\times n} \setminus (\mathcal{O}_0 \cup \mathcal{O}_\infty) \\ \infty \text{ if } & (\rho, m) \in \mathcal{O}_\infty, \end{cases}$$

and

$$F_0(\rho, m) = \begin{cases} \frac{1}{2}\langle m; m\rho^{-1}\rangle \text{ if } & (\rho, m) \in \mathcal{O}_0 \\ \infty \text{ if } & (\rho, m) \in \mathcal{H} \times \mathbb{C}^{nN\times n} \setminus \mathcal{O}_0. \end{cases}$$

We show (cf. Lemma 11) that F is a convex functional and then we characterise the minimisers of

$$\inf_{(\rho,m)} \left\{ \int_0^1 F(\rho, m)dt \,\middle|\, (6.4) \text{ and } (6.5) \quad \text{hold} \right\}. \tag{6.3}$$

Here, the infimum is performed over the set of pairs (ρ, m) satisfying the requirements

$$\rho \in W^{1,2}(0, 1; \mathcal{H}), \quad m \in L^2\big(0, 1; \mathbb{C}^{nN\times n}\big), \tag{6.4}$$

$$\rho(0) = \rho_0, \ \rho(1) = \rho_1, \quad \text{and} \quad \dot{\rho} = \frac{1}{2}\nabla_L^*(m - m_*) \tag{6.5}$$

in the sense of distributions on (0, 1).

Under technical conditions, the characterising of the minimisers (ρ, m) of (6.3) is equivalent to characterising the minimisers (ρ, v) of (6.1). We will make use of the set of paths $\lambda : [0, 1] \to \mathcal{H}$ such that

$$\lambda \in W^{1,1}(0, 1; \mathcal{H}) \tag{6.6}$$

and

$$\dot{\lambda} + \frac{1}{2}(\nabla_L\lambda)^*(\nabla_L\lambda) \le 0 \quad \text{a.e. on} \quad (0, 1). \tag{6.7}$$

Lemma 11 *The function F is convex and lower semicontinuous and equals the convex envelope of F_0. In addition, the Legendre transform of F is*

$$F^*(a, b) = \begin{cases} 0 \textit{ if } & a + \frac{b^*b}{2} \le 0 \\ \infty & \textit{otherwise.} \end{cases} \tag{6.8}$$

Proof Observe that $\mathcal{O}_0$ is a convex set. For $(a, b) \in \mathcal{H} \times \mathbb{C}^{nN\times n}$, we have

$$F_0^*(a, b) = \sup_{\rho,m} \left\{ \langle a; \rho\rangle + b \cdot m - \frac{1}{2}\langle m; m\rho^{-1}\rangle \mid \rho > 0, \ m \in \mathbb{C}^{nN\times n} \right\}.$$

But

$$b\cdot m-\frac{1}{2}\langle m;m\rho^{-1}\rangle=-\frac{1}{2}\|m\rho^{-\frac{1}{2}}-b\rho^{\frac{1}{2}}\|^2+\frac{1}{2}\langle\rho;b^*b\rangle.$$

Hence,

$$F_0^*(a,b)=\sup_\rho\Big\{\langle a;\rho\rangle+\frac{1}{2}\langle\rho;b^*b\rangle\mid\rho>0\Big\}=\begin{cases}0 \text{ if } & a+\frac{b^*b}{2}\le 0\\ \infty & \text{otherwise.}\end{cases}\tag{6.9}$$

Denote by F_0^{**} the Legendre transform of F_0^*. If $(\rho,m)\in\mathcal{H}\times\mathbb{C}^{nN\times n}$, we use (6.9) to obtain

$$F_0^{**}(\rho,m)=\sup_{a,b}\Big\{\langle a;\rho\rangle+b\cdot m\mid(a,b)\in\mathcal{H}\times\mathbb{C}^{nN\times n},\ a+\frac{b^*b}{2}\le 0\Big\}.\tag{6.10}$$

If $(\rho,m)\in\mathcal{O}_0$, we can set

$$b=m\rho^{-1},\ a=-\frac{1}{2}\rho^{-1}m^*m\rho^{-1}\in\mathcal{H}.$$

Clearly $a+\frac{b^*b}{2}=0$ and so, by (6.10)

$$F_0^{**}(\rho,m)\ge-\Big\langle\frac{1}{2}\rho^{-1}m^*m\rho^{-1};\rho\Big\rangle+\langle m;m\rho^{-1}\rangle=\frac{1}{2}\langle m;m\rho^{-1}\rangle=F_0(\rho,m).\tag{6.11}$$

If $(\rho,m)\in\mathcal{O}_\infty$, then there exists $x\in\mathbb{C}^n$ such that $\langle\rho x;x\rangle<0$. Set

$$(a_\lambda,b):=(-\lambda x\otimes x,0).$$

Observe that for any $\lambda\ge 0$, we have $a_\lambda+\frac{b^*b}{2}\le 0$. Thus, by (6.10)

$$F_0^{**}(\rho,m)\ge\lim_{\lambda\to\infty}\langle a_\lambda,\rho\rangle+b\cdot m=\lim_{\lambda\to\infty}-\lambda\langle\rho x;x\rangle=\infty.\tag{6.12}$$

Since, in general $F_0^{**}\le F_0$, (6.11) and (6.12) imply that

$$F_0^{**}=F_0=F\quad\text{on}\quad\mathcal{O}_0\cup\mathcal{O}_\infty.\tag{6.13}$$

Observe that F is lower semicontinuous. We next claim that since (6.13) holds, F is convex. Indeed, let $t\in(0,1)$, let $\rho^0,\rho^1\in\mathcal{H}$ and let $m^0,m^1\in\mathbb{C}^{nN\times n}$. We are to show that

$$F(\rho^t,m^t)\le(1-t)F(\rho^0,m^0)+tF(\rho^1,m^1),\tag{6.14}$$

where

$$(\rho^t,m^t):=\Big((1-t)\rho^0+t\rho^1,(1-t)m^0+tm^1\Big).$$

Clearly, (6.14) holds if either $(\rho^0,m^0)\in\mathcal{O}_\infty$ or $(\rho^1,m^1)\in\mathcal{O}_\infty$. Since $\mathcal{O}_0$ is a convex set and F_0^{**} is a convex function, we use (6.13) to conclude that (6.14) holds if $(\rho^0,m^0)\in\mathcal{O}_0$ and $(\rho^1,m^1)\in\mathcal{O}_0$. It remains to prove (6.14) when we have either $(\rho^0,m^0)\notin(\mathcal{O}_0\cup\mathcal{O}_\infty)$ and $(\rho^1,m^1)\notin\mathcal{O}_\infty$ or $(\rho^1,m^1)\notin(\mathcal{O}_0\cup\mathcal{O}_\infty)$ and $(\rho^0,m^0)\notin\mathcal{O}_\infty$. In these latter cases, there exist sequences $(\rho^0_\ell,m^0_\ell)_\ell\subset\mathcal{O}_0$ converging to (ρ^0,m^0) and $(\rho^1_\ell,m^1_\ell)_\ell\subset\mathcal{O}_0$ converging to (ρ^1,m^1) such that by (6.13) and the definition of F,

$$F(\rho^0,m^0)=\lim_{\ell\to\infty}F_0^{**}(\rho^0_\ell,m^0_\ell)\quad\text{and}\quad F(\rho^1,m^1)=\lim_{\ell\to\infty}F_0^{**}(\rho^1_\ell,m^1_\ell).\tag{6.15}$$

Note that

$$\left(\rho_\ell^t, m_\ell^t\right) := \left((1-t)\rho_\ell^0 + t\rho_\ell^1, (1-t)m_\ell^0 + tm_\ell^1\right)_\ell \subset \mathcal{O}_0 \tag{6.16}$$

and the sequence in (6.16) converges to (ρ^t, m^t). Thus, using the definition of F, (6.13) and the convexity property of F_0^{**}, we have

$$F(\rho^t, m^t) \le \liminf_{\ell\to\infty} F_0^{**}\left(\rho_\ell^t, m_\ell^t\right) \le \liminf_{\ell\to\infty} \left\{(1-t)F_0^{**}(\rho_\ell^0, m_\ell^0) + tF_0^{**}(\rho_\ell^1, m_\ell^1)\right\}.$$

This, together with (6.15), yields (6.14). Thus, F is convex and so, since F is also lower semicontinuous, we have $F = F^{**}$.

Note that the complement of $\mathcal{O}_0 \cup \mathcal{O}_\infty$ is contained in the boundary of $\mathcal{O}_0 \cup \mathcal{O}_\infty$ and so, since F_0^{**} is lower semicontinuous, (6.13), in view of the definition (6.2) of G, implies

$$F_0^{**} \le G \quad \text{on} \quad \mathcal{H} \times \mathbb{C}^{nN\times n} \setminus \mathcal{O}_0 \cup \mathcal{O}_\infty. \tag{6.17}$$

By (6.13) and (6.17), $F_0^{**} \le F$ and so, $F_0^{**} \le F^{**}$. The fact that $F \le F_0$ yields the reversed inequality to ensure that $F^{**} = F_0^{**}$. As a consequence, $F^* = F_0^{***} = F_0^*$ and so, by (6.9) we obtain (6.8). □

Lemma 12 *Let $\rho \in \mathcal{H}_+$, $X \in \mathcal{H}$ and set $m := \nabla_L X\,\rho$, then we have*

$$F(\rho, m) = \frac{1}{2}\langle \nabla_L X\,\rho; \nabla_L X\rangle$$

and

$$(a, b) := \left(\frac{1}{2}(\nabla_L X)^*(\nabla_L X), \nabla_L X\right) \in \partial_- F(\rho, m). \tag{6.18}$$

Proof For any $\epsilon > 0$, we have $\rho + \epsilon I \in \mathcal{H}_{++}$ and both ρ and $(\rho + \epsilon I)^{-1}$ have the same eigenspaces. Thus ρ and $(\rho + \epsilon I)^{-1}$ commute and so, $(\rho + \epsilon I)^{-1}\rho \in \mathcal{H}$. If $\lambda_1, \ldots, \lambda_n \ge 0$ are the eigenvalues of ρ, then $\lambda_1/(\lambda_1 + \epsilon), \ldots, \lambda_n/(\lambda_n + \epsilon) \ge 0$ are those of $(\rho + \epsilon I)^{-1}\rho$ and so, $(\rho + \epsilon I)^{-1}\rho \in \mathcal{H}_+$. Thus,

$$0 \le \langle (\nabla_L X)^* \nabla_L X; \rho(\rho + \epsilon I)^{-1}\rangle = \langle \nabla_L X(\rho + \epsilon I)^{-1}; \nabla_L X \rho\rangle. \tag{6.19}$$

Since, F is lower semicontinuous, we have

$$F(\rho, m) \le \liminf_{\epsilon\to 0^+} F(\rho + \epsilon I, \nabla_L X\,\rho) = \frac{1}{2}\liminf_{\epsilon\to 0^+} \langle \nabla_L X\,\rho; \nabla_L X\,\rho(\rho + \epsilon I)^{-1}\rangle \tag{6.20}$$

But by (6.19)

$$\langle \nabla_L X\,\rho; \nabla_L X\,\rho(\rho + \epsilon I)^{-1}\rangle \le \langle \nabla_L X\,\rho; \nabla_L X\,(\rho + \epsilon I)(\rho + \epsilon I)^{-1}\rangle = \frac{1}{2}\langle \nabla_L X\,\rho; \nabla_L X\rangle.$$

This, together with (6.20), implies

$$F(\rho, m) \le \frac{1}{2}\langle \nabla_L X\,\rho; \nabla_L X\rangle. \tag{6.21}$$

On the other hand, with (a, b) as in (6.18), we have

$$\langle a; \rho\rangle + b \cdot m = -\frac{1}{2}\langle (\nabla_L X)^* \nabla_L X; \rho\rangle + \langle \nabla_L X; \nabla_L X \rho\rangle = \frac{1}{2}\langle \nabla_L X \rho; \nabla_L X\rangle.$$

We use (6.21) and the fact that $F^*(a, b) = 0$ (cf. by Lemma 11) to conclude that

$$\langle a; \rho\rangle + b\cdot m = \frac{1}{2}\langle \nabla_L X\rho; \nabla_L X\rangle \geq F(\rho, m) + F^*(a, b) \geq \langle a; \rho\rangle + b\cdot m.$$

Thus, $F(\rho, m) = \frac{1}{2}\langle \nabla_L X\rho; \nabla_L X\rangle$ and

$$\langle a; \rho\rangle + b\cdot m = F(\rho, m) + F^*(a, b)$$

implying (6.18). □

Lemma 13 *We have the following:*

(i) *If $\rho \in \mathcal{H}_+\backslash\{0\}$ and $m \in \mathbb{C}^{nN\times n}$, then*

$$F(\rho, m) \geq \frac{||m||^2}{2\,\mathrm{tr}(\rho)}. \tag{6.22}$$

(ii) *Assume $\rho \in C([0, 1]; \mathcal{H}_+)$ and $m : (0, 1) \to \mathbb{C}^{nN\times n}$ is a Borel map such that*

$$\dot\rho = \frac{1}{2}\nabla_L^*(m - m_*)$$

in the sense of distributions on $(0, 1)$ and $F(\rho, m) \in L^1(0, 1)$. Then, $\dot\rho \in L^2(0, 1; \mathcal{H})$ and there exists a constant c_L independent of (ρ, m) such that

$$c_L \int_0^1 F(\rho, m)dt \geq \int_0^1 ||\dot\rho||^2 dt.$$

Furthermore,

$$\mathrm{tr}(\rho)(t) = \mathrm{tr}(\rho)(0).$$

Proof (i) When $\rho \in \mathcal{H}_{++}$, (6.22) is a direct consequence of the fact that $\rho^{-1}\,\mathrm{tr}(\rho) \geq I$. Since F is defined through the liminf in (6.2), we conclude that if $\rho \in \mathcal{H}_+\backslash\{0\}$ and $m \in \mathbb{C}^{nN\times n}$, then (6.22) continues to hold.
(ii) Let c_L be such that $||\nabla_L^*(m - m_*)|| \leq 2c_L||m||$. Under the assumptions in (ii), we have that for almost every $t \in (0, 1)$

$$\int_0^1 ||\dot\rho||^2 dt = \int_0^1 \left\|\frac{1}{2}\nabla_L^*(m - m_*)\right\|^2 dt \leq c_L^2 \int_0^1 ||m||^2 dt.$$

This, together with (i), implies the last inequality in (ii). The conservation of $\mathrm{tr}(\rho)(t)$ is due to the fact that $\mathrm{tr}(\nabla_L^*(m - m_*)) \equiv 0$. □

7 Strong duality

In this section, we state and prove our main results namely a duality as expressed by (7.2) and (7.3). This will be crucial in our further analysis of the proposed Wasserstein distance on $\mathcal{D}_+$ and its Riemannian structure.

We fix $\rho_0, \rho_1 \in \mathcal{D}_+$ such that

$$\rho_0 - \rho_1 \in \ker(\nabla_L)^\perp. \tag{7.1}$$

One of the aims of this section is to show under appropriate conditions that the convex variational problems,

$$\inf_{(\rho,m)}\left\{\int_0^1 F(\rho,m)dt \,\Big|\, (6.4) \text{ and } (6.5) \text{ hold}\right\} =: i_0 \tag{7.2}$$

and

$$\sup_{\lambda}\left\{\langle\lambda(1);\rho_1\rangle-\langle\lambda(0);\rho_0\rangle \,\Big|\, (6.6) \text{ and } (6.7) \text{ hold}\right\} =: j_0, \tag{7.3}$$

are dual to each other. Recall that one of our goals is to define a Riemannian metric on $\mathcal{D}_+$. In order to have finite value for i_0, in view of Lemma 8, it is necessary to assume that $\ker(\nabla_{\rm L})$ is spanned by the identity matrix I. All the analysis, however, goes through without this assumption.

Proposition 2 *Let λ satisfy (6.6) and (6.7) and (ρ,m) satisfy (6.4) and (6.5).*

(i) *Then*

$$\langle\lambda(1);\rho_1\rangle-\langle\lambda(0);\rho_0\rangle \le \int_0^1 F(\rho,m)dt. \tag{7.4}$$

(ii) *If equality holds in (7.4), then λ is a maximiser in (7.3) and (ρ,m) is a minimiser in (7.2).*

Proof Note that (ii) is a direct consequence of (i) and so, the only proof to supply is that of (i). Since λ satisfies (6.6) and (6.7), we use Lemma 11 to conclude that $F(\dot\lambda,\nabla_L\lambda)\equiv 0$ and so,

$$\langle\dot\lambda;\rho\rangle+\nabla_L\lambda\cdot m \le F(\rho,m)+F^*(\dot\lambda,\nabla_L\lambda)=F(\rho,m) \quad \text{a.e. on} \quad (0,1). \tag{7.5}$$

Note that by (6.5)

$$\nabla_L\lambda\cdot m=\frac{1}{2}\langle\nabla_L\lambda;m-m_*\rangle=\frac{1}{2}\langle\lambda;\nabla_L^*(m-m_*)\rangle=\langle\lambda;\dot\rho\rangle,$$

thus, (7.5) implies

$$\frac{d}{dt}\langle\lambda;\rho\rangle=\langle\dot\lambda;\rho\rangle+\langle\lambda;\dot\rho\rangle\le F(\rho,m) \quad \text{a.e. on} \quad (0,1). \tag{7.6}$$

The pointwise derivative of $\langle\lambda;\rho\rangle$ coinciding with its distributional derivative, we integrate (7.6) to discover that

$$\langle\lambda(1);\rho(1)\rangle-\langle\lambda(0);\rho(0)\rangle\le\int_0^1 F(\rho,m)dt.$$

□

By the above proposition, we have the following.

Remark 14 *Assume that $\lambda\in W^{1,1}(0,1;\mathcal{H})$ satisfies (6.7) and $(\rho,m)\in W^{1,2}(0,1;\mathcal{H})\times L^2(0,1;\mathbb{C}^{nN\times n})$ is such that (6.5) holds. Then,*

$$\langle\lambda(1);\rho(1)\rangle-\langle\lambda(0);\rho(0)\rangle=\int_0^1 F(\rho,m)dt,$$

 https://doi.org/10.1017/S0956792519000172

if and only if

$$\frac{d}{dt}\langle\lambda;\rho\rangle = F(\rho,m) \quad a.e. \tag{7.7}$$

Lemma 15 *Let $(\rho,m)\in\mathcal{H}_{++}\times\mathbb{C}^{nN\times n}$.*

(i) *The partial derivatives of F with respect to m and ρ are*

$$\nabla_m F(\rho,m) = m\rho^{-1} \quad and \quad \nabla_\rho F(\rho,m) = -\frac{1}{2}(m\rho^{-1})^*(m\rho^{-1}). \tag{7.8}$$

(ii) *They satisfy the relation*

$$\nabla_\rho F + \frac{1}{2}(\nabla_m F)^*(\nabla_m F) = 0 \quad on \quad \mathcal{O}_0. \tag{7.9}$$

Proof (i) For $r\in\mathcal{H}$ with $||r|| << 1$, we have

$$(\rho+r)^{-1} = (I+\rho^{-1}r)^{-1}\rho^{-1} = \Big(I+\sum_{l=1}^{\infty}(-1)^l(\rho^{-1}r)^l\Big)\rho^{-1} = \rho^{-1}-\rho^{-1}r\rho^{-1}+o(||r||).$$

Hence,

$$F(\rho+r,m) = \frac{1}{2}\langle m; m(\rho+r)^{-1}\rangle = F(\rho,m) - \frac{1}{2}\langle r;(m\rho^{-1})^*(m\rho^{-1})\rangle + o(||r||),$$

which gives the second identity in (7.8). The first identity is obtained in a similar manner.

(ii) By direct computations, (7.9) is obtained from (7.8). □

Theorem 16 *Let $\rho_0,\rho_1\in\mathcal{D}_+$ (recall that throughout this section (7.1) is assumed to hold).*

(i) *The problem (7.2) admits a minimiser (ρ,m).*

(ii) *Let $J := \{t\in(0,1)\mid \det(\rho(t))>0\}$. Then J is an open set and there exists a measurable map $\hat{\lambda}: J\to \ker(\nabla_L)^\perp$ such that for almost every $t\in J$*

$$m = \nabla_L\hat{\lambda}\,\rho \quad on \quad J. \tag{7.10}$$

(iii) *If $\epsilon>0$ and we set $J_\epsilon := \{t\in(0,1)\mid \det(\rho(t))>\epsilon\}$, then $\hat{\lambda}\in L^1(J_\epsilon;\mathcal{H})$. Extend $\hat{\lambda}$ to $(0,1)$ by setting $\hat{\lambda}$ to 0 on $(0,1)\backslash J$, and let $\lambda(t)=\hat{\lambda}(t)+\Lambda(t)$ where*

$$\Lambda(t) = \int_0^t -\frac{1}{2}\mathrm{proj}\big((\nabla_L\hat{\lambda})^*(\nabla_L\hat{\lambda})\big)dt.$$

Then, $m=\nabla_L\lambda\rho$ on J, $\lambda\in L^1(J_\epsilon;\mathcal{H})$ and

$$\dot{\lambda} + \frac{1}{2}(\nabla_L\lambda)^*(\nabla_L\lambda) = 0 \quad on \quad J \tag{7.11}$$

in the sense of distributions.

Proof (i) By Lemma 8 and the fact that $\mathrm{tr}(\rho_0-\rho_1)=0$, we have $i_0<\infty$. Let $(\rho_\ell,m_\ell)_\ell$ be a minimising sequence of (7.2). Using Lemma 13 and the fact that

$$\sup_\ell \int_0^1 F(\rho_\ell,m_\ell)dt < \infty,$$

we conclude that

$$\mathrm{tr}(\rho_\ell) \equiv 1,$$

$(m_\ell)_\ell$ is a bounded sequence in $L^2(0, 1; \mathbb{C}^{nN\times n})$ and $(\rho_\ell)_\ell$ is a bounded sequence in $W^{1,2}(0, 1; \mathcal{H})$. Extracting subsequences if necessary, we assume without loss of generality that $(m_\ell)_\ell$ converges weakly to some m in $L^2(0, 1; \mathbb{C}^{nN\times n})$, $(\rho_\ell)_\ell$ converges strongly to some ρ in $L^2(0, 1; \mathcal{H})$ and $(\dot\rho_\ell)_\ell$ converges weakly to $\dot\rho$ in $L^2(0, 1; \mathcal{H})$. Since (ρ_ℓ, m_ℓ) satisfies (6.4) and (6.5), so does (ρ, m). By Lemma 11, F is convex and lower semicontinuous and so,

$$i_0 = \liminf_{\ell\to\infty} \int_0^1 F(\rho_\ell, m_\ell)dt \ge \int_0^1 F(\rho, m)dt \ge i_0.$$

The first equality in the above is due to the fact that $(\rho_\ell, m_\ell)_\ell$ is a minimising sequence in (7.2). The first inequality is due to standard results of the calculus of variations (cf., e.g., [9]) which ensure lower semicontinuity of functionals for weak topologies. The last inequality is due to the definition of i_0. This proves that (ρ, m) is a minimiser in (7.2).

(ii) Since $\rho \in W^{1,2}(0, 1; \mathcal{H}_+)$, $t \to \det(\rho(t))$ is a continuous function on $[0, 1]$ and so, the set J is an open set. The last identity in (6.5), which holds in the sense of distributions, also holds pointwise almost everywhere. Hence, for almost every $t \in J$, $m(t)$ minimises $F(\rho(t), w)$ over the set of $w \in \mathbb{C}^{nN\times n}$ such that

$$\dot\rho(t) = \frac{1}{2}\nabla_L^*(w - w_*).$$

By Proposition 1, for these t, there exists a unique $\hat\lambda(t) \in \ker(\nabla_L)^\perp$ such that $m(t) = \nabla_L\hat\lambda(t)\rho(t)$. By Corollary 7, the map $\hat\lambda : J \to \ker(\nabla_L)^\perp$ is measurable.

(iii) We first establish $\hat\lambda \in L^1(J_\epsilon, \mathcal{H})$. By Lemma 13 (i) and the fact that $F(\rho, m) \in L^1(0, 1)$, we have $m \in L^2(0, 1; \mathbb{C}^{nN\times n})$, and therefore, $\nabla_L\hat\lambda \in L^2(J_\epsilon; \mathbb{C}^{nN\times n})$. It follows

$$||\nabla_L\hat\lambda||^2 \in L^1(J_\epsilon; \mathbb{C}^{nN\times n}).$$

Now, we apply the Poincaré–Wirtinger inequality (cf. Theorem 3) with $\mathbb{K} := \{I\}$ to conclude that there exists a constant $c_{\mathbb{K}}$ independent of ϵ and $\hat\lambda$ such that

$$c_{\mathbb{K}} \int_{J_\epsilon} ||\hat\lambda||^2 dt \le \int_{J_\epsilon} ||\nabla_L\hat\lambda||^2 dt.$$

Therefore, $\hat\lambda \in L^1(J_\epsilon; \mathcal{H})$. Since $\nabla_L\hat\lambda \in L^2(J_\epsilon; \mathbb{C}^{nN\times n})$, we have $\Lambda \in L^1(0, 1)$. It follows $\lambda \in L^1(J_\epsilon; \mathcal{H})$. Recalling that $\Lambda(t) \in \ker(\nabla_L)$ for any $t \in (0, 1)$, we have $m = \nabla_L\hat\lambda\rho = \nabla_L\lambda\rho$. Proving (7.11) amounts to proving that, for any arbitrary $f \in C_c^1(J; \mathcal{H})$,

$$\int_J \langle \dot f; \lambda\rangle dt = \frac{1}{2}\int_J \langle f; (\nabla_L\lambda)^*(\nabla_L\lambda)\rangle dt. \tag{7.12}$$

For $f \in C_c^1(J; \ker(\nabla_L))$, since $\hat\lambda(t) \in \ker(\nabla_L)^\perp$, $\Lambda(t) \in \ker(\nabla_L)$ and $f(t) \in \ker(\nabla_L)$, we have

$$\int_J \langle \dot f; \lambda\rangle dt = \int_J \langle \dot f; \Lambda\rangle dt = \int_J \langle f; \mathrm{proj}\Big(\frac{1}{2}(\nabla_L\hat\lambda)^*(\nabla_L\hat\lambda)\Big)\rangle dt = \int_J \langle f; \Big(\frac{1}{2}(\nabla_L\hat\lambda)^*(\nabla_L\hat\lambda)\Big)\rangle dt,$$

which proves (7.12). Therefore, it remains to consider $f \in C_c^1(J; \ker(\nabla_L)^\perp)$. Fix such an f and denote the support of f by $\mathrm{spt}(f)$. To avoid technical difficulties, we assume without loss of generality that $\mathrm{spt}(f)$ is contained in some J_ϵ. Extend f to $[0, 1]$ by setting $f(t) \equiv 0$ on $[0, 1] \setminus J$ and observe that the extension, which we still denote by f, satisfies $f \in C_c^1([0, 1]; \ker(\nabla_L)^\perp)$. By Proposition 1 and Corollary 7, there exists a unique map $\beta \in C(J; \ker(\nabla_L)^\perp)$ such that

$$\dot{f} = \frac{1}{2}\nabla_L^*(\nabla_L\beta\rho + \rho\nabla_L\beta) \quad \text{on} \quad J.$$

By its uniqueness property on J, we have $\beta(t) = 0$ for $t \in J \setminus \mathrm{spt}(f)$. Set $\beta(t) = 0$ for $t \in [0, 1] \setminus J$ and continue to denote the extension by β to observe that $\beta \in C([0, 1]; \ker(\nabla_L)^\perp)$ and

$$\dot{f} = \frac{1}{2}\nabla_L^*(\nabla_L\beta\rho + \rho\nabla_L\beta) \quad \text{on} \quad (0, 1). \tag{7.13}$$

We set

$$\rho_\epsilon := \rho + \epsilon f, \quad m_\epsilon := m + \epsilon\nabla_L\beta\rho.$$

Since $\mathrm{spt}(f)$ is a compact subset of J there exists $c > 0$ such that $\rho \geq c$ on $\mathrm{spt}(f)$. We have

$$0 \leq \int_0^1 F(\rho_\epsilon, m_\epsilon)dt - \int_0^1 F(\rho, m)dt = \int_{\mathrm{spt}(f)} \big(F(\rho_\epsilon, m_\epsilon) - F(\rho, m)\big)dt$$

and so, the function $\epsilon \to \int_0^1 F(\rho_\epsilon, m_\epsilon)dt$ achieves its minimum at $\epsilon = 0$. Since $m \in L^2(0, 1; \mathbb{C}^{nN\times n})$ and F is differentiable on $\{r \in \mathcal{H} \mid r \geq c\} \times \mathbb{C}^{nN\times n}$ with its derivatives given by (7.8), we conclude that $\int_0^1 F(\rho_\epsilon, m_\epsilon)dt - \int_0^1 F(\rho,\ m)dt$ is differentiable at $\epsilon = 0$ with a null derivative there. More precisely,

$$\begin{aligned} 0 &= \int_{\mathrm{spt}(f)} \big(\langle\nabla_\rho F(\rho,\ m); f\rangle + \nabla_m F(\rho, m)\cdot\nabla_L\beta\rho\big)dt && (7.14)\\ &= \int_J \big(\langle\nabla_\rho F(\rho, m); f\rangle + \nabla_m F(\rho, m)\cdot\nabla_L\beta\rho\big)dt. && (7.15)\end{aligned}$$

This, together with (7.8) and the fact that $m = \nabla_L\lambda\rho$ on J, yields

$$\begin{aligned} 0 &= \int_J \big(\langle -\frac{1}{2}(\nabla_L\lambda)^*\nabla_L\lambda; f\rangle + \nabla_L\lambda\cdot\nabla_L\beta\rho\big)dt\\ &= \int_J \big(\langle -\frac{1}{2}(\nabla_L\lambda)^*\nabla_L\lambda; f\rangle + \frac{1}{2}\langle\nabla_L\lambda; \nabla_L\beta\rho + \rho\nabla_L\beta\rangle\big)dt\\ &= \int_J \big(\langle -\frac{1}{2}(\nabla_L\lambda)^*\nabla_L\lambda; f\rangle + \frac{1}{2}\langle\lambda; \nabla_L^*\big(\nabla_L\beta\rho + \rho\nabla_L\beta\big)\rangle\big)dt.\end{aligned}$$

We then use (7.13) to obtain (7.12). □

Corollary 17 *Let $\rho_0, \rho_1 \in \mathcal{D}_+$ and let (ρ, m) be such that (6.4) and (6.5) hold.*

(i) *If there exists $\lambda \in W^{1,1}(0, 1; \mathcal{H})$ such that*

$$\dot{\lambda} + \frac{1}{2}(\nabla_L\lambda)^*(\nabla_L\lambda) = 0 \quad \textit{on} \quad (0, 1) \tag{7.16}$$

in the sense of distributions and

$$m = \nabla_L\lambda\rho \quad \textit{a.e. on} \quad (0, 1), \tag{7.17}$$

then (ρ, m) minimises (7.2).

(ii) *Conversely, assume that* (ρ, m) *minimises (7.2) and the range of* ρ *is contained in* $\mathcal{D}_+$. *Then, there exists* $\lambda \in W^{1,1}(0,1;\mathcal{H})$ *such that (7.16) holds.*

(iii) *Any minimiser* (ρ, m) *of (7.2) whose range is contained in* $\mathcal{D}_+$ *must be of class* C^∞.

Proof (i) Assume there exists $\lambda \in W^{1,1}(0,1;\mathcal{H})$ such that (7.16) holds. Since (7.16) holds almost everywhere and m satisfies (7.17), we have

$$\langle\dot{\lambda};\rho\rangle + \nabla_L\lambda\cdot m = -\left\langle\frac{1}{2}(\nabla_L\lambda)^*(\nabla_L\lambda);\rho\right\rangle + \langle\nabla_L\lambda;\nabla_L\lambda\rho\rangle = \frac{1}{2}\langle\nabla_L\lambda;\nabla_L\lambda\rho\rangle.$$

Hence by Lemma 12, we have

$$\langle\dot{\lambda};\rho\rangle + \nabla_L\lambda\cdot m = F(\rho,m) + F^*(\dot{\lambda},\nabla_L\lambda) = F(\rho,m) \quad \text{a.e. on} \quad (0,1). \tag{7.18}$$

Since

$$\nabla_L\lambda\cdot m = \frac{1}{2}\langle\nabla_L\lambda;m-m_*\rangle = \frac{1}{2}\langle\lambda;\nabla_L^*(m-m_*)\rangle,$$

we combine (6.5) and (7.18) to conclude that

$$\frac{d}{dt}\langle\lambda;\rho\rangle = \langle\dot{\lambda};\rho\rangle + \langle\lambda;\dot{\rho}\rangle = F(\rho,m) + F^*(\dot{\lambda},\nabla_L\lambda) = F(\rho,m) \quad \text{a.e. on} \quad (0,1). \tag{7.19}$$

The pointwise derivative of $\langle\lambda;\rho\rangle$ coinciding with its distributional derivative, we integrate (7.19) to discover that

$$\langle\lambda(1);\rho(1)\rangle - \langle\lambda(0);\rho(0)\rangle = \int_0^1 F(\rho,m)dt. \tag{7.20}$$

We use Proposition 2 to conclude that (ρ, m) minimises (7.2).

(ii) Assume that (ρ, m) minimises (7.2) and the range of ρ is contained in $\mathcal{D}_+$. By Theorem 16, there exists $\lambda : [0,1] \to \mathcal{H}$ such that (7.17) holds. Since ρ is continuous, its range is a compact set and so, the range of $\det(\rho)$ is a compact subset of $(0,\infty)$. Since $m \in L^2(0,1;\mathbb{C}^{nN\times n})$, we have $\nabla_L\lambda \in L^2(0,1;\mathbb{C}^{nN\times n})$. Thus, in view of (7.16), $\lambda \in W^{1,1}(0,1;\mathcal{H})$.

(iii) Assume (ρ, m) is a minimiser in (7.2) and the range of ρ is contained in $\mathcal{D}_+$. Since by (ii) λ is continuous, (7.16) implies that $\dot{\lambda}$ is continuous and so, λ is of class C^1. We repeat the procedure to conclude that λ is of class C^∞. Since (7.17) holds and both ρ and λ are continuous, we obtain that m is continuous. By (6.5), $\dot{\rho}$ is continuous and so, ρ is of class C^1. Because, λ has been shown to be of class C^∞, (7.17) implies that m is of class C^1. We use again (6.5) to conclude that $\dot{\rho}$ is of class C^1 and so, ρ is of class C^2. We repeat the procedure to conclude that ρ is of class C^∞. □

Remark 18 *Let* $\rho_0, \rho_1 \in \mathcal{D}_+$. *By Theorem 16, (7.2) admits a minimiser* $(\tilde{\rho}, \tilde{m})$. *Observe that thanks to Corollary 17, we have proven that if the range of* $\tilde{\rho}$ *is contained in* $\mathcal{H}_{++}$, *then we have the duality result*

$$\min_{(\rho,m)}\left\{\int_0^1 F(\rho,m)dt \,\middle|\, (6.4) \textit{ and } (6.5) \textit{ hold}\right\} = \max_{\lambda}\left\{\langle\lambda(1);\rho_1\rangle - \langle\lambda(0);\rho_0\rangle \,\middle|\, (6.6) \textit{ and } (6.7) \textit{ hold}\right\}.$$

Our goal is to extend the duality result in Remark 18 without assuming that the range of $\tilde{\rho}$ is contained in $\mathcal{H}_{++}$, at some expense. It is convenient to introduce the sets

$$\mathcal{A} := \{\rho \in L^2(0,1;\mathcal{H}) \mid \operatorname{tr}(\rho) \equiv 1\} \times L^2(0,1;\mathbb{C}^{nN\times n}),$$

$$\mathcal{A}_1 := \{\rho \in L^2(0,1;\mathcal{H}) \mid \operatorname{tr}(\rho) \leq 1\} \times L^2(0,1;\mathbb{C}^{nN\times n}),$$

$$\mathcal{A}_\infty := L^2(0,1;\mathcal{H}) \times L^2(0,1;\mathbb{C}^{nN\times n})$$

and

$$\mathcal{B} := W^{1,2}(0,1;\mathcal{H}), \quad \mathcal{B}_\ell := \left\{\lambda \in W^{1,2}(0,1;\mathcal{H}) \mid ||\lambda||_{W^{1,2}} \leq \ell^2\right\},$$

where

$$||\lambda||^2_{W^{1,2}} := \int_0^1 (||\lambda||^2 + ||\dot{\lambda}||^2)dt.$$

We also set

$$J(a,b) = \inf_{(\rho,m)} \left\{F(\rho,m) - \langle \rho; a\rangle - \frac{1}{2}\langle m - m_*; b\rangle \mid (\rho,m) \in \mathcal{A}\right\}.$$

and for $\beta \in \{1,\infty\}$,

$$J_\beta(a,b) = \inf_{(\rho,m)} \left\{F(\rho,m) - \langle \rho; a\rangle - \frac{1}{2}\langle m - m_*; b\rangle \mid (\rho,m) \in \mathcal{A}_\beta\right\}.$$

Remark 19 *Let $\lambda \in W^{1,2}(0,1;\mathcal{H})$, let $\alpha \in W^{1,2}(0,1)$ and set $\bar{\lambda} := \lambda + \alpha I$, where I is the identity matrix. Then*

(i) *Since F is 1-homogeneous, $J_\infty = -F^*$ and $J_1 = -\sup_{0\leq\mu\leq 1}\{-\mu J\} = -J_-$.*
(ii) *$J(\dot{\lambda}, \nabla_L \lambda) \in L^2(0,1)$.*
(iii) *Since $I \in \ker(\nabla_{\mathrm{L}})$, $\nabla_L\lambda = \nabla_L\bar{\lambda}$. One may easily check that $J(\dot{\bar{\lambda}}, \nabla_L\bar{\lambda}) = J(\dot{\lambda}, \nabla_L\lambda) - \dot{\alpha}$.*

Lemma 20 *For any $\lambda \in W^{1,2}(0,1;\mathcal{H})$, there exists $\bar{\lambda} \in W^{1,2}(0,1;\mathcal{H})$ such that $J(\dot{\bar{\lambda}}, \nabla_L\bar{\lambda}) \geq 0$ and*

$$\inf_{(\rho,m)\in\mathcal{A}_1} \mathcal{L}(\rho,m,\lambda) = \langle\bar{\lambda}(1);\rho_1\rangle - \langle\bar{\lambda}(0);\rho_0\rangle.$$

Proof Set

$$F := \{t \in (0,1) \mid J(\dot{\lambda}(t), \nabla_L\lambda(t)) < 0\}, \quad \alpha(t) := \int_0^t \chi_F(s) J(\dot{\lambda}(s), \nabla_L\lambda(s))ds \quad \forall\, t \in (0,1).$$

By Remark 19, $\alpha \in W^{1,2}(0,1)$ and $\bar{\lambda} := \lambda + \alpha I$ satisfy the desired properties. □

Corollary 21 *We have*

$$\sup_{\lambda\in\mathcal{B}} \inf_{(\rho,m)\in\mathcal{A}} \mathcal{L}(\rho,m,\lambda) = \sup_{\lambda\in\mathcal{B}} \inf_{(\rho,m)\in\mathcal{A}_1} \mathcal{L}(\rho,m,\lambda) = \sup_{\lambda\in\mathcal{B}} \inf_{(\rho,m)\in\mathcal{A}_\infty} \mathcal{L}(\rho,m,\lambda)$$

Proof Since by Lemma 20

$$\sup_{\lambda\in\mathcal{B}}\inf_{(\rho,m)\in\mathcal{A}_1}\mathcal{L}(\rho,m,\lambda)=\sup_{\lambda\in\mathcal{B}}\left\{\inf_{(\rho,m)\in\mathcal{A}_1}\mathcal{L}(\rho,m,\lambda)\mid J(\dot{\lambda},\nabla_L\lambda)\geq 0\right\}$$

we use Remark 19 to conclude that

$$\sup_{\lambda\in\mathcal{B}}\inf_{(\rho,m)\in\mathcal{A}_1}\mathcal{L}(\rho,m,\lambda)=\sup_{\lambda\in\mathcal{B}}\left\{\langle\lambda(1);\rho_1\rangle-\langle\lambda(0);\rho_0\rangle\mid J(\dot{\lambda},\nabla_L\lambda)\geq 0\right\}.$$

Similarly, Lemma 20 and Remark 19 imply that

$$\inf_{(\rho,m)\in\mathcal{A}_1}\mathcal{L}(\rho,m,\lambda)=\inf_{(\rho,m)\in\mathcal{A}}\mathcal{L}(\rho,m,\lambda).$$

□

Theorem 22 *Let $\rho_0,\rho_1\in\mathcal{D}_+$. We have*

$$\min_{(\rho,m)\in\mathcal{A}}\left\{\int_0^1 F(\rho,m)dt\;\middle|\;(6.5)\text{ holds}\right\}=\sup_{\lambda\in\mathcal{B}}\left\{\langle\lambda(1);\rho_1\rangle-\langle\lambda(0);\rho_0\rangle\;\middle|\;(6.7)\text{ holds}\right\}.$$

Proof We endow $\mathcal{A}$ and $\mathcal{B}$ with their respective weak topologies and for $(\rho,m)\in\mathcal{A}$ and $\lambda\in\mathcal{B}$, we define

$$\mathcal{L}(\rho,m,\lambda):=\langle\lambda(1);\rho_1\rangle-\langle\lambda(0);\rho_0\rangle+\int_0^1\Big(F(\rho,m)-\langle\rho;\dot{\lambda}\rangle-\frac{1}{2}\langle m-m_*;\nabla_L\lambda\rangle\Big)dt.$$

For $\ell\in(0,\infty)$, $\mathcal{B}_\ell$ is a compact convex topological space. Let $(\rho^0,m^0)\in\mathcal{A}$ and $\lambda^0\in\mathcal{B}_\ell$. For any c, the set $\{\lambda\in\mathcal{B}_\ell\mid\mathcal{L}(\rho^0,m^0,\lambda)\geq c\}$ is a closed convex set in $\mathcal{B}_\ell$ while the set $\{(\rho,m)\in\mathcal{A}\mid\mathcal{L}(\rho,m,\lambda^0)\leq c\}$ is a closed convex set in $\mathcal{A}$. Thus, by Theorem 1.6 in [16]

$$\inf_{\mathcal{A}}\sup_{\mathcal{B}_\ell}\mathcal{L}=\sup_{\mathcal{B}_\ell}\inf_{\mathcal{A}}\mathcal{L}. \tag{7.21}$$

Set

$$\mathcal{E}(\rho,m):=\sup_{\lambda\in\mathcal{B}_1}\langle\lambda(1);\rho_1\rangle-\langle\lambda(0);\rho_0\rangle-\int_0^1\Big(\langle\rho;\dot{\lambda}\rangle+\frac{1}{2}\langle m-m_*;\nabla_L\lambda\rangle\Big)dt.$$

Then, $\mathcal{E}$ is a non-negative convex function such that

$$\mathcal{E}(\rho,m)=0\quad\Longleftrightarrow\quad\rho(1)=\rho_1,\ \rho(0)=\rho_0\text{ and }\dot{\rho}=\frac{1}{2}\nabla_L^*(m-m_*) \tag{7.22}$$

in the sense of distributions on $(0,1)$.

For any $(\rho,m)\in\mathcal{A}$

$$\sup_{\lambda\in\mathcal{B}_\ell}\mathcal{L}(\rho,m)=\int_0^1 F(\rho,m)dt+\ell\mathcal{E}(\rho,m).$$

Let $(\rho_\ell,m_\ell)\in\mathcal{A}$ be such that

$$\inf_{(\rho,m)\in\mathcal{A}}\sup_{\lambda\in\mathcal{B}_\ell}\mathcal{L}(\rho,m,\lambda)=\int_0^1 F(\rho_\ell,m_\ell)dt+\ell\mathcal{E}(\rho_\ell,m_\ell).$$

Since $\mathrm{tr}(\rho_\ell) \equiv 1$, by Lemma 13, $(m_\ell)_\ell$ is bounded in $L^2(0, 1; \mathbb{C}^{nN\times n})$. The fact that $\rho_\ell \geq 0$ yields that $(\rho_\ell)_\ell$ is bounded in $L^2(0, 1; \mathcal{H})$. Thus, there exists a subsequence $(\rho_{\ell_k}, m_{\ell_k})_k$ such that $(\rho_{\ell_k})_k$ converges weakly to some ρ_∞ is $L^2(0, 1; \mathcal{H})$ and $(m_{\ell_k})_k$ converges weakly to some m_∞ is $L^2(0, 1; \mathbb{C}^{nN\times n})$. Clearly, we have $\mathrm{tr}(\rho_\infty) \equiv 1$.

Let $(\tilde{\rho}, \tilde{m})$ be a minimiser of (7.2) as given by Theorem 22. By (7.22), $\mathcal{E}(\tilde{\rho}, \tilde{m}) = 0$ and so,

$$\int_0^1 F(\rho_\ell, m_\ell)dt + \ell\mathcal{E}(\rho_\ell, m_\ell) \leq \int_0^1 F(\tilde{\rho}, \tilde{m})dt. \tag{7.23}$$

Hence, by the weak lower semicontinuity property of $\mathcal{E}$, we have

$$\mathcal{E}(\rho_\infty, m_\infty) \leq \liminf_{k\to\infty} \mathcal{E}(\rho_{\ell_k}, m_{\ell_k}) \leq \liminf_{k\to\infty} \frac{1}{\ell_k}\int_0^1 F(\tilde{\rho}, \tilde{m})dt = 0.$$

We conclude that $\mathcal{E}(\rho_\infty, m_\infty) = 0$ and so, by (7.22)

$$\rho_\infty(1) = \rho_1, \ \rho_\infty(0) = \rho_0 \text{ and } \dot{\rho}_\infty = \frac{1}{2}\nabla_L^*(m_\infty - (m_\infty)_*) \tag{7.24}$$

in the sense of distributions on (0, 1). By (7.23)

$$\int_0^1 F(\rho_\infty, m_\infty)dt \leq \liminf_{k\to\infty}\int_0^1 F(\rho_{\ell_k}, m_{\ell_k})dt \leq \int_0^1 F(\tilde{\rho}, \tilde{m})dt. \tag{7.25}$$

Since (ρ_∞, m_∞) satisfies (7.24), (7.25) shows that its is also a minimiser in (7.2).

By the definition of (ρ_ℓ, m_ℓ) and then (7.21), we have

$$\int_0^1 F(\rho_{\ell_k}, m_{\ell_k})dt + \ell_k\mathcal{E}(\rho_{\ell_k}, m_{\ell_k}) = \sup_{\mathcal{B}_{\ell_k}}\inf_{\mathcal{A}}\mathcal{L} \leq \sup_{\mathcal{B}}\inf_{\mathcal{A}}\mathcal{L}$$

and so,

$$\int_0^1 F(\rho_{\ell_k}, m_{\ell_k})dt \leq \sup_{\mathcal{B}}\inf_{\mathcal{A}}\mathcal{L}.$$

This, together with (7.25) and Corollary 21, implies

$$\int_0^1 F(\rho_\infty, m_\infty)dt \leq \sup_{\lambda\in B}\inf_{(\rho,m)\in\mathcal{A}}\mathcal{L} = \sup_{\lambda\in\mathcal{B}}\inf_{(\rho,m)\in\mathcal{A}_\infty}\mathcal{L}(\rho, m, \lambda)$$

$$= \sup_{\lambda\in\mathcal{B}}\left\{\langle\lambda(1); \rho_1\rangle - \langle\lambda(0); \rho_0\rangle - \int_0^1 F^*(\dot{\lambda}, \nabla_L\lambda)dt\right\}.$$

Hence, using Lemma 11, we conclude that

$$\int_0^1 F(\rho_\infty, m_\infty)dt \leq \sup_{\lambda\in\mathcal{B}}\left\{\langle\lambda(1); \rho_1\rangle - \langle\lambda(0); \rho_0\rangle \mid \dot{\lambda} + \frac{1}{2}(\nabla_L\lambda)^*(\nabla_L\lambda) \leq 0\right\}.$$

This, together with Proposition 2, yields the desired result. □

8 Conservation of the Hamiltonian

In this section, we are ready to complete the analysis of the Wasserstein distance W_2 on positive densities. The next result has three parts: constancy of F along optimal trajectories, W_2 gives well-defined Riemannian-type metric on $\mathcal{D}_+$ and a Lax–Oleinik–Hopf result. More precisely, we state and prove the following key theorem:

Theorem 23 (Conservation of the Hamiltonian) *Let $\rho_0, \rho_1 \in \mathcal{D}_+$ and assume (ρ, m) minimises (7.2). Then*
(i)

$$F(\rho(t), m(t)) \equiv F(\rho(0), m(0)).$$

(ii) If $0 \leq s \leq t \leq 1$, then

$$W_2(\rho(s), \rho(t)) = (t-s)\sqrt{2F(\rho(t), m(t))} = (t-s)W_2(\rho_0, \rho_1).$$

(iii) If we further assume that $\lambda \in W^{1,1}(0, 1; \mathcal{H})$ is a maximiser in (7.3), then

$$\langle \lambda(t); \rho(t)\rangle = \langle \lambda(0); \rho_0\rangle + \frac{W_2(\rho_0, \rho(t))^2}{2t}, \quad t \in (0, 1].$$

Proof (i) Let $\zeta \in C_c^1(0, 1)$ be arbitrary and set $S(t) = t + \epsilon\zeta(t)$. We have $S(0) = 0$, $S(1) = 1$ and $\dot{S}(t) = 1 + \epsilon\dot{\zeta}(t) > 1/2$ for $|\epsilon| << 1$. Thus, $S : [0, 1] \to [0, 1]$ is a diffeomorphism. Let $T := S^{-1}$ and set

$$f(s) = \rho(T(s)), \quad w(s) = \dot{T}(s)m(T(s)).$$

We have

$$\dot{f} = \frac{1}{2}\nabla_L^*(w - w_*), \quad f(0) = \rho_0, \quad f(1) = \rho_1.$$

Thus,

$$\int_0^1 F(\rho, m)dt \leq \int_0^1 F(f, w)ds = \int_0^1 \dot{T}^2(s)F(\rho(T(s)), m(T(s)))ds.$$

We use the fact that $dt = \dot{T}(s)ds$ and $\dot{T}(S(t))\dot{S}(t) = 1$ to conclude that

$$\int_0^1 F(\rho, m)dt \leq \int_0^1 \frac{1}{\dot{S}(t)}F(\rho(t), m(t))dt = \int_0^1 (1 - \epsilon\dot{\zeta}(t) + o(\epsilon))F(\rho(t), m(t))dt.$$

Since $\epsilon \to \int_0^1(1 - \epsilon\dot{\zeta}(t) + o(\epsilon))F(\rho(t), m(t)dt$ admits its minimum at 0, we conclude that its derivative there is null, i.e.,

$$\int_0^1 \dot{\zeta}(t)F(\rho(t), m(t))dt = 0.$$

This proves that the distributional derivative of $F(\rho(t), m(t))$ is null and so, $F(\rho(t), m(t))$ is independent of t.

(ii) Recalling the definition of W_2 in (2.10), we have

$$W_2(\rho_0, \rho_1)^2 = \int_0^1 2F(\rho(\tau), m(\tau))d\tau.$$

Due to the time homogeneity of the definition, and the optimality of (ρ, m), one can clearly see, for $0 \le s \le t \le 1$,

$$W_2(\rho(s), \rho(t))^2 = (t-s)\int_s^t 2F(\rho(\tau), m(\tau))d\tau.$$

We use these, together with (i), to conclude the proof of (ii).

(iii) We use Remark 14 and the duality result in Theorem 22 to conclude that

$$\frac{d}{dt}\langle\rho;\lambda\rangle = F(\rho, m) \quad \text{a.e..}$$

Thus, if $0 \le s < t \le 1$, then

$$\langle\lambda(t);\rho(t)\rangle - \langle\lambda(s);\rho(s)\rangle = \int_s^t F(\rho(\tau), m(\tau))d\tau.$$

We apply (i) and use (ii) to conclude that

$$\langle\lambda(t);\rho(t)\rangle - \langle\lambda(s);\rho(s)\rangle = (t-s)F(\rho(0), m(0)) = \frac{W_2(\rho(s), \rho(t))^2}{2(t-s)}.$$

□

9 Conclusions and further research

This note continues our study of a quantum mechanical approach to (non-commutative) optimal mass transport between density matrices initiated in [4]. In particular, we prove a duality result that elucidates the connection of our set-up to Monge–Kantorovich theory [14], in particular Kantorovich duality as well as a Poincaré–Wirtinger-type result. For applications, it is important to note that our methodology leads to convex optimisation problems that may be implemented and numerically solved on computer.

It is of interest to explore further potential implications of this construction to quantum channels and quantum information. It is not clear if a static formulation of the problem that leads to the Benamou–Brenier matrix formulation described in the present work can be introduced. At this point, the best result (to the best of our knowledge) is along the lines of Theorem 23. In this sense, it may be that the dynamic Benamou–Brenier approach to mass transport may be the more versatile formulation for defining the Wasserstein metric as compared to the classical Monge–Kantorovich approach. Finally, much of the theory should go through in the infinite dimensional case. This is another area we plan to further explore.

References

[1] AMBROSIO, L. & GIGLI, N. (2012) A user's guide to optimal mass transport. In: *Modelling and Optimisation of Flows on Networks*, Lecture Notes in Mathematics, Vol. 2062, Springer, New York, pp. 1–155.

[2] CARLEN, E. & MAAS, J. (2014) An analog of the 2-Wasserstein metric in non-Commutative probability under which the fermionic Fokker–Planck equation is gradient flow for the entropy. *Commun. Math. Phys.* **331**(3), 887–926.

[3] CARLEN, E. & MAAS, J. (2017) Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance. *J. Funct. Anal.* **273**, 1810–1869. Original version https://arxiv.org/abs/1609.01254, 2016.

[4] CHEN, Y., GEORGIOU, T. & TANNENBAUM, A. (2018) Matrix optimal mass transport: a quantum mechanical approach. *IEEE Trans. Autom. Control* **63**, 2612–2619. Original version https://arxiv.org/abs/1610.03041, 2016.

[5] CHEN, Y., GANGBO, W., GEORGIOU, T. & TANNENBAUM, A. (2017) On the matrix Monge-Kantorovich problem. https://arxiv.org/abs/1701.02826.

[6] CHEN, Y., HABER, E., YAMAMOTO, K., GEORGIOU, T. & TANNENBAUM, A. (2018) An efficient algorithm for matrix-valued and vector-valued optimal mass transport. *J. Sci. Comput.* **77**, 79–100. https://arxiv.org/abs/1706.08841.

[7] BENAMOU, J.-D. & BRENIER, Y. (2000) A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numerische Math.* **84**, 375–393.

[8] DITTMANN, J. (1993) On the Riemannian geometry of finite dimensional mixed states. *Semin. Sophus Lie* **3**, 73–87.

[9] EKELAND, I. & TEMAN, R. (1976) *Convex Analysis and Variational Problems,* SIAM, Philadelphia.

[10] GANGBO, W. & MCCANN, R. J. (1996) The geometry of optimal transportation. *Acta Math.* **177**(2), 113–161.

[11] GOZLAN, N., ROBERTO, C. & SAMSON, P.-M. (2012) Hamilton-Jacobi equations on metric spaces and transport-entropy inequalities. https://arxiv.org/abs/1203.2783.

[12] GUSTAFSON, S. & SIGAL, I. M. (2011) *Mathematical Concepts of Quantum Mechanics*, Springer, New York.

[13] JORDAN, R., KINDERLEHRER, D. & OTTO, F. (1998) The variational formulation of the Fokker–Planck equation. *SIAM J. Math. Anal.* **29**(1), 1–17.

[14] KANTOROVICH, L. V. (1948) On a problem of Monge. *Uspekhi Mat. Nauk.* **3**, 225–226.

[15] MCCANN, R. (1997) A convexity principle for interacting gases. *Adv. Math.* **128**, 153–179.

[16] MERTENS, J.-F., SORIN S. & ZAMIR, S. (1994) *Repeated Games. Part A: Background Material*. CORE Discussion Paper No 9420, Université Catholique de Louvain.

[17] MITTNENZWEIG, M. & MIELKE, A. (2016) An entropic gradient structure for Lindblad equations and couplings of quantum systems to macroscopic models. https://arxiv.org/abs/1609.05765.

[18] NING, L. & GEORGIOU, T. T. (2014) Metrics for matrix-valued measures via test functions. In: *53rd IEEE Conference on Decision and Control*, pp. 2642–2647.

[19] NING, L., GEORGIOU, T. T. & TANNENBAUM, A. (2015) On matrix–valued Monge-Kantorovich optimal mass transport. *IEEE Trans. Autom. Control* **60**(2), 373–382.

[20] OTTO, F. (2001) The geometry of dissipative evolution equations: the porous medium equation. *Commun. Partial Differ. Equ.* **26**, 101–174.

[21] RACHEV, S. & RÜSCHENDORF, L. (1998) *Mass Transportation Problems*, Probability and Its Applications, Vols. I and II, Springer, New York.

[22] TANNENBAUM, E., GEORGIOU, T. & TANNENBAUM, A. (2010) Signals and control aspects of optimal mass transport and the Boltzmann entropy. In: *49th IEEE Conference on Decision and Control (CDC)*.

[23] VILLANI, C. (2003) *Topics in Optimal Transportation,* Graduate Studies in Mathematics, Vol. 58, AMS, Providence, RI.