Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 20–22

THE SCHRÖDINGER PROBLEM ON THE NON-COMMUTATIVE FISHER-RAO SPACE

LÉONARD MONSAINGEON AND DMITRY VOROTNIKOV

ABSTRACT: We present a self-contained and comprehensive study of the Fisher-Rao space of matrix-valued non-commutative probability measures, and of the related Hellinger space. Our non-commutative Fisher-Rao space is a natural generalization of the classical commutative Fisher-Rao space of probability measures and of the Bures-Wasserstein space of Hermitian positive-definite matrices. We introduce and justify a canonical entropy on the non-commutative Fisher-Rao space, which differs from the von Neumann entropy. We consequently derive the analogues of the heat flow, of the Fisher information, and of the dynamical Schrödinger problem. We show the Γ-convergence of the ε-Schrödinger problem towards the geodesic problem for the Fisher-Rao space, and, as a byproduct, the strict geodesic convexity of the entropy.

KEYWORDS: entropic regularization, matrix-valued measure, optimal transport, gamma-convergence.
MATH. SUBJECT CLASSIFICATION (2010): 28A33, 47A56, 49J45, 49Q20, 58B20.

1. Introduction

Recently, deep connections have been established between *optimal transport* and *quantum information theory*, two very active and proficient fields. The prototypical quantities involved are the quadratic *Wasserstein distance* between two probability measures $\rho_0, \rho_1 \in \mathcal{P}(\mathcal{X})$ over a polish space $(\mathcal{X}, d)$

$$W_2^2(\rho_0, \rho_1) = \min_{\pi} \left\{ \int_{\mathcal{X}\times\mathcal{X}} d^2(x,y)\, \mathrm{d}\pi(x,y) : \right.$$

Received 17 July 2020.
Credit is due to Aymeric Baradat for our construction of recovery sequences, which was improved and adapted from [8]. LM wishes to thank Jean-Claude Zambrini for numerous and fruitful discussions on the Schrödinger problem, and acknowledges support from the Portuguese Science Foundation through FCT project PTDC/MAT-STA/28812/2017 *SchröMoka*. DV was partially supported by CMUC – UID/MAT/00324/2020, funded by the Portuguese Government through FCT/MCTES and co-funded by the European Regional Development Fund through the Partnership Agreement PT2020, and by the FCT project PTDC/MAT-PUR/28686/2017.

$$\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{X}) \text{ has marginals } \rho_0, \rho_1 \Bigg\}$$

and the *Bures distance* between Hermitian positive semi-definite (PSD) matrices A_0, A_1 (or more generally two quantum operators)

$$d_B^2(A_0, A_1) = \min_R \left\{ \left| \sqrt{A_0} - R\sqrt{A_1} \right|^2 : \qquad R \text{ is unitary} \right\}.$$

Clearly these two problems share a similar variational structure, where one aims at finding an optimal coupling (π or R) between a given pair of points. The Wasserstein distance plays a significant role in probability theory, partial differential equations, geometry, etc [69, 68, 63]. The Bures distance appears in quantum information theory [11] as a quantum equivalent of the Fisher information metric in information geometry [7], and yields a noncommutative generalization of the Hellinger distance between probability distributions.

The connection between these objects is three-fold. Firstly, when ρ_0, ρ_1 are Gaussian measures with covariance matrices A_0, A_1, both distances agree as $W_2^2(\rho_0, \rho_1) = d_B^2(A_0, A_1)$ up to some irrelevant multiplicative factor [3, 65]. Hence the Bures distance can be considered simply as Gaussian optimal transport (at least for real, nondegenerate *positive*-definite matrices). Secondly, recent attempts have been made to develop an optimal transport theory for *quantum* objects, namely measures whose values are PSD matrices (or operators) [26, 58, 19, 20, 25, 22, 60, 54, 38]. Finally, let us mention that in the scalar case both worlds have been unified into a single theory, by now referred to as *unbalanced optimal transport* [45, 51, 28] and aiming at providing a transport framework between nonnegative measures with unequal masses. The unbalanced optimal transport of matrix-valued measures has recently been introduced in [27, 15].

The current theory however has two significant stipulations: First of all, the rigorous Riemannian geometric perspective [55, 65, 12, 53] only works for nonsingular matrices (a covariance matrix must be positive-definite), and the analysis is limited to finite-dimensional statistical manifolds (the space of Gaussian measures, parametrized by the finite-dimensional manifold of symmetric positive-definite matrices $\mathcal{S}^{++}$, endowed with the Bures Riemannian metric). In this work we aim at extending the framework and further studying the Bures-Wasserstein geometry in infinite-dimensional counterparts of these statistical manifolds. We consider complex Gaussians and possibly degenerate (*semi*-definite) matrices, and we treat the case of matrix-valued *measures*. Some related attempts in more restrictive settings have been made in [35] (in the context of signal processing) and recently by the second author in [15, Remark 2.9]. PSD-valued densities naturally arise in signal processing, geometry (Riemannian metrics) and other applications. Our two spaces of interest here will be the Hellinger space of PSD-valued measures, and its Fisher-Rao subspace consisting in matrix-measures whose scalar trace integrates to one. We think of the Hellinger space as a noncommutative version of nonnegative measures, while the normalized Fisher-Rao space can be thought of as the subspace of noncommutative *probability* measures, see Remark 5.1. It is worth pointing that both spaces are built upon *complex* Hermitian matrices, but will be considered here as *real* (formal) manifolds, cf. [12].

Our first contribution in this paper is a self-contained and comprehensive study of these two spaces, in particular we will reveal a very geometric structure by showing that the non-commutative Hellinger space is a *metric cone* over the Fisher-Rao subspace. In other words, the Fisher-Rao space can be viewed as a unit sphere in the ambient Hellinger space, and we shall sometimes accordingly speak of the *spherical* Fisher-Rao space as opposed to the *conic* Hellinger space. In the commutative case (matrices of size 1), our Hellinger and Fisher-Rao distances coincide with the classical Hellinger and Fisher-Rao distances [7, 43, 46]. Some of our results are perhaps known to experts in the field for finite-dimensional Bures manifolds, but in the infinite-dimensional measure-valued setting the analysis requires significant technical work and we could not find the corresponding statements anywhere in the literature. Our starting point will consist in rewriting the *static* Bures-Wasserstein distance as a *dynamical* and more geometric problem, which is nothing but the celebrated Benamou-Brenier formulation [10] of optimal transport restricted to Gaussian measures. The resulting minimization immediately extends to the infinite-dimensional setting, thus giving rise to a geodesic problem: the minimization of an L^2 kinetic energy in the space of matrix-measure-valued curves, computed with respect to an infinite-dimensional version of the

Bures metric. This shows that the Hellinger space is a (formal) Riemannian manifold endowed with this extended quantum Fisher-Rao metric, and the Fisher-Rao space is simply a submanifold with induced Riemannian metric. This Riemannian structure allows for Riemannian computations in the spirit of Otto [59].

A physical quantity often appearing both in quantum information theory and in optimal transport is entropy. Entropy conveys significant information about the underlying geometric structures, in the sense that is is canonically associated with a corresponding geometric heat flow. This heat flow is nothing but the (negative) gradient-flow of the entropy, and in optimal transport the groundbreaking paper [42] led to whole variety of results ranging from applied PDEs [63] to Lott-Sturm-Villani synthetic curvature theory [52]. In the Bures-Wasserstein setting it turns out that the relevant notion of entropy is not the usual von Neumann entropy from quantum statistical mechanics, but is rather induced by the restriction of the Boltzmann-Shannon entropy (Kullback-Leibler divergence) to Gaussian optimal transport [9, 55]. In the infinite-dimensional case, the corresponding entropy is very much related to the classical Itakura-Saito divergence [39, 57, 41] from signal processing.

Over the last few years, a particular regularization of the optimal transport problem, the so-called *entropic optimal transport*, has received considerable attention and allowed for efficient numerical computations and theoretical advances [61, 21, 30]. This regularization is related to the Schrödinger problem [24, 18, 23, 48] and to *Euclidean quantum dynamics* [2, 71], and can be considered as a blurred version of deterministic optimal transport. Based on our dynamical framework and inspired from the very generic Schrödinger problem discussed formally in [47], our second main contribution in this paper consists in justifying an adapted definition of a canonical entropy on the non-commutative Fisher-Rao space, the derivation of the analogous heat flow and Fisher information, and of the corresponding *dynamical* Schrödinger problem. Let us point out that a related *static* Schrödinger problem has been considered very recently in [40], but only on the finite-dimensional statistical manifold of real nonsingular matrices $\mathcal{S}^{++}$. We would also like to mention the contributions [29, 64] that discuss an entropic regularization of the static commutative unbalanced optimal transport. A related regularization of the static non-commutative unbalanced optimal transport that employs the von Neumann entropy was suggested in [60].

Given a fixed functional on a Riemannian manifold, one can construct [44] two canonical evolutionary processes: the associated gradient flow (a dissipative system) and Newton's equation (a Hamiltonian system). The dynamical Schrödinger problem can be viewed as a third "sibling" in this geometric family, since it merely suffices to fix a functional (entropy) in order to define it, cf. [47]. Moreover, any trajectory of the gradient flow solves an appropriate Schrödinger problem, cf. [56]. On the other hand, a Schrödinger problem may be viewed as a Newton equation driven by a suitably defined analogue of the Fisher information. In this connection, let us mention the discussion of several Newton equations on the commutative Fisher-Rao space in [44].

Our second main contribution will therefore consist in the study of the Schrödinger problem on the Fisher-Rao space. We will prove that, as the temperature parameter $\varepsilon \to 0$, the ε-Schrödinger functional Gamma-converges towards the Fisher-Rao L^2 kinetic functional. As a consequence the Schrödinger minimizers converge to geodesics. An indirect byproduct of our explicit construction of the recovery sequences for the Γ-limit will be the $\frac{1}{2}$-geodesic convexity of the entropy in the Fisher-Rao space. This has been studied in [55] on the finite-dimensional cone space by direct computations involving second order derivatives, but our additional mass constraint and the measure-theoretic setting both make the analysis more delicate here. We note that our construction of recovery sequences for the Γ-convergence is reminiscent of [8], and we will treat abstract metric spaces in our upcoming work [56].

The paper is organized as follows: In section 2 we fix the notations and define the basic concepts to be used throughout. Section 3 contains preliminary material on general metric cones, the finite-dimensional Bures-Wasserstein distance, and the associated entropy, Fisher information, heat-flow, and Schrödinger problem. We then proceed with our comprehensive study of the general matrix-valued-measure setting and we introduce the Hellinger space in Section 4. Section 5 is concerned with the Fisher-Rao space, and also details the cone structure. In Section 6 we discuss the spherical Fisher-Rao Riemannian structure, we explain how to carry out the corresponding variant of the Otto calculus and how to compute Fisher-Rao gradients of functionals of measures. We also define our canonical entropy, we study its associated heat flow, and we show that the corresponding Schrödinger problem is well-posed. Our last Section 7 focuses on the Γ-convergence of the ε-Schrödinger problem towards the geodesic problem as well as the geodesic convexity of the entropy. Finally, we opted for postponing some technical statements and proofs to the Appendix A.

2. Notations and conventions

We will use the following basic notation:

- The space Ω is a fixed separable, locally compact, metrizable topological space.
- $\mathbb{C}^{d\times d}$ is the space of $d\times d$ complex matrices, equipped with the *real* Frobenius product and norm
$$\Phi:\Psi = \operatorname{Re}\operatorname{tr}(\Phi^*\Psi) \quad \text{and} \quad |\Phi|_2 = \sqrt{\Phi:\Phi}.$$
- $A^{Sym} := \frac{1}{2}(A+A^*)$ will denote the Hermitian part of $A\in\mathbb{C}^{d\times d}$.
- $I\in\mathbb{C}^{d\times d}$ is the identity matrix.
- $\mathcal{H}$ is the subspace of Hermitian $d\times d$ matrices.
- $\mathcal{H}^+$ is the subspace of Hermitian positive-semidefinite (PSD) $d\times d$ matrices.
- $\mathcal{H}^{++}$ is the subspace of Hermitian positive-definite $d\times d$ matrices.
- $\mathcal{S}$ is the subspace of real symmetric $d\times d$ matrices.
- $\mathcal{S}^+$ is the subspace of real symmetric positive-semidefinite matrices of size d.
- $\mathcal{S}^{++}$ is the subspace of real symmetric positive-definite $d\times d$ matrices.
- $\mathcal{P}$ is the subspace of Hermitian PSD matrices of size d and of unit trace.
- For $A\in\mathcal{S}^{++}$ we write $\mathcal{N}(A)=\mathcal{N}(0,A)$ for the Gaussian distribution on $\mathbb{R}^d$ with mean 0 and covariance A.
- We recall that for $U,V\in\mathcal{H}$ one has $\operatorname{tr}(UV)\in\mathbb{R}$. If the matrices U,V are positive-semidefinite, then $\operatorname{tr}(UV)\geq 0$. By the Cauchy-Schwarz inequality, $|gU:V|\leq |gU:U|^{\frac{1}{2}}|gV:V|^{\frac{1}{2}}$ for all $U,V\in\mathcal{H}$ and a PSD matrix $g\in\mathcal{H}^+$. We recall moreover the elementary inequality
$$|gU:U|\leq (\operatorname{tr} g)|U|_2^2. \tag{2.1}$$

- The square root $\sqrt{g}$ of a PSD matrix $g \in \mathcal{H}^+$ is
$$\sqrt{g} = R\,\mathrm{diag}(\sqrt{D_{ii}})R^*,$$
where $D = \mathrm{diag}(D_{ii})$ and $g = RDR^*$ is a spectral decomposition of g.
- The total variation norm of an $\mathcal{H}$-valued Radon measure G on Ω is
$$\|G\|_{TV} := \sup_{\Psi \in C_b(\Omega;\mathcal{H}), \|\Psi\|_\infty \le 1} \left| \int_\Omega \mathrm{d}G(x) : \Psi(x) \right|,$$
were $\|\Psi\|_\infty = \sup_{x\in\Omega} |\Psi(x)|_2$ is computed relatively to the Frobenius norm on $\mathcal{H}$. Clearly this norm is equivalent to
$$\sum_{i,j} \|G_{ij}\|_{TV},$$
where each $\|G_{ij}\|_{TV}$ is the total variation norm (in the usual sense) of the complex measure G_{ij}. Note that only Hermitian test-functions are needed since G itself is Hermitian.
- $\mathbb{H}$ is the set of $\mathcal{H}$-valued *finite* Radon measures G on Ω, i-e with $\|G\|_{TV} < +\infty$.
- $\mathbb{H}^+ \subset \mathbb{H}$ is the set of $\mathcal{H}^+$-valued finite Radon measures G on Ω.
- $\mathbb{P} \subset \mathbb{H}^+$ is the set of $\mathcal{H}^+$-valued measures G on Ω with
$$\mathrm{tr}(G(\Omega)) = 1.$$
These will be our main objects, and can be considered as non-commutative Probability measures (hence the notation), cf. Remark 5.1.
- $\mathbb{P}_\Lambda \subset \mathbb{P}$ is the set of $\mathcal{H}^+$-valued measures G on Ω with unit trace and absolutely continuous w.r.t. to a reference scalar measure Λ, in the sense that $\gamma = \mathrm{tr}\, G \ll \Lambda$.
- $\mathbb{P}^{++}_\Lambda \subset \mathbb{P}_\Lambda$ is the set of $\mathcal{H}^{++}$-valued measures $G \in \mathbb{P}_\Lambda$.
- For $G \in \mathbb{H}^+$ we denote the trace measure $\gamma := \mathrm{tr}\, G$, which is a non-negative scalar measure. The mass of G is $m = \|\gamma\|_{TV} = \gamma(\Omega)$. Note that the (trace) Schatten 1-norm $|g|_1 := \mathrm{tr}(\sqrt{gg^*}) = \mathrm{tr}\, g \ge C_d |g|_2$ controls the Frobenius norm on $\mathcal{H}^+$, hence the mass controls $\|G\|_{TV} \lesssim m$ on $\mathbb{H}^+$.
- Whenever λ is a positive scalar measure and $G \in \mathbb{H}^+$ we abuse the notations and write $G \ll \lambda$ or $G \ll \lambda I$ for $\gamma = \mathrm{tr}\, G \ll \lambda$. (This is *not* equivalent to requiring that $|G_{ij}| \ll \lambda$ for all i, j.) In particular

we always have $G \ll \gamma = \operatorname{tr} G$. The Radon-Nikodym density $g(x) = \frac{\mathrm{d}G}{\mathrm{d}\gamma}(x) \in \mathcal{P}$ is unitary in the sense that automatically $\operatorname{tr} g(x) \equiv 1$.

- We say that a sequence of matrix-valued measures $G^k \in \mathbb{H}^+$ converges *strongly*

$$G^k \to G \quad \text{if} \quad \|G^k - G\|_{TV} \to 0.$$

- We say that a sequence of matrix-valued measures $G^k \in \mathbb{H}^+$ converges weakly-$*$

$$G^k \overset{*}{\rightharpoonup} G \quad \text{if} \quad \int_\Omega \varphi : \mathrm{d}G^k \to \int_\Omega \varphi : \mathrm{d}G_t$$

for all $\varphi \in C_0(\Omega;\mathcal{H})$, the closure (w.r.t. the uniform norm) of compactly supported functions $C_c(\Omega;\mathcal{H})$. (This is the predual convergence on $\mathbb{H}^+ \subset (C_0(\Omega;\mathcal{H}))^*$.)
- For curves $t \in [0,1] \mapsto G_t \in \mathbb{H}^+$ we write $G \in C([0,1];\mathbb{H}^+_{TV})$ for the continuity with respect to the strong TV topology. We should anticipate at this stage that we will define two d_H, d_{FR} (Hellinger and Fisher-Rao) distances on $\mathbb{H}^+, \mathbb{P}$, but the corresponding topologies will always be sequentially equivalent to TV. As a consequence all curves will be continuous indistinctly w.r.t. any of these topologies and we will often omit the index for brevity.
- Whenever $G \in \mathbb{H}^+$ one can define the $L^2(\mathrm{d}G;\mathcal{H})$ space of $\mathcal{H}$-valued functions through $\|U\|^2_{L^2(\mathrm{d}G)} := \int_\Omega \mathrm{d}GU : U$. The Hermitian inner product is defined accordingly, and we reger to [33] for further properties of noncommutative L^p spaces.
- For a given curve $G \in C([0,1];\mathbb{H}^+)$ the weighted

$$L^2(0,1;L^2(\mathrm{d}G_t;\mathcal{H}))$$

space is defined by disintegration $\mathrm{d}t \otimes \mathrm{d}G_t$, with

$$\|U\|^2_{L^2(0,1;L^2(\mathrm{d}G_t))} := \int_0^1 \int_\Omega \mathrm{d}G_t U_t : U_t \,\mathrm{d}t.$$

Similar definitions can be used for $L^p(0,1;L^2(\mathrm{d}G_t;\mathcal{H}))$, $p \in [1,\infty]$.
- In a given metric space (X,d) we say that a curve $x : t \in [0,1] \mapsto x_t \in X$ is $AC^p(0,1;X)$ (absolutely continuous) if there exists an $L^p(0,1)$

function $\eta(\tau) \geq 0$ such that

$$d(x_s, x_t) \leq \int_s^t \eta(\tau)\,\mathrm{d}\tau, \qquad \forall\, 0 \leq s \leq t \leq 1.$$

In this case the *metric derivative*

$$|\dot{x}_t| := \lim_{\delta \to 0} \frac{d(x_t, x_{t+\delta})}{\delta}$$

exists almost everywhere, and it is the smallest function $\eta(\tau)$ satisfying the above inequality, [5, thm. 1.1.4].

- By *geodesics* we always mean constant-speed, minimizing metric geodesics.
- We will use three separate distances d_H, d_{FR}, TV on the spaces $\mathbb{H}^+, \mathbb{P}$, and the corresponding metric speeds $|\dot{G}_t|_H, |\dot{G}_t|_{FR}, |\dot{G}_t|_{TV}$ of a single curve $(G_t)_{t\in[0,1]}$ may a priori differ. We will try to emphasize the difference by writing

$$AC^p(0,1;\mathbb{P}_H),\, AC^p(0,1;\mathbb{P}_{FR}),\, AC^p(0,1;\mathbb{P}_{TV})$$

depending on the context. (H and FR will however coincide.)
- C is a generic positive constant.

Finally, we fix once and for all a reference scalar, non-negative Radon measure Λ on Ω such that $\mathrm{tr}(\Lambda I)$ is a probability measure. (The reader might think of Λ as of a normalized Lebesgue measure.)

3. Preliminaries

3.1. Metric cones. We recall [16, 17] that, given a metric space (X, d_X) of diameter $\leq \pi$, one can define another metric space

$$(\mathfrak{C}(X), d_{\mathfrak{C}(X)}),$$

called a *cone* over X, in the following manner. Consider the quotient $\mathfrak{C}(X) := X \times [0,\infty)/X \times \{0\}$, that is, all points of the fiber $X \times \{0\}$ constitute a single point of the cone called the apex. In other words, points in $\mathfrak{C}(X)$ are of the form $[x, r]$ and we identify points $[x_0, 0] \sim [x_1, 0]$ for all $x_0, x_1 \in X$. Now set

$$d^2_{\mathfrak{C}(X)}([x_0, r_0], [x_1, r_1]) := r_0^2 + r_1^2 - 2 r_0 r_1 \cos(d_X(x_0, x_1)). \tag{3.1}$$

Very few metric spaces are actually cones, and this property provides neat scaling and other nice geometric features [46]. A particularly regular situation appears when the diameter of X is strictly less than π, since in this case there is a one-to-one correspondence between the geodesics in X and $\mathfrak{C}(X)$. Given a cone $Y = \mathfrak{C}(X)$, X can be embedded canonically as a *sphere* into Y via $X \cong \{[x,1] : x \in X\} \subset \mathfrak{C}(X)$. In particular,

$$d_{\mathrm{conic}}(x_0,x_1) := d_{\mathfrak{C}(X)}([x_0,1],[x_1,1])$$

defines a distance on the sphere X. On the sphere, the relation (3.1) may be inverted:

$$d_X(x_0,x_1) = \arccos\left(1 - \frac{1}{2} d^2_{\mathrm{conic}}(x_0,x_1)\right). \tag{3.2}$$

Lemma 3.1 ([15]). *If X is a length space, and $Y = \mathfrak{C}(X)$, then the distance $d_X(x_0,x_1)$ coincides with the infimum of Y-lengths of continuous curves $([x_t,1])_{t\in[0,1]}$ joining $[x_0,1]$ and $[x_1,1]$ and lying within $X \times \{1\}$.*

Lemma 3.2 ([46]). *Let $([x_t,r_t])_{t\in[0,1]}$ be a constant-speed geodesic in the metric cone $Y = \mathfrak{C}(X)$. Then*

$$r_t^2 = tr_1^2 + (1-t)r_0^2 - t(1-t)d_Y^2([x_0,r_0],[x_1,r_1]) \le tr_1^2 + (1-t)r_0^2. \tag{3.3}$$

Lemma 3.3. *The distances d_X and d_{conic} are Lipschitz-equivalent on X: namely, one has $d_{\mathrm{conic}} \le d_X \le c d_{\mathrm{conic}}$. Here $c = \frac{\arccos\left(1-D^2/2\right)}{D} \le \frac{\pi}{2}$, where $D \le 2$ is the diameter of (X, d_{conic}).*

The proof is immediate by Lemma 3.1 and convexity of

$$z \mapsto \arccos\left(1 - z^2/2\right).$$

3.2. The Bures-Wasserstein distance. A notable finite-dimensional example of a metric cone structure is provided by the Bures-Wasserstein distance [12, 32, 11] on the space $\mathcal{H}^+$ of positive-semidefinite Hermitian matrices. We recall that the Bures-Wasserstein distance can be computed in three different equivalent ways. The first one is the explicit formula

$$\begin{aligned} d_B^2(A_0,A_1) :=& \ \operatorname{tr} A_0 + \operatorname{tr} A_1 - 2\operatorname{tr}\sqrt{\sqrt{A}_0 A_1 \sqrt{A}_0} \\ =& \ \operatorname{tr} A_1 + \operatorname{tr} A_0 - 2\operatorname{tr}\sqrt{\sqrt{A}_1 A_0 \sqrt{A}_1} \end{aligned} \tag{3.4}$$

Note that when A_0 and A_1 commute this reads explicitly $d_B^2(A_0,A_1) = |\sqrt{A}_1 - \sqrt{A}_0|_2^2$.

The second equivalent formulation is more geometric.

Proposition 3.4. *The Bures-Wasserstein space is a geodesic space. Moreover, we have*

$$d_B^2(A_0,A_1) = \frac{1}{4}\min_{\mathcal{A}(A_0,A_1)} \int_0^1 A_t U_t : U_t \mathrm{d}t, \tag{3.5}$$

where the admissible set $\mathcal{A}(A_0,A_1)$ consists of all pairs $(A_t,U_t)_{t\in[0,1]}$ such that

$$\begin{cases} A \in C^1([0,1];\mathcal{H}^+), U \in C^1([0,1];\mathcal{H}) \\ A|_{t=0} = A_0; \quad A|_{t=1} = A_1, \end{cases} \quad \textit{and} \quad \frac{dA_t}{dt} = (A_t U_t)^{Sym}. \tag{3.6}$$

The Bures-Wasserstein geodesics in $\mathcal{H}^+$ correspond to minimizers of (3.5)*.*

We believe that the claim is known to the experts in the field, at least for the non-singular matrices A_0, A_1. We however failed to find an explicit reference (in particular for possibly degenerate matrices), and we provide an independent proof in the Appendix.

The third possible formulation only works for *nonsingular* (definite) matrices $A \in \mathcal{H}^{++}$, and provides an explicit relation with the quadratic Wasserstein distance W_2 on the space of probability measures on Euclidean spaces. Let us start from the more classical real-valued case $A \in \mathcal{S}^{++}(d)$, and recall that we write $\mathcal{N}(A) = \mathcal{N}(0,A)$ for the corresponding multivariate normal distribution with mean $0 \in \mathbb{R}^d$ and covariance A. Then

Proposition 3.5 ([3, 65])**.** *For any $A_0, A_1 \in \mathcal{S}^{++}$, we have*

$$d_B(A_0,A_1) = W_2\Big(\mathcal{N}(A_0), \mathcal{N}(A_1)\Big). \tag{3.7}$$

Actually, one can go beyond (3.7) and show that the *real* subspace $\mathcal{S}^{++}$ of the Bures space may be viewed as a totally geodesic submanifold of the space of probability measures on $\mathbb{R}^d$ equipped with the Otto-Wasserstein Riemannian structure, cf. [55, 40].

Let us now treat the general (complex) case $A \in \mathcal{H}^{++}(d)$. It is natural to substitute every complex entry of a Hermitian matrix A with the real (2×2) block

$$a_{ij} = x_{ij} + i y_{ij} \longrightarrow \begin{bmatrix} x_{ij} & -y_{ij} \\ y_{ij} & x_{ij} \end{bmatrix} \tag{3.8}$$

This defines an inclusion function $\mathfrak{r} : \mathcal{H}^{++}(d) \to \mathcal{S}^{++}(2d)$. Using the definition (3.4) it is easy to check that

$$d_B^2(\mathfrak{r}(A_0), \mathfrak{r}(A_1)) = 2d_B^2(A_0, A_1), \tag{3.9}$$

and employing Proposition 3.5 we immediately conclude that

Proposition 3.6. *For any $A_0, A_1 \in \mathcal{H}^{++}$, we have*

$$d_B(A_0, A_1) = \frac{1}{\sqrt{2}} W_2\Big(\mathcal{N}(\mathfrak{r}(A_0)), \mathcal{N}(\mathfrak{r}(A_1))\Big). \tag{3.10}$$

Consequently, $\mathcal{H}^{++}$ may be viewed as a totally geodesic submanifold of the Otto-Wasserstein space of probability measures on $\mathbb{R}^{2d}$.

For any element $A \in \mathcal{H}^+$, we set

$$r = r(A) = \sqrt{\operatorname{tr} A}.$$

Then we can identify A with a pair $[A/r^2, r] \in \mathfrak{C}(\mathcal{P})$. (The first factor is normalized to unity in the sense that $\operatorname{tr}\big(A/r^2\big) = 1$ and we think of $\mathcal{P} = \{A \in \mathcal{H}^+ : \operatorname{tr} A = 1\}$ as the sphere of radius $r = 1$ in $\mathcal{H}^+$.) The related *spherical* Bures distance [67, 11] on $\mathcal{P}$ is defined by

$$d_{SB}^2(A_0, A_1) := \arccos\left(1 - \frac{1}{2} d_B^2(A_0, A_1)\right), \qquad A_0, A_1 \in \mathcal{P}, \tag{3.11}$$

consistently with (3.2). The fact that the Bures space is actually a metric cone might be well-known to the community but we never saw it explicitly written down.

Proposition 3.7. *The space $(\mathcal{H}^+, d_B)$ is a metric cone over $(\mathcal{P}, d_{SB})$, where $\mathcal{H}^+$ is identified with $\mathfrak{C}(\mathcal{P})$ via $A \simeq [A/r^2, r]$.*

We omit the proof since a more general statement will be provided in Theorem 5.

The one-to-one correspondence between Hermitian PSD matrices and real Gaussian distributions from Proposition 3.6 allows to define an analogue of the Boltzmann entropy on the Bures-Wasserstein space, cf. [55]. Indeed, the usual (negative) Boltzmann entropy $H(\rho) = \int_{\mathbb{R}^{2d}} \rho \log \rho$ of a multivariate Gaussian distribution $\rho = \mathcal{N}(A)$ reads explicitly [1]

$$H(\mathcal{N}(A)) = -\frac{1}{2}(\log \det(2\pi e A)).$$

This suggests defining the Bures entropy of a PSD matrix $A \in \mathcal{H}^{++}$ as the Boltzmann entropy of the associated Gaussian distribution

$$E_B(A) := -\frac{1}{2}(\log\det(2\pi e \mathfrak{r}(A))),$$

cf. [34, 9]. It will be more convenient for our purpose to use a "modulated" version (*à la* Bregman), making the entropy non-negative and attaining its minimum value (zero) at I:

$$E(A) := \frac{1}{2}\left[\operatorname{tr}(\mathfrak{r}(A)) - \operatorname{tr}(\mathfrak{r}(I)) - \log\det(\mathfrak{r}(A))\right] = \operatorname{tr} A - \operatorname{tr} I - \log\det A. \quad (3.12)$$

Note that for $A \in \mathcal{S}^{++}$ this is exactly the (doubled) Kullback–Leibler divergence

$$E(A) = 2D_{KL}\Big(\mathcal{N}(A)\Big|\mathcal{N}(I)\Big) = 2\int_{\mathbb{R}^d} \frac{\rho_A}{\rho_I}\log\left(\frac{\rho_A}{\rho_I}\right) \mathrm{d}\rho_I$$

from $\rho_A = \mathcal{N}(A)$ to the standard normal distribution $\rho_I = \mathcal{N}(I)$. Note also that this differs from the von Neumann entropy

$$S(A) = \operatorname{tr}(A\log A) = \sum_{\lambda\in\sigma_A} \lambda\log\lambda.$$

The entropy S is not geodesically convex on the Bures spaces $(\mathcal{H}^+, d_B)$ and $(\mathcal{P}, d_{SB})$ (One can mimic the argument from [50, Section 5.1] to show this for the cone $\mathcal{H}^+$; The case of the sphere $\mathcal{P}$ is left as an exercise for the reader.) However, our entropy E in (3.12) turns out to be geodesically convex both on $(\mathcal{H}^+, d_B)$ and $(\mathcal{P}, d_{SB})$. For $\mathcal{H}^+$ this is proved in [55] (for real-valued matrices), and for $\mathcal{P}$ it will follow from our more general Theorem 8.

3.3. The heat flow and the Schrödinger problem on the Bures-Wasserstein space. The usual heat flow on $\mathbb{R}^d$

$$\partial_t\rho_t = \Delta\rho_t \quad (3.13)$$

is known [42, 5, 68] to be the gradient flow of the Boltzmann entropy $H(\rho) = \int \rho\log\rho$ on the Otto-Wasserstein space of probability measures on $\mathbb{R}^d$. It is easy to check by hand that the set of multivariate normal distributions of zero mean is invariant w.r.t. (3.13). Thus, one can consider the restriction $\rho_t = \mathcal{N}(A_t)$ of the heat flow onto the real Bures subspace $\mathcal{S}^{++}$, cf. [55]. The resulting evolution reads

$$\frac{dA_t}{dt} = 2I, \tag{3.14}$$

cf. [53] (the covariance grows linearly). Similarly, the Wasserstein gradient flow of the relative entropy (Kullback–Leibler divergence) $\int_{\mathbb{R}^d} \rho \log \frac{\rho}{\mathcal{N}(I)}$ is the Fokker-Planck equation

$$\partial_t \rho_t = \Delta \rho_t - \operatorname{div}(\rho_t \nabla \log \mathcal{N}(I)),$$

cf. [42, 5, 68]. Note that this is exactly the heat flow if $\mathbb{R}^d$ is viewed as a Riemannian manifold with Gaussian volume form, cf. [69]. One easily checks that if the initial datum $\rho_0 = \mathcal{N}(A_0)$ is Gaussian then the solution remains Gaussian $\rho_t = \mathcal{N}(A_t)$, and the corresponding flow on the Bures space reads [55]

$$\frac{dA_t}{dt} = 2(I - A_t). \tag{3.15}$$

Both (3.14) and (3.15) actually make sense on the whole Bures space $\mathcal{H}^+$, i-e even when A are complex-valued and merely semi-definite. Moreover, it can be rigorously justified (e.g., by mimicking [55]), at least if we restrict ourselves to $\mathcal{H}^{++}$, that the extended flows (3.14) and (3.15) are exactly the Bures-Wasserstein gradient flows of the entropies $\frac{1}{2}E_B(A)$ and $\frac{1}{2}E(A)$, respectively.

The production of the Boltzmann entropy $-\frac{d}{dt}\int_{\mathbb{R}^d} \rho_t \log \rho_t = \int_{\mathbb{R}^d} \frac{|\nabla \rho_t|^2}{\rho_t}$ along the heat flow (3.13) is the celebrated Fisher information. It is a key ingredient in the formulation of the time-symmetric dynamical version of the Schrödinger problem [24, 18, 23, 47]:

$$\frac{1}{2}\int_0^1 \left(\int_{\mathbb{R}^d} \left(\rho_t |u_t|^2 + \epsilon^2 \frac{|\nabla \rho_t|^2}{\rho_t} \right) \mathrm{d}x \right) \mathrm{d}t \qquad \to \min. \tag{3.16}$$

Here the unknown probability density ρ_t and velocity field u_t are related by the transport equation $\partial_t \rho_t + \operatorname{div}(\rho_t u_t) = 0$; the initial and final configurations ρ_0, ρ_1 are prescribed and $\varepsilon > 0$ is a temperature parameter. Just as the heat flow, the Schrödinger problem also leaves the the Gaussian manifold invariant:

Proposition 3.8. *Assume that $\rho_0 = \mathcal{N}(A_0), \rho_1 = \mathcal{N}(A_1)$ are Gaussians. Then the solution to* (3.16) *remains Gaussian, $\rho_t = \mathcal{N}(A_t)$ for some explicitly computable $A_t \in \mathcal{S}^{++}$.*

We could not find this statement anywhere in the literature and postpone a sketch of proof to the appendix.

In the spirit of [47], this suggests considering an analogue of the dynamical Schrödinger problem (3.16) on the Bures space $\mathcal{H}^+$, just as we have done above for the heat flow. Related issues for the static Schrödinger problem were discussed very recently in [40]. The corresponding Fisher information, i-e the production of the entropy $\frac{1}{2}E_B(A)$ along the restricted heat flow (3.14), turns out to be $\operatorname{tr} A^{-1}$. The resulting Schrödinger problem reads

$$\frac{1}{2}\int_0^1 \left(A_t U_t : U_t + \varepsilon^2 \operatorname{tr} A_t^{-1}\right) \mathrm{d}t \qquad \to \min, \tag{3.17}$$

where $(A_t, U_t)_{t\in[0,1]}$ belongs to the admissible set $\mathcal{A}(A_0, A_1)$ from Proposition 3.4, i-e satisfy the continuity equation $\frac{dA_t}{dt} = (A_t U_t)^{Sym}$. As the temperature $\varepsilon \to 0$ the solutions (*Schrödinger bridges*) are expected to approximate the geodesics on $(\mathcal{H}^+, d_B)$, cf. (3.5). We will prove a related but more general statement later on, Corollary 7.7.

4. The Hellinger distance

The Hellinger distance on $\mathbb{H}^+$ can be defined as follows.

Definition 4.1 (Hellinger distance)**.** *Given two matrix measures $G_0, G_1 \in \mathbb{H}^+$ we define*

$$d_H^2(G_0, G_1) := \inf_{\mathcal{A}(G_0,G_1)} \int_0^1 \left(\int_\Omega \mathrm{d}G_t U_t : U_t\right) \mathrm{d}t, \tag{4.1}$$

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

$$\begin{cases} G \in C([0,1]; \mathbb{H}^+_{TV}), \\ G|_{t=0} = G_0; \quad G|_{t=1} = G_1, \quad \text{and } \partial_t G_t = (G_t U_t)^{Sym} \text{ in the weak sense.} \\ U \in L^2(0,1; L^2(\mathrm{d}G_t; \mathcal{H})) \end{cases} \tag{4.2}$$

For purely aesthetic reasons we dropped a 1/4 factor in (4.1) compared to (3.5).

We say that $\partial_t G_t = (G_t U_t)^{Sym}$ holds in the weak sense if for all test-functions $\Psi \in C^1([0,1]; C_b(\Omega; \mathcal{H}))$ there holds

$$\int_\Omega \Psi_t : \mathrm{d}G_t - \int_\Omega \Psi_s : \mathrm{d}G_s = \int_s^t \left(\int_\Omega \mathrm{d}G_\tau : \partial_\tau \Psi_\tau + \mathrm{d}G_\tau U_\tau : \Psi_\tau \right) \mathrm{d}\tau, \qquad \forall s,t \in [0,1]. \quad (4.3)$$

We will see shortly that this automatically implies that G is absolutely continuous in time w.r.t. the TV norm, hence the ODE $\partial_t G_t = (G_t U_t)^{Sym}$ can also be understood as pointwise a.e. $t \in (0,1)$ with values in the Banach space $\mathbb{H}_{TV}$ and boundary data taken in the strong sense. More precisely,

Lemma 4.2. *Let $(G_t, U_t)_{t\in[0,1]}$ be a solution of (4.3) with $G \in C([0,1]; \mathbb{H}^+_{TV})$ and mass $m_t := \|\operatorname{tr} G_t\|_{TV}$. Set*

$$E := \|U\|^2_{L^2(0,1;L^2(dG_t))} \quad \text{and} \quad M := \max_{t\in[0,1]} m_t.$$

Then the map $t \mapsto G_t$ is AC^2 for the TV norm, there holds

$$\|G_t - G_s\|_{TV} \le \sqrt{ME}|t-s|^{\frac{1}{2}}, \qquad \forall\, s,t \in [0,1] \quad (4.4)$$

and we have the mass estimate

$$M \le E + 2\min\{m_0, m_1\}. \quad (4.5)$$

Proof: Note first that by continuity the mass is bounded from above, i-e $M = \max m_t < +\infty$. In order to estimate M we first take $\psi \equiv I$ in (4.3). The Cauchy-Schwarz inequality yields

$$\begin{aligned} |m_t - m_s| &= \left| \int_s^t \int_\Omega \mathrm{d}G_\tau U_\tau : I \,\mathrm{d}\tau \right| \\ &\le \left(\int_s^t \int_\Omega \mathrm{d}G_\tau U_\tau : U_\tau \,\mathrm{d}\tau \right)^{\frac{1}{2}} \left(\int_s^t \int_\Omega \mathrm{d}G_\tau I : I \,\mathrm{d}\tau \right)^{\frac{1}{2}} \\ &= \sqrt{E} \times \left(\int_s^t \|\operatorname{tr} G_\tau\|_{TV} \right)^{\frac{1}{2}} \le \sqrt{E} \times (M|t-s|)^{\frac{1}{2}}, \end{aligned}$$

where we used $\|\operatorname{tr} G_\tau\|_{TV} = m_\tau \le M$. Taking now $s = 0$ and picking any time t where $M = \max m_\tau = m_t$ is attained gives $M \le m_0 + \sqrt{E}\sqrt{M}|t-0|^{\frac{1}{2}} \le m_0 + \sqrt{E}\sqrt{M}$. Elementary algebra guarantees in turn $M \le E + 2m_0$. By symmetry we get $M \le E + 2m_1$ as well and (4.5) follows.

In order to get the absolute continuity in time, take now a time-independent test-function $\psi \in C_b(\Omega;\mathcal{H})$ and fix any $s \le t$. Then (4.3) and (2.1) imply

$$\begin{aligned}\left|\int_\Omega \mathrm{d}(G_t - G_s) : \psi\right| &= \left|\int_s^t \int_\Omega \mathrm{d}G_\tau U_\tau : \psi \,\mathrm{d}\tau\right| \\ &\le \int_s^t \left(\int_\Omega \mathrm{d}G_\tau U_\tau : U_\tau\right)^{\frac12} \left(\int_\Omega \mathrm{d}G_\tau \psi : \psi\right)^{\frac12} \mathrm{d}\tau \\ &\le \int_s^t \left(\int_\Omega \mathrm{d}G_\tau U_\tau : U_\tau\right)^{\frac12} \left(m_\tau \|\psi\|_\infty^2\right)^{\frac12} \mathrm{d}\tau \\ &\le \sqrt{M}\|\psi\|_\infty \int_s^t \left(\int_\Omega \mathrm{d}G_\tau U_\tau : U_\tau\right)^{\frac12} \mathrm{d}\tau.\end{aligned}$$

Taking the supremum over ψ's such that $\|\psi\|_\infty \le 1$ gives

$$\|G_t - G_s\|_{TV} \le \sqrt{M} \int_s^t \underbrace{\left(\int_\Omega \mathrm{d}G_\tau U_\tau : U_\tau\right)^{\frac12}}_{\in L^2(0,1)} \mathrm{d}\tau$$

and entails the AC^2 regularity in total variation. Applying finally the Cauchy-Schwarz inequality (in time) gives

$$\|G_t - G_s\|_{TV} \le \sqrt{M}\left(\int_s^t \int_\Omega \mathrm{d}G_\tau U_\tau : U_\tau \mathrm{d}\tau\right)^{\frac12} |t-s|^{\frac12} \le \sqrt{ME}|t-s|^{\frac12}$$

and concludes the proof. ∎

Theorem 1. *d_H is a distance on $\mathbb{H}^+$.*

Proof: The argument is quite standard and we only sketch the details. First of all, given any G_1 it is easy to see that $(G_t, U_t) := (t^2 G_1, \frac{2}{t} I)$ gives an admissible path connecting 0 to G_1 with finite cost. As a consequence any two points G_0, G_1 can be connected going through zero, $G_0 \rightsquigarrow 0 \rightsquigarrow G_1$ each in time 1/2, and $d_H^2(G_0, G_1)$ is therefore always finite.

Assume now that $d_H^2(G_0, G_1) = 0$, and let $(G_t^n, U_t^n)_{t\in[0,1]}$ be a minimizing curve with energy $E_n = E[G^n, U^n] \to 0$. Then Lemma 4.2 guarantees that

$$\|G_1 - G_0\|_{TV} \le \sqrt{E^n + 2\min\{m_0, m_1\}}\sqrt{E_n} \to 0$$

hence $G_0 = G_1$.

Finally for the triangular inequality, fix any $G_0, G_1 \in \mathbb{H}^+$ and take any $\tilde{G} \in \mathbb{H}^+$. Consider two minimizing sequences (G^{n0}, U^{n0}) and (G^{n1}, U^{n1}) in the definitions of $d_H^2(G_0, \tilde{G})$ and $d_H^2(\tilde{G}, G_1)$, respectively, both in time $t \in [0,1]$. For any fixed $\theta \in (0,1)$ it is easy to scale the path G^{n0} in time $[0,\theta]$, rescale the path G^{n1} in time $[\theta,1]$, and concatenate them to produce an admissible path $(\check{G}^n, \check{U}^n)_{t\in[0,1]}$ connecting $\check{G}^n_0 = G_0 \rightsquigarrow \check{G}^n_\theta = \tilde{G} \rightsquigarrow \check{G}^n_1 = G_1$. The scaling property shows that the energy of the resulting path is

$$d_H^2(G_0, G_1) \le E[\check{G}^n, \check{U}^n] = \frac{1}{\theta} E[G^{n0}, U^{n0}] + \frac{1}{1-\theta} E[G^{n1}, U^{n1}]$$

and therefore taking $n \to \infty$ gives

$$d_H^2(G_0, G_1) \le \frac{1}{\theta} d_H^2(G_0, \tilde{G}) + \frac{1}{1-\theta} d_H^2(\tilde{G}, G_1).$$

Choosing $\theta = \frac{d_H(G_0,\tilde{G})}{d_H(G_0,\tilde{G}) + d_H(\tilde{G},G_1)}$ finally yields

$$\frac{1}{\theta} d_H^2(G_0, \tilde{G}) + \frac{1}{1-\theta} d_H^2(\tilde{G}, G_1) = \left[d_H(G_0, \tilde{G}) + d_H(\tilde{G}, G_1) \right]^2$$

and achieves the proof. ■

Recalling that we write $d_B(g_0, g_1)$ for the Bures distance between $g_0, g_1 \in \mathcal{H}^+$, we have next

Lemma 4.3. *The Hellinger distance can be computed by the formula*

$$d_H^2(G_0, G_1) = 4 \int_\Omega d_B^2(G_0, G_1). \tag{4.6}$$

Note that this integral is well-defined by 1-homogeneity of d_B^2 (which is clear from (3.4)), i-e we mean here

$$\int_\Omega d_B^2(G_0, G_1) := \int_\Omega d_B^2\left(\frac{\mathrm{d}G_0}{\mathrm{d}\lambda}(x), \frac{\mathrm{d}G_1}{\mathrm{d}\lambda}(x) \right) \mathrm{d}\lambda(x)$$

for any positive scalar measure λ dominating simultaneously (the traces of) G_0, G_1. (The integral is independent of the choice of λ.)

Proof: Let first us prove that $d_H^2 \le 4 \int d_B^2$. To this end, fix any scalar measure λ dominating G_0, G_1, and denote the corresponding $\mathcal{H}^+$-valued densities $g_0(x) := \frac{\mathrm{d}G_0}{\mathrm{d}\lambda}(x), g_1(x) := \frac{\mathrm{d}G_1}{\mathrm{d}\lambda}(x)$. For λ-a.e. $x \in \Omega$ fixed, consider a minimizing pair curve/potential in the dynamical definition (Proposition 3.4) of the Bures distance between $g_0(x), g_1(x)$: This is a pair $U_t = U_t(x)$ with values in $\mathcal{H}$ and $g_t = g_t(x)$ with values in $\mathcal{H}^+$ such that

$$d_B^2(g_0, g_1) = \frac{1}{4}\int_0^1 g_t U_t : U_t \,\mathrm{d}t, \qquad \frac{d}{dt} g_t = (g_t U_t)^{Sym}.$$

Defining the matrix-valued measures

$$\mathrm{d}G_t := g_t(x) \cdot \mathrm{d}\lambda$$

it is easy to see that the weak formulation (4.3) of $\partial_t G_t = (G_t U_t)^{Sym}$ is satisfied, and the curve $(G_t, U_t)_{t\in[0,1]}$ is an admissible competitor in the minimization (4.1) for $d_H^2(G_0, G_1)$. By definition of G_t and Fubini's theorem we get

$$\begin{aligned}
d_H^2(G_0, G_1) \leq \int_0^1 \int_\Omega \mathrm{d}G_t(x) U_t(x) : U_t(x) \,\mathrm{d}t \\
&= \int_0^1 \int_\Omega \mathrm{d}\lambda(x) g_t(x) U_t(x) : U_t(x) \,\mathrm{d}t \\
&= \int_\Omega \left(\int_0^1 g_t(x) U_t(x) : U_t(x) \,\mathrm{d}t \right) \mathrm{d}\lambda(x) \\
&= 4 \int_\Omega d_B^2(g_0(x), g_1(x)) \,\mathrm{d}\lambda(x) = 4 \int_\Omega d_B^2(G_0, G_1).
\end{aligned}$$

Let us now establish the reversed inequality $4\int d_B^2 \leq d_H^2$. To this end, pick a minimizing sequence (G_t^n, U_t^n) in the definition (4.1) of $d_H^2(G_0, G_1)$, and fix any positive scalar measure λ dominating simultaneously G_0, G_1. We claim that we can assume $G_t^n \ll \lambda$ as well for any intermediate time $t \in (0,1)$. For if not, the linearity of the Lebesgue decomposition $G_t^n = (G_t^n)^\lambda + (G_t^n)^\perp$ with respect to λ (for any fixed time) easily shows that $\partial_t G_t^n = (G_t^n U_t^n)^{Sym}$ is equivalent to the two separate ODEs $\partial_t (G_t^n)^\lambda = ((G_t^n)^\lambda U_t^n)^{Sym}$ and $\partial_t (G_t^n)^\perp = ((G_t^n)^\perp U_t^n)^{Sym}$. Since $G_0^\perp = G_1^\perp = 0$ clearly $((G_n^n)^\lambda, U_t^n)_{t\in[0,1]}$ is an admissible path connecting $G_0 = G_0^\lambda, G_1 = G_1^\lambda$. Of course we have $(G_t^n)^\lambda \ll \lambda$ for all t, and

$$\int_0^1 \int_\Omega \mathrm{d}(G_n^n)^\lambda U_t^n : U_t^n \,\mathrm{d}t \leq \int_0^1 \int_\Omega \mathrm{d}G_t^n U_t^n : U_t^n \,\mathrm{d}t$$

gives a lesser cost so $((G_t^n)^\lambda, U_t^n)$ is a better competitor.

Thus assuming that $G^n_t \ll \lambda$ for all times, and writing as before $g^n_t(x) := \frac{\mathrm{d}G^n_t}{\mathrm{d}\lambda}(x)$ for the corresponding densities, the measure-valued ODE simply means now

$$\frac{d}{dt}g^n_t(x) = (g^n_t(x)U^n_t(x))^{Sym} \quad \text{with endpoints} \quad g_0(x), g_1(x)$$

for λ-a.e. $x \in \Omega$. In particular (g^n_t, U^n_t) is an admissible curve connecting g_0, g_1 (in each fiber $x \in \Omega$), it is therefore an admissible competitor in the characterization (3.5) of $d^2_B(g_0(x), g_1(x))$ for λ-a.e. x, thus

$$\begin{aligned}
4\int_\Omega d^2_B(G_0, G_1) = 4\int_\Omega d^2_B(g_0(x), g_1(x))\,\mathrm{d}\lambda(x) \\
&\leq \int_\Omega \left(\int_0^1 g^n_t(x)U^n_t(x) : U^n_t(x)\,\mathrm{d}t\right)\mathrm{d}\lambda(x) \\
&= \int_0^1 \int_\Omega g^n_t(x)U^n_t(x) : U^n_t(x)\,\mathrm{d}\lambda(x)\,\mathrm{d}t \\
&= \int_0^1 \int_\Omega \mathrm{d}G^n_t U^n_t : U^n_t\,\mathrm{d}t \xrightarrow[n\to\infty]{} d^2_H(G_0, G_1)
\end{aligned}$$

as desired. ∎

An immediate and important consequence of this is:

Theorem 2 (Existence of Hellinger geodesics)**.** *$(\mathbb{H}^+, d_H)$ is a geodesic space, i-e for all $G_0, G_1 \in \mathbb{H}^+$ the infimum in (4.1) is always a minimum. Moreover:*

(i) A particular minimizer is given by $\mathrm{d}G_t = g_t(x)\cdot \mathrm{d}\lambda$ and $U_t(x)$, where (g_t, U_t) is a Bures geodesic from $g_0(x) := \frac{\mathrm{d}G_0}{\mathrm{d}\lambda}(x)$ to $g_1(x) := \frac{\mathrm{d}G_1}{\mathrm{d}\lambda}(x)$ in λ-a.e. fiber $x \in \Omega$ and λ is any positive scalar measure dominating G_0, G_1.

(ii) Any minimizer is a d_H-Lipschitz curve $t \mapsto G_t$ such that $d_H(G_t, G_s) = |t-s| d_H(G_0, G_1)$ with potential $U \in L^\infty(0,1; L^2(\mathrm{d}G_t; \mathcal{H}))$ such that $\|U_t\|^2_{L^2(\mathrm{d}G_t;\mathcal{H})} = cst = d^2_H(G_0, G_1)$ for a.e. $t \in [0,1]$.

Proof: Fix G_0, G_1 once and for all.

(i) As in the previous proof it is easy to see that the particular choice (G_t, U_t) with $\mathrm{d}G_t = g_t(x)\cdot \mathrm{d}\lambda$ (and U_t optimal from g_0 to g_1 for a.e. x) connects G_0, G_1 and is therefore an admissible competitor in (4.1). Moreover (4.6) gives

$$
\begin{aligned}
d_H^2(G_0,G_1) = 4\int_\Omega d_B^2(g_0(x),g_1(x))\,\mathrm{d}\lambda(x) \\
&= \int_\Omega\left(\int_0^1 g_t(x)U_t(x):U_t(x)\,\mathrm{d}t\right)\mathrm{d}\lambda(x) \\
&= \int_0^1\int_\Omega \mathrm{d}\lambda(x)g_t(x)U_t(x):U_t(x)\,\mathrm{d}t = \int_0^1\int_\Omega \mathrm{d}G_tU_t:U_t\,\mathrm{d}t
\end{aligned}
$$

thus (G_t,U_t) is a minimizer.

(ii) Pick any minimizer (G_t,U_t). The constant-speed property easily follows from the fact that $t\mapsto \|U_t\|_{L^2(\mathrm{d}G_t)}$ is constant in time: For if not, an easy arc length reparametrization (Lemma A.1 in the appendix) would give an admissible curve with strictly lesser energy.

■

We can now establish a comparison between TV and d_H, which will be technically convenient and used repeatedly in the sequel.

Theorem 3. *For $G_0,G_1\in\mathbb{H}^+$ with masses m_0,m_1 there holds*

$$
\frac{1}{4\sqrt{d}}d_H^2(G_0,G_1)\le\|G_1-G_0\|_{TV}\le\sqrt{m_0+m_1}\,d_H(G_0,G_1), \tag{4.7}
$$

and d_H is topologically (sequentially) equivalent to the total variation distance on $\mathbb{H}^+$.

Note that this estimate has the correct scaling with respect to mass, i-e it is 1-homogeneous (since the Hellinger distance, roughly speaking, scales as the square-root of the mass).

Proof: Let us start with the lower bound. From [12, theorem 1] we have that

$$
d_B^2(g_0,g_1)\le\left|\sqrt{g_1}-\sqrt{g_0}\right|_2^2
$$

for any $g_0,g_1\in\mathcal{H}^+$, where $\sqrt{g}$ denotes the canonical square root of a PSD matrix g. Moreover, the Powers-Størmer inequality [62] reads exactly

$$
\left|\sqrt{g_1}-\sqrt{g_0}\right|_2^2\le|g_1-g_0|_1\,,
$$

where $|g|_1=\mathrm{tr}(\sqrt{gg^*})=\mathrm{tr}\,g$ is the Schatten (trace) 1-norm of g. Since we have $|g|_1\le\sqrt{d}|g|_2$ we get immediately from Lemma 4.3

$$\begin{aligned} d_H^2(G_0,G_1) = 4\int_\Omega d_B^2(G_0,G_1) &= 4\int_\Omega d_B^2(g_0(x),g_1(x))\mathrm{d}\lambda(x) \\ &\leq 4\int_\Omega \left|\sqrt{g_1}(x)-\sqrt{g_0}(x)\right|_2^2 \mathrm{d}\lambda(x) \leq 4\int_\Omega |g_1(x)-g_0(x)|_1\,\mathrm{d}\lambda(x) \\ &\leq 4\sqrt{d}\int_\Omega |g_1(x)-g_0(x)|_2\,\mathrm{d}\lambda(x) = 4\sqrt{d}\|G_1-G_0\|_{TV} \end{aligned}$$

as desired.
Let us now turn to the upper bound. Fix any positive scalar measure such that $G_0,G_1 \ll \lambda$. From Theorem 2 we know that there exists a minimizing curve $(G_t,U_t)_{t\in[0,1]}$ in (4.1) with $\mathrm{d}G_t = g_t(x)\cdot\mathrm{d}\lambda$ and (g_t,U_t) a Bures geodesic from $g_0 = \frac{\mathrm{d}G_0}{\mathrm{d}\lambda}$ to $g_1 = \frac{\mathrm{d}G_1}{\mathrm{d}\lambda}$ in a.e. every fiber. It follows from Lemma 3.2 that $\operatorname{tr} g_t \leq \max\{\operatorname{tr} g_0, \operatorname{tr} g_1\}$ for all $t\in[0,1]$ along this geodesic and for λ-a.e. x. Integrating and using the very rough bound $\max\{a,b\}\leq a+b$ gives here

$$m_t = \int_\Omega \operatorname{tr} g_t(x)\,\mathrm{d}\lambda(x) \leq \int_\Omega \{\operatorname{tr} g_0(x) + \operatorname{tr} g_1(x)\}\,\mathrm{d}\lambda(x) = m_0+m_1.$$

Our estimate (4.4), with here the mass control $m_t \leq M \leq m_0+m_1$ and the energy of the geodesic $E = \|U\|^2_{L^2(0,1;L^2(\mathrm{d}G_t))} = d_H^2(G_0,G_1)$, gives

$$\|G_1-G_0\|_{TV} \leq \sqrt{ME} \leq \sqrt{m_0+m_1}\,d_H(G_0,G_1)$$

as desired.
For the topological equivalence, note that the lower bound in (4.7) immediately shows that TV is stronger than d_H. Conversely, assume that $d_H(G^n,G)\to 0$. We first claim that the masses remain bounded, i-e $m^n = \operatorname{tr} G^n(\Omega) \leq M$ for some M. Since $d_H(G^n,G)\leq 1$ for n large enough and $|m^n-m| \leq \|G^n-G\|_{TV}$ (test $\psi = I$), (4.7) gives

$$|m^n-m| \leq \|G^n-G\|_{TV} \leq \sqrt{m^n+m}\,d_H(G^n,G) \leq \sqrt{m^n+m},$$

which then guarantees the boundedness of $\{m^n\}$ as claimed. One last use of (4.7) finally gives $\|G^n-G\|_{TV} \leq \sqrt{m^n+m}\,d_H(G^n,G) \leq \sqrt{M+m}\,d_H(G^n,G)\to 0$ and the proof is complete. ∎

Proposition 4.4 (Upper bound of the distance)**.** *For every pair $G_0,G_1\in\mathbb{H}^+$ with masses m_0,m_1 one has*

$$d_H^2(G_0,G_1) \leq 4(m_0+m_1). \tag{4.8}$$

Proof: Take any reference measure λ dominating G_0, G_1, and let $g_0(x), g_1(x)$ be the corresponding Radon-Nikodym densities. From [12, thm. 1] we have

$$d_B^2(g_0(x), g_1(x)) \le |\sqrt{g_1}(x) - \sqrt{g_0}(x)|_2^2 \le \operatorname{tr} g_0(x) + \operatorname{tr} g_1(x),$$

where the second inequality follows from simple linear algebra. Integrating w.r.t λ over Ω and appealing to Lemma 4.3 gives

$$\begin{aligned} d_H^2(G_0, G_1) = 4 \int_\Omega d_B^2(g_0(x), g_1(x))\, \mathrm{d}\lambda(x) \\ \le 4 \int_\Omega (\operatorname{tr} g_0(x) + \operatorname{tr} g_1(x))\, \mathrm{d}\lambda(x) = 4(m_0 + m_1) \end{aligned}$$

as desired. ■

Remark 4.5 (Optimality). The constant in (4.8) is optimal since for every $G \in \mathbb{H}^+$ with mass m we have

$$d_H^2(G, 0) = 4m. \tag{4.9}$$

The proof is immediate from Lemma 4.3 and (3.4) with $A_1 = 0$.

Lemma 4.6 (Characterization of AC^2 curves)**.** *A curve $G : [0,1] \to \mathbb{H}^+$ is AC^2 w.r.t. d_H if and only if there exists $U \in L^2(0,1; L^2(\mathrm{d}G_t; \mathcal{H}))$ such that the weak formulation (4.3) of $\partial_t G_t = (G_t U_t)^{Sym}$ holds, in which case U_t is uniquely defined for a.a. $t \in (0,1)$ as an element of $L^2(\mathrm{d}G_t)$ and the d_H metric speed is*

$$|\dot{G}_t|_H^2 = \int_\Omega \mathrm{d}G_t U_t : U_t \quad \text{for a.a.} \quad t \in (0,1).$$

Consequently, $G : [0,1] \to \mathbb{H}^+$ is Lipschitz if and only if the corresponding U belongs to $L^\infty(0,1; L^2(\mathrm{d}G_t; \mathcal{H}))$, and the corresponding Lipschitz constant coincides with $\|U\|_{L^\infty(0,1;L^2(\mathrm{d}G_t;\mathcal{H}))}$.

Proof: Notice that our claim about Lipschitz curves immediately follows from the stronger AC^2 part of our statement so we only need to establish the latter.

Let us start with the easiest implication, and assume that (G, U) satisfy $\partial_t G_t = (G_t U_t)^{Sym}$ with $U \in L^2(0,1; L^2(\mathrm{d}G_t))$. Fix any $[s,t] \subset [0,1]$. Changing variables $\tau = s + \theta(t-s)$ and rescaling time gives an admissible curve $(\widetilde{G}_\theta, \widetilde{U}_\theta)_{\theta \in [0,1]} = (G_\tau, (t-s)U_\tau)_{\tau \in [s,t]}$ connecting $\widetilde{G}_0 = G_s$ to $\widetilde{G}_1 = G_t$.

By Lemma A.1 we can reparametrize $(\widetilde{G}_\theta, \widetilde{U}_\theta)_{\theta\in[0,1]} \rightsquigarrow (\check{G}_\theta, \check{U}_\theta)_{\theta\in[0,1]}$ with constant speed $\|\check{U}_\theta\|_{L^2(\mathrm{d}\check{G}_\theta)} \equiv cst$. Taking into account the $(t-s)$ scaling in $\tau = s + \theta(t-s)$ gives exactly

$$d_H^2(G_s, G_t) \leq \int_0^1 \|\check{U}_\theta\|^2_{L^2(\mathrm{d}\check{G}_\theta)} \mathrm{d}\theta$$
$$= \left(\int_0^1 \|\widetilde{U}_\theta\|_{L^2(\mathrm{d}\widetilde{G}_\theta)} \mathrm{d}\theta\right)^2 = \left(\int_s^t \|U_\tau\|_{L^2(\mathrm{d}G_\tau)} \mathrm{d}\tau\right)^2$$

and therefore G is AC^2 relatively to the d_H distance.

Let us now turn to the converse implication, and let $(G_t)_{t\in[0,1]}$ be an arbitrary AC^2 curve. We first observe that

$$d_H^2(G_t, G_0) \leq \int_0^1 |\dot{G}_\tau|^2 \mathrm{d}\tau < +\infty$$

and (4.9) control

$$m_t := \operatorname{tr} G_t(\Omega) = \frac{1}{4} d_H^2(G_t, 0) \leq \frac{1}{4}\left(d_H(G_t, G_0) + d_H(G_0, 0)\right)^2 =: M.$$

Exploiting this mass bound in Theorem 3, (4.7) guarantees that $\|G_t - G_s\|_{TV} \leq \sqrt{2M} d_H(G_s, G_t)$. Hence $t \mapsto G_t$ is absolutely continuous w.r.t. the TV norm and $\partial_t G_t$ is a well-defined and finite ($\mathcal{H}$-valued) Radon measure for a.a. $t \in (0,1)$. For any fixed $\varphi \in C_b(\Omega; \mathcal{H})$ we set

$$\Phi(t) := \int_\Omega \mathrm{d}G_t : \varphi,$$

and observe that $\Phi \in AC([0,1])$ with

$$\Phi'(t) = \int_\Omega \mathrm{d}(\partial_t G_t) : \varphi \qquad \text{for a.a. } t \in (0,1). \tag{4.10}$$

Fix now any point of differentiability $t \in (0,1)$ of Φ (the set of such points has full measure, and is actually independent of φ). Let h be small enough, and pick from Theorem 2 a geodesic $(\tilde{G}_s, \tilde{U}_s)_{s\in[0,1]}$ from G_t to G_{t+h}, satisfying in particular $\partial_s \tilde{G}_s = (\tilde{G}_s \tilde{U}_s)^{Sym}$. Then

$$\left|\frac{\Phi(t+h) - \Phi(t)}{h}\right| = \frac{1}{h}\left|\int_\Omega \mathrm{d}G_{t+h} : \varphi - \int_\Omega \mathrm{d}G_t : \varphi\right|$$

$$
\begin{aligned}
&= \frac{1}{h}\left|\int_\Omega \mathrm{d}\tilde{G}_1 : \varphi - \int_\Omega \mathrm{d}\tilde{G}_0 : \varphi\right| \\
&= \frac{1}{h}\left|\int_0^1 \int_\Omega \mathrm{d}\tilde{G}_s \tilde{U}_s : \varphi \,\mathrm{d}s\right| \\
&\leq \frac{1}{h}\left(\int_0^1 \int_\Omega \mathrm{d}\tilde{G}_s \tilde{U}_s : \tilde{U}_s \,\mathrm{d}s\right)^{\frac{1}{2}} \left(\int_0^1 \int_\Omega \mathrm{d}\tilde{G}_s \varphi : \varphi \,\mathrm{d}s\right)^{\frac{1}{2}} \\
&= \underbrace{\frac{d_H(G_t, G_{t+h})}{h}}_{:=A_h} \underbrace{\left(\int_0^1 \int_\Omega \mathrm{d}\tilde{G}_s \varphi : \varphi \,\mathrm{d}s\right)^{\frac{1}{2}}}_{:=B_h} \qquad (4.11)
\end{aligned}
$$

By standard properties of the metric speed [5, thm. 1.1.2] the first term $A_h \to |\dot{G}_t|_H$ as $h \to 0$. In order to take the limit in the second B_h term, it is first easy to argue as before and conclude from (4.9) that the mass $\tilde{m}_s \leq \tilde{M}$ is bounded uniformly in $s \in (0,1)$ and $h \to 0$. Observing also that $d_H(\tilde{G}_0, \tilde{G}_s) \leq d_H(\tilde{G}_0, \tilde{G}_1) = d_H(G_t, G_{t+h}) \to 0$ as $h \to 0$ for all $s \in [0,1]$, (4.7) shows that

$$
\tilde{G}_s \xrightarrow[h\to 0]{TV} \tilde{G}_0 = G_t \qquad \text{for a.a. } s \in [0,1].
$$

As a consequence for fixed $\varphi \in C_b$ the inner integral in B_h converges pointwise in time

$$
\int_\Omega \mathrm{d}\tilde{G}_s \varphi : \varphi \xrightarrow[h\to 0]{} \int_\Omega \mathrm{d}G_t \varphi : \varphi \qquad \text{for a.a. } s \in [0,1].
$$

An easy application of Lebesgue's dominated convergence (with the previous mass bound $\left|\int \mathrm{d}\tilde{G}_s \varphi : \varphi\right| \leq \tilde{m}_s |\varphi|_\infty^2 \leq C$ uniformly in s) finally gives

$$
B_h = \left(\int_0^1 \int_\Omega \mathrm{d}\tilde{G}_s \varphi : \varphi \,\mathrm{d}s\right)^{\frac{1}{2}} \xrightarrow[h\to 0]{} \left(\int_0^1 \int_\Omega \mathrm{d}G_t \varphi : \varphi \,\mathrm{d}s\right)^{\frac{1}{2}} = \|\varphi\|_{L^2(\mathrm{d}G_t)}.
$$

Taking the limit in (4.11) with (4.10) we get

$$
\left|\int_\Omega \mathrm{d}(\partial_t G_t) : \varphi\right| \leq |\dot{G}_t| \cdot \|\varphi\|_{L^2(\mathrm{d}G_t)} \qquad \text{for all } \varphi \in C_b(\Omega; \mathcal{H}). \tag{4.12}
$$

hence by density the linear map $\varphi \mapsto \int_\Omega \mathrm{d}(\partial_t G_t) : \varphi$ is continuous for the $L^2(\mathrm{d}G_t)$ norm. The Riesz representation theorem therefore gives a unique element $U_t \in L^2(\mathrm{d}G_t)$ such that

$$\int_\Omega \mathrm{d}(\partial_t G_t) : V = \int_\Omega \mathrm{d}G_t U_t : V \qquad \text{for all } V \in L^2(\mathrm{d}G_t).$$

This means of course $\partial_t G_t = (G_t U_t)^{Sym}$ for a.a. $t \in (0,1)$, from which it is easy to check that the weak formulation (4.3) is satisfied. Moreover from (4.12) we see that

$$\|U_t\|_{L^2(\mathrm{d}G_t)} \leq |\dot{G}_t| \in L^2(0,1). \tag{4.13}$$

Fix now any Lebesgue point $t \in (0,1)$ for $\tau \mapsto \|U_\tau\|_{L^2(\mathrm{d}G_\tau)}$ and take h small. Since $(G_\tau, U_\tau)_{\tau\in[t,t+h]}$ is an admissible curve connecting G_t and G_{t+h} we get, by definition of d_H^2 and after a suitable scaling in time,

$$d_H^2(G_t, G_{t+h}) \leq h \int_t^{t+h} \int_\Omega \mathrm{d}G_\tau U_\tau : U_\tau \mathrm{d}\tau.$$

Dividing by h^2, taking the limit, and observing that $\frac{d_H(G_t,G_{t+h})}{h} \to |\dot{G}_t|$ in the left-hand side because $G \in AC^2$, we get from (4.13)

$$|\dot{G}_t|^2 \leq \lim_{h\to 0} \frac{1}{h} \int_t^{t+h} \int_\Omega \mathrm{d}G_\tau U_\tau : U_\tau \mathrm{d}\tau = \|U_t\|^2_{L^2(G_t)} \leq |\dot{G}_t|^2$$

and the proof is complete. ■

5. The Fisher-Rao distance

In this section we will mainly take interest in the space

$$\mathbb{P} = \{G \in \mathbb{H}^+, \quad \mathrm{tr}\, G(\Omega) = 1\}.$$

Recall that we wish to view $\mathbb{P}$ as a sphere of radius one in $\mathbb{H}^+$.

Remark 5.1. This is compatible with free probability theory (cf. [66]) in the following sense: Fix $P \in \mathbb{P}$, and consider the $*$-algebra of (generalized) random matrices $\mathcal{M} := L^{\infty-}(\mathrm{d}P; \mathbb{C}^{d\times d})$ with identity I. Defining the *trace* (in the sense of the free probability theory) by $\tau(\xi) := \int_\Omega \mathrm{d}P : \xi$, we get a non-commutative probability space $(\mathcal{M}, \tau)$. This space is faithful but in general not tracial. For $d = 1$ we recover the classical (commutative) probabilistic setting. Moreover, the Fisher-Rao distance that we are ready to introduce will coincide with the classical Fisher-Rao distance [7, 43] from information geometry for $d = 1$. In [13], a free-probabilistic analogue of the Wasserstein distance was introduced. It would be interesting to go beyond our matricial setting and try to define a free probabilistic counterpart of the Fisher-Rao distance, but this lies completely out of the scope of this article.

The Fisher-Rao distance on $\mathbb{P}$ can be defined as follows.

Definition 5.2 (Fisher-Rao distance). *Given $G_0, G_1 \in \mathbb{P}$ we set*

$$d^2_{FR}(G_0, G_1) := \inf_{\mathcal{A}_1(G_0,G_1)} \int_0^1 \left(\int_\Omega \mathrm{d}G_t U_t : U_t \right) \mathrm{d}t, \tag{5.1}$$

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

$$\begin{cases} G \in C([0,1]; \mathbb{P}_{TV}), \\ G|_{t=0} = G_0; \quad G|_{t=1} = G_1, \\ U \in L^2(0,1; L^2(\mathrm{d}G_t; \mathcal{H})) \end{cases} \quad \text{and } \partial_t G_t = (G_t U_t)^{Sym} \text{ in the weak sense.}$$

Comparing with Definition 4.1 it is clear that, at least formally, we view $\mathbb{P}$ as a submanifold of $\mathbb{H}^+$ with the induced Riemannian metric.

Theorem 4. *d_{FR} is a distance on $\mathbb{P}$.*

Proof: The identity of indiscernibles, symmetry, triangular inequality can be proved exactly as in Theorem 1 and we omit the details (the a priori bound on the masses $m_t = \operatorname{tr} G_t(\Omega) = 1$ even simplifies some parts). The only delicate point is to check that the infimum in (5.1) is finite, i-e that there is at least one admissible curve remaining in $\mathbb{P}$ while joining G_0, G_1 with finite cost. As already mentioned, we will later on view $(\mathbb{P}, d_{FR})$ as a unit sphere in the ambient cone-space $(\mathbb{H}^+, d_H)$, in which we already proved existence of geodesics. The natural thing to do is therefore to project down these conic Hellinger geodesics onto the sphere, i-e renormalize to unit masses.

To this end, take from Theorem 2 a geodesic $(G_t, U_t)_{t\in[0,1]}$ from G_0 to G_1 with $\partial_t G_t = (G_t U_t)^{Sym}$. By Corollary 5.6 we control the masses from below as

$$m_t = \operatorname{tr} G_t(\Omega) \geq \underline{m} = \frac{1}{2}.$$

(There is of course no circular argument in this anticipated use of Corollary 5.6, see Remark 5.7.) It is then a simple exercise to check that

$$(\tilde{G}_t, \tilde{U}_t) := \left(\frac{1}{m_t} G_t, U_t - \frac{\dot{m}_t}{m_t} I \right)$$

satisfies $\partial_t \tilde{G}_t = (\tilde{G}_t \tilde{U}_t)^{Sym}$. Moreover taking the trace in $\partial_t G_t = (G_t U_t)^{Sym}$ we know that $\dot{m}_t = \frac{d}{dt}\int_\Omega \mathrm{d}G_t : I = \int_\Omega \mathrm{d}G_t U_t : I$ and therefore

$$|\dot{m}_t|^2 = \left|\int_\Omega \mathrm{d}G_t U_t : I\right|^2 \leq \left(\int_\Omega \mathrm{d}G_t U_t : U_t\right)\left(\int_\Omega \mathrm{d}G_t I : I\right) = m_t \int_\Omega \mathrm{d}G_t U_t : U_t. \quad (5.2)$$

This allows to estimate the energy of the rescaled path as

$$\begin{aligned}
\int_0^1 \int_\Omega \mathrm{d}\tilde{G}_t \tilde{U}_t : \tilde{U}_t \,\mathrm{d}t &= \int_0^1 \int_\Omega \frac{1}{m_t} \mathrm{d}G_t \left(U_t - \frac{\dot{m}_t}{m_t} I\right) : \left(U_t - \frac{\dot{m}_t}{m_t} I\right) \mathrm{d}t \\
&\leq \frac{2}{\underline{m}}\left[\int_0^1 \int_\Omega \mathrm{d}G_t U_t : U_t \,\mathrm{d}t + \int_0^1 \int_\Omega \left|\frac{\dot{m}_t}{m_t}\right|^2 \mathrm{d}G_t I : I \,\mathrm{d}t\right] \\
&= \frac{2}{\underline{m}}\left[\int_0^1 \int_\Omega \mathrm{d}G_t U_t : U_t \,\mathrm{d}t + \int_0^1 \left|\frac{\dot{m}_t}{m_t}\right|^2 m_t \,\mathrm{d}t\right] \\
&\overset{(5.2)}{\leq} \frac{4}{\underline{m}} \int_0^1 \int_\Omega \mathrm{d}G_t U_t : U_t \,\mathrm{d}t < +\infty
\end{aligned}$$

and the proof is complete. ■

Remark 5.3. Lemma 4.6 holds in particular when the curve G takes values in $\mathbb{P} \subset \mathbb{H}^+$: As a consequence the characterization of AC^2 curves in $(\mathbb{P}, d_{FR})$ is exactly identical. (Of course this stems from the fact that the Riemannian metric on $\mathbb{P}$ is simply induced by the overlying metric on $\mathbb{H}^+$.)

5.1. Conic structure. We are going to show that the abstract metric cone over our Fisher-Rao space $(\mathbb{P}, d_{FR}/2)$ coincides with the Hellinger space $(\mathbb{H}^+, d_H/2)$. In other words, $\mathbb{P}$ is a unit sphere in the cone $\mathbb{H}^+$. Firstly, for any element $G \in \mathbb{H}^+$, we set

$$r = r(G) := \sqrt{m} = \sqrt{\operatorname{tr} G(\Omega)}.$$

Then we can identify G with a pair $[G/r^2, r] \in \mathfrak{C}(\mathbb{P})$, where the first factor is again normalized to unity, $\int_\Omega \operatorname{tr} G/r^2 = 1$.

Theorem 5 (Conic structure). *The space $(\mathbb{H}^+, d_H/2)$ is a metric cone over $(\mathbb{P}, d_{FR}/2)$, where $\mathbb{H}^+$ is identified with $\mathfrak{C}(\mathbb{P})$ via $G \simeq [G/r^2, r]$.*

Note that this covers one-point spaces $\Omega = \{x\}$ and implies Proposition 3.7, up to a minor dimensional scaling issue (the details of which are left to the reader).

Proof: *Step 1.* We first observe that it suffices to establish the weaker claim that $(\mathbb{H}^+, d_H/2)$ is a metric cone over *some* metric space (which, due to the identification above, is nothing but $\mathbb{P}$ equipped with *some* distance d). Indeed, by Proposition 4.4, for any two matrix measures $G_0, G_1 \in \mathbb{P}$ one has

$$\frac{1}{2} d_H(G_0, G_1) \leq \sqrt{2}. \tag{5.3}$$

If $(\mathbb{H}^+, d_H/2)$ is to be a cone over $(\mathbb{P}, d)$ as claimed, (5.3) and (3.1) imply that $\cos(d(G_0, G_1)) \geq 0$, whence the diameter of $(\mathbb{P}, d)$ would be controlled from above by $\pi/2 < \pi$. By Theorem 2 $(\mathbb{H}^+, d_H)$ is a geodesic space, hence from [16, Corollary 5.11] $(\mathbb{P}, d)$ would also necessarily be a geodesic space. Evoking Lemma 3.1 and Definition 5.2, we see that d should actually coincide with $d_{FR}/2$ as claimed (note that the infimum of the lengths coincides with the infimum of the AC^2-energies due to Lemma A.1).

Step 2. In view of (5.3) and [46, Theorem 2.2], in order to prove the weaker claim in step 1 it suffices to establish the following scaling property that fully characterizes cones:

$$d_H^2(r_0^2 G_0, r_1^2 G_1) = r_0 r_1 d_H^2(G_0, G_1) + 4(r_1 - r_0)^2, \tag{5.4}$$

for all $G_0, G_1 \in \mathbb{P}$ and $r_0, r_1 \geq 0$. In the case $r_0 r_1 = 0$ the claim is immediate by Remark 4.5 and our choice of $r^2 = \operatorname{tr} G(\Omega) = m$. We can thus assume that $r_0 r_1 > 0$. Consider the monotone increasing function

$$a(t) := \frac{r_1 t}{(1-t) r_0 + t r_1},$$

and observe that

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

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

Let (G_t, U_t) be any admissible path joining $G_0, G_1 \in \mathbb{P}$. Then the rescaled path $(\tilde{G}_t, \tilde{U}_t)$ defined as

$$\tilde{G}_t := [(1-t) r_0 + t r_1]^2 G_{a(t)}, \qquad \tilde{U}_t := a'(t) U_{a(t)} + 2 \frac{(r_1 - r_0)}{(1-t) r_0 + t r_1} I \tag{5.5}$$

connects $r_0^2G_0$ and $r_1^2G_1$ in $\mathbb{H}^+$. A straightforward computation shows that $(\tilde G_t, \tilde U_t)$ satisfies $\partial_t \tilde G_t = (\tilde G_t \tilde U_t)^{Sym}$. Testing (4.3) with $\Psi = \Phi_a = (r_0 + (r_1 - r_0)t(a))I$ and $a \in [0,1]$, we infer

$$\begin{aligned}(r_0+(r_1-r_0)t(1))\int_\Omega \mathrm{d}G_1 : I - (r_0+(r_1-r_0)t(0))\int_\Omega \mathrm{d}G_0 : I \\ -(r_1-r_0)\int_0^1 t'(a)\int_\Omega \mathrm{d}G_a : I\,\mathrm{d}a \\ = \int_0^1 (r_0+(r_1-r_0)t(a))\int_\Omega \mathrm{d}G_a : U_a\,\mathrm{d}a. \quad (5.6)\end{aligned}$$

Let us compute the energy of the path $\tilde G_t$, employing (5.6):

$$\begin{aligned}&\int_0^1 \left(\int_\Omega \mathrm{d}\tilde G_t \tilde U_t : \tilde U_t\right)\mathrm{d}t \\ &= r_0 r_1 \int_0^1 a'(t)\left(\int_\Omega \mathrm{d}G_{a(t)}U_{a(t)} : U_{a(t)}\right)\mathrm{d}t \\ &+4(r_1-r_0)\int_0^1 a'(t)(r_0+(r_1-r_0)t)\int_\Omega \mathrm{d}G_{a(t)} : U_{a(t)}\,\mathrm{d}t \\ &+4(r_1-r_0)^2\int_0^1\int_\Omega \mathrm{d}G_{a(t)} : I\,\mathrm{d}t \\ &= r_0r_1\int_0^1\left(\int_\Omega \mathrm{d}G_aU_a : U_a\right)\mathrm{d}a \\ &+4(r_1-r_0)\int_0^1 (r_0+(r_1-r_0)t(a))\int_\Omega \mathrm{d}G_a : U_a\,\mathrm{d}a \\ &+4(r_1-r_0)^2\int_0^1 t'(a)\int_\Omega \mathrm{d}G_a : I\,\mathrm{d}a \\ &= r_0r_1\int_0^1\left(\int_\Omega \mathrm{d}G_tU_t : U_t\right)\mathrm{d}t \\ &+4(r_1-r_0)(r_0+(r_1-r_0)t(1))\int_\Omega \mathrm{d}G_1 : I \\ &-4(r_1-r_0)(r_0+(r_1-r_0)t(0))\int_\Omega \mathrm{d}G_0 : I\end{aligned}$$

$$= r_0 r_1 \int_0^1 \left(\int_\Omega \mathrm{d}G_t U_t : U_t \right) \mathrm{d}t + 4(r_1 - r_0)^2.$$

Consequently, $d_H^2(r_0^2 G_0, r_1^2 G_1) \leq r_0 r_1 d_H^2(G_0, G_1) + 4(r_1 - r_0)^2$.

The opposite inequality is proved in a similar fashion: Taking any path $(\tilde{G}, \tilde{U})$ connecting $r_0^2 G_0$ to $r_1^2 G_1$ in $\mathbb{H}^+$ and undoing the change of variables (5.5) gives an admissible path (G, U) connecting G_0, G_1 in $\mathbb{P}$, whose cost can be computed explicitly as above. ■

Note that we also proved along the way

Corollary 5.4. *The space $(\mathbb{P}, d_{FR})$ has diameter $\leq \pi$.*

Another useful consequence is

Corollary 5.5 (Existence of Fisher-Rao geodesics)**.** *$(\mathbb{P}, d_{FR})$ is a geodesic space, i-e for all $G_0, G_1 \in \mathbb{P}$ the infimum in (5.1) is always a minimum. Moreover any minimizer is a d_{FR}-Lipschitz curve such that $d_{FR}(G_t, G_s) = |t - s| d_{FR}(G_0, G_1)$ with potential $U \in L^2(0, 1; L^2(\mathrm{d}G_t; \mathcal{H}))$ such that $\|U_t\|^2_{L^2(\mathrm{d}G_t;\mathcal{H})} = cst = d_{FR}^2(G_0, G_1)$ for a.e. $t \in [0, 1]$.*

Proof: We have already proved in Theorem 2 that $(\mathbb{H}^+, d_H)$ is a geodesic space. Owing to the cone structure (Theorem 5), $(\mathbb{P}, d_{FR})$ is automatically a geodesic space (see the discussion in step 1 of the proof of Theorem 5). The rest follows by Lemma 4.6. ■

Corollary 5.6. *Let G_t be a d_H-geodesic in $\mathbb{H}^+$ joining G_0, $G_1 \in \mathbb{P}$. Then the corresponding masses m_t are bounded from below:*

$$m_t \geq 1 - 2t(1 - t). \tag{5.7}$$

The proof is immediate by Lemma 3.2, the bound (5.3), and the fact that $m_0 = m_1 = 1$.

Remark 5.7. There was no circular reasoning in applying Corollary 5.6 in the proof of Theorem 4, since (5.7) merely relies on (5.3) and step 2 of the proof of Theorem 5, both of which have nothing to do with Theorem 4.

5.2. Topological properties. The very particular conic structure automatically entails nice topological properties:

Corollary 5.8. *The distance d_{FR} is topologically equivalent to the TV distance on $\mathbb{P}$.*

Proof: It is immediate from Lemma 3.3 that the spherical and the conic distances d_{FR}, d_H are topologically equivalent on the sphere, but by Theorem 3 the latter one is topologically equivalent to the total variation distance. ■

Proposition 5.9. *The metric space* $(\mathbb{P}, d_{FR})$ *is complete.*

Proof: Take any d_{FR}-Cauchy sequence G^n. Since G^n has unit mass $m^n = \operatorname{tr} G^n(\Omega) = 1$ for all n, (4.7) shows that this sequence is also Cauchy for the TV distance. Since $\mathbb{P}$ is complete for the total variation we see that the G^n converges in TV, and therefore in d_{FR} too owing to Corollary 5.8. ■

Corollary 5.10. *The metric space* $(\mathbb{H}^+, d_H)$ *is complete.*

Proof: By [16, Proposition 5.9] a metric cone $(\mathfrak{C}(X), d_{\mathfrak{C}(X)})$ is complete if and only if (X, d_X) is complete, hence the result immediately follows from the cone structure (Theorem 5) and Proposition 5.9. ■

A last result will turn out to be useful later on for technical purposes:

Lemma 5.11 (Lower-semicontinuity)**.** *The distance d_{FR} is sequentially lower semicontinuous with respect to the weak-∗ topology on* $\mathbb{P}$.

Proof: Consider any two converging sequences of measures from $\mathbb{P}$,

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

and assume that $\liminf d_{FR}(G_0^k, G_1^k) < +\infty$ (otherwise there is nothing to prove). Up to extraction of a subsequence if needed we can moreover take $\lim d_{FR}(G_0^k, G_1^k) = \liminf d_{FR}(G_0^k, G_1^k) < +\infty$. For each k, the endpoints G_0^k and G_1^k can be joined by a geodesic $(G_t^k, U_t^k)_{t\in[0,1]}$, whose energies are therefore bounded as

$$E[G^k; U^k] = d_{FR}^2(G_0^k, G_1^k) \leq E$$

uniformly in $k \in \mathbb{N}$.

By the fundamental estimate (4.4) with $m_t^k = M = 1$ on $\mathbb{P}$ we get

$$\forall\, t, s \in [0,1], \forall k \in \mathbb{N}: \qquad \|G_s^k - G_t^k\|_{TV} \leq C|t-s|^{1/2}.$$

By the (classical) Banach-Alaoglu theorem, $\mathbb{P} \subset (C_0(\Omega; \mathcal{H}))^*$ is moreover weakly-∗ sequentially relatively compact. The previous $\frac{1}{2}$-Hölder bound and the sequential lower semicontinuity of d_{TV} with respect to the weak-∗

convergence allow us to apply a refined version of the Arzelà-Ascoli theorem (Lemma A.3 in the Appendix) to conclude that there exists a TV-continuous curve $(G_t)_{t\in[0,1]}$ connecting G_0 and G_1 such that

$$\forall t \in [0,1]: \qquad G_t^k \to G_t \quad \text{weakly-}* \tag{5.8}$$

along some subsequence $k \to \infty$ (not relabeled here) and with $\|G_s - G_t\| \le C|t-s|^{\frac{1}{2}}$. Let $\mu^k = \mathrm{d}t \otimes \mathrm{d}G_t^k$ be the matricial measure on $Q := [0,1]\times\Omega$ defined by disintegration as

$$\forall\, \phi \in C_c(Q;\mathcal{H}): \qquad \int_Q \phi(t,x) : \mathrm{d}\mu^k(t,x) := \int_0^1 \left(\int_\Omega \phi(t,.) : \mathrm{d}G_t^k \right) \mathrm{d}t.$$

Leveraging the pointwise convergence (5.8) and the uniform bounds on the mass $m_t^k = \operatorname{tr} G_t^k(\Omega) = 1$, a simple application of Lebesgue's dominated convergence guarantees that

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

where the finite measure $\mu^0 = \mathrm{d}t\otimes\mathrm{d}G_t \in \mathbb{P}(Q)$ is defined by duality in terms of the weak-$*$ limit $G_t = \lim G_t^k$ (as was μ^k in terms of G_t^k), and, moreover,

$$\mu^k \llcorner [t_0,t_1]\times\Omega \to \mu^0 \llcorner [t_0,t_1]\times\Omega \quad \text{weakly-}* \qquad \forall\, t_0,t_1 \in [0,1].$$

Let

$$X \subset L^\infty(Q;\mathcal{H})$$

be the linear span of the functions of the form

$$\Psi(t,x) = \Phi(t,x)\mathbf{1}_{[t_0,t_1]\times\Omega}(t,x), \quad \Phi \in C_c^{1,0}(Q;\mathcal{H}), \quad t_0,t_1 \in \mathbb{Q}\cap[0,1].$$

We are going to apply a refined Banach-Alaoglu theorem (Lemma A.2 in the Appendix) on the space X equipped with the norm $\|\cdot\| := \|\cdot\|_{L^\infty(Q)}$. To this end, it is easy to see that $(X,\|\cdot\|)$ is separable. Consider the following norms on X

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

and the linear forms

$$\varphi_k(\Phi) = \int_Q \mathrm{d}\mu^k U^k : \Phi, \qquad k = 1,2,\dots.$$

The weak-* convergence of μ^k and the Cauchy-Schwarz inequality imply that the hypotheses of Lemma A.2 are met with

$$c_k := \|\varphi_k\|_{(X,\|.\|_k)^*} \le \sqrt{E[G^k;U^k]} = d_{FR}(G_0^k, G_1^k).$$

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

$$\forall \Phi \in C_c^1(Q;\mathcal{H}): \qquad \int_0^1 \left(\int_\Omega \mathrm{d}G_t^k U_t^k : \Phi_t\right)\mathrm{d}t \xrightarrow[k\to\infty]{} \varphi_0(\Phi)$$

with moreover

$$\|\varphi_0\|_{(X,\|\cdot\|_0)^*} \le \liminf_{k\to\infty} d_{FR}(G_0^k, G_1^k). \tag{5.9}$$

Let $N_0 \subset X$ be the kernel of the seminorm $\|\cdot\|_0$. By the Riesz representation theorem, the dual $(X, \|\cdot\|_0)^* = (X/N_0, \|\cdot\|_0)^*$ can be isometrically identified with the completion $\overline{X/N_0}$ of X/N_0 with respect to $\|\cdot\|_0$. One can check that this completion is exactly $L^2(0,1;L^2(\mathrm{d}G_t;\mathcal{H}))$.

Consequently, there exists $U \in L^2(0,1;L^2(\mathrm{d}G_t;\mathcal{H}))$ such that

$$\varphi_0(\Phi) = \int_Q \mathrm{d}\mu^0 U : \Phi = \int_0^1 \left(\int_\Omega \mathrm{d}G_t U_t : \Phi_t\right)\mathrm{d}t$$

and

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

Moreover, (G,U) is an admissible curve joining G_0, G_1. Indeed, the established convergences are enough to pass to the limit in the constraint (4.3) inside time intervals $[s,t]$ with $s,t \in \mathbb{Q} \cap [0,1]$ and $\Psi \in C_c^1(Q;\mathcal{H})$. Since G_t is a TV-continuous matrix function, an easy approximation argument shows that (4.3) actually holds for any $s,t \in [0,1]$ and $\Psi \in C_b^1(Q;\mathcal{H})$.

Recalling (5.9), it remains to take into account that

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

■

6. The spherical heat flow and Schrödinger problem

6.1. Otto calculus. At least formally, it is clear from the above construction that one can view $\mathbb{P}$ as a real Riemannian manifold such that d_{FR} becomes the Riemannian distance. This is very similar to the celebrated Otto calculus [59, 68, 69], in particular the tangent space at a point $G \in \mathbb{P}$ is

$$T_G\mathbb{P} := \left\{ \Xi = (GU)^{Sym} - G\int_\Omega \mathrm{d}G : U \middle| \qquad U \in L^2(\mathrm{d}G, \mathcal{H}) \right\}$$

and the (squared) tangent norm reads

$$\|\Xi\|^2_{T_G\mathbb{P}} := \int_\Omega \mathrm{d}GU : U - \left(\int_\Omega \mathrm{d}G : U\right)^2. \tag{6.1}$$

The gradients of functionals $f : \mathbb{P} \to \mathbb{R}$ are

$$\nabla_{FR} f(G) = \left[G\frac{\delta f}{\delta G}\right]^{Sym} - G\int_\Omega \mathrm{d}G : \frac{\delta f}{\delta G}, \tag{6.2}$$

where $\frac{\delta f}{\delta G}$ stands for the usual first variation. We omit the details and refer to [15, Appendix C] for similar considerations.

6.2. The heat flow. As motivated in Section 3, cf. (3.12), the relevant entropy here is not the classical von Neumann entropy $S(G) = \mathrm{tr}\int_\Omega \frac{\mathrm{d}G}{\mathrm{d}\Lambda}\log\frac{\mathrm{d}G}{\mathrm{d}\Lambda}\mathrm{d}\Lambda$, but rather

$$\begin{aligned} E(G) &= \int_\Omega \left(\mathrm{tr}\,\frac{\mathrm{d}G}{\mathrm{d}\Lambda} - \mathrm{tr}\, I - \log\det\frac{\mathrm{d}G}{\mathrm{d}\Lambda}\right)\mathrm{d}\Lambda \\ &= \int_\Omega -\log\det\left(\frac{\mathrm{d}G}{\mathrm{d}\Lambda}\right)\mathrm{d}\Lambda = -\mathrm{tr}\int_\Omega \log\left(\frac{\mathrm{d}G}{\mathrm{d}\Lambda}\right)\mathrm{d}\Lambda. \end{aligned} \tag{6.3}$$

By construction, E is a non-negative functional that achieves its minimum and vanishes at $G = \Lambda I$. It is very similar to the classical Itakura-Saito divergence [39, 57, 41] from signal processing.

Here we slightly abused the notations and implicitly assumed that $G \in \mathbb{P}_\Lambda$ was absolutely continuous w.r.t Λ for these formulas to make sense. However, E can be extended from $\mathbb{P}_\Lambda$ to the whole space $\mathbb{P}$ in the following natural way. Indeed, let

$$\mathrm{d}G = \mathrm{d}G^\Lambda + \mathrm{d}G^\perp = \frac{\mathrm{d}G^\Lambda}{\mathrm{d}\Lambda}\mathrm{d}\Lambda + \mathrm{d}G^\perp \tag{6.4}$$

be the (Radon-Nikodym)-Lebesgue decomposition of G w.r.t. Λ. Since the function $f(A) = -\log\det(A)$ is sublinear at infinity, its recession function $f^{\infty}(A) := \lim\limits_{t\to+\infty} \frac{f(tA)}{t}$ is identically zero. Employing the standard definition for a convex functional of measures, cf. [37, 31], we can legitimately set

$$E(G) := \int_\Omega -\log\det\left(\frac{dG^\Lambda}{d\Lambda}\right) d\Lambda \qquad \text{for all } G \in \mathbb{P}, \tag{6.5}$$

where the integral may be infinite. Note that this definition automatically makes E lower semicontinuous w.r.t. the weak-$*$ convergence of measures, [4, thm. 2.34].

As motivated in Section 3, the analogue of the heat flow in our Fisher-Rao space can be defined as the gradient flow $\partial_t G = -\nabla_{FR} E(G)$. In order to compute explicitly the latter FR gradient, note first that the Lebesgue decomposition $G \mapsto G^\Lambda$ is linear. Accordingly, and since the first variation of $A \in \mathcal{H}^+ \mapsto \log\det A$ is A^{-1}, it is not difficult to check that the L^2-variation of $G \mapsto \int_\Omega -\log\det\left(\frac{dG^\Lambda}{d\Lambda}\right) d\Lambda$ is the $\mathcal{H}$-valued function $x \mapsto -\left(\frac{dG^\Lambda}{d\Lambda}\right)^{-1}(x)$. Since $\int_\Omega d\Lambda I : I = 1$ and $\frac{dG^\Lambda}{d\Lambda}(x) = 0$ for $dG^\perp$-a.e. x, we compute explicitly

$$\begin{aligned}
\nabla_{FR} E(G) &= -\left[G\left(\frac{dG^\Lambda}{d\Lambda}\right)^{-1}\right]^{Sym} + G\int_\Omega dG : \left(\frac{dG^\Lambda}{d\Lambda}\right)^{-1} \\
&= -\left[\left(\Lambda\frac{dG^\Lambda}{d\Lambda} + G^\perp\right)\times\left(\frac{dG^\Lambda}{d\Lambda}\right)^{-1}\right]^{Sym} + G\int_\Omega \left(d\Lambda\frac{dG^\Lambda}{d\Lambda} + dG^\perp\right) : \left(\frac{dG^\Lambda}{d\Lambda}\right)^{-1} \\
&= -[\Lambda I + 0]^{Sym} + G\left[\int_\Omega d\Lambda\frac{dG^\Lambda}{d\Lambda} : \left(\frac{dG^\Lambda}{d\Lambda}\right)^{-1} + 0\right] = -\Lambda I + G.
\end{aligned}$$

Thus the heat flow is

$$\partial_t G = \Lambda I - G \tag{6.6}$$

Remark 6.1. We point out that, although the entropy E itself only sees the absolute continuous part $\frac{dG^\Lambda}{d\Lambda}$ of G, its gradient really does depend on the singular part as well through the full G term in (6.6). This might sound surprising at first, but can be explained recalling that $G \in \mathbb{P}$ must satisfy the mass constraint $\operatorname{tr} G(\Omega) = 1$. One therefore cannot perturb the singular part without (in general) perturbing the absolutely continuous part as well in order to comply with the mass constraint, and it becomes clear that the gradient should depend on both the absolutely continuous and singular parts of G. (This would be different if we worked in the conic Hellinger space, where no correction term is needed to enforce mass conservation. In that case the Hellinger gradient would only depend on G^Λ, not on $G^\perp$.)

As usual, the corresponding Fisher information is then defined as the production of the entropy E along its own (negative) gradient flow (6.6). To make this more explicit, assume for simplicity that $G = G^\Lambda$ is absolutely continuous. Then by definition we set

$$\begin{aligned} F(G) &:= -\frac{d}{dt}E(G) = \frac{d}{dt}\int_\Omega \log\det\left(\frac{\mathrm{d}G}{\mathrm{d}\Lambda}\right)\mathrm{d}\Lambda \\ &= \int_\Omega \mathrm{d}\Lambda\left(I - \frac{\mathrm{d}G}{\mathrm{d}\Lambda}\right) : \left(\frac{\mathrm{d}G}{\mathrm{d}\Lambda}\right)^{-1} = \operatorname{tr}\int_\Omega \mathrm{d}\Lambda\left[\left(\frac{\mathrm{d}G}{\mathrm{d}\Lambda}\right)^{-1} - I\right]. \end{aligned} \tag{6.7}$$

With (6.7) at hand, observe that the function $A \mapsto A^{-1} - I$ is sublinear and therefore its recession function vanishes. Just like we did for the entropy in (6.5), we can lawfully extend the definition of the Fisher information to all $G \in \mathbb{P}$ by setting

$$F(G) := \operatorname{tr}\int_\Omega \mathrm{d}\Lambda\left[\left(\frac{\mathrm{d}G^\Lambda}{\mathrm{d}\Lambda}\right)^{-1} - I\right], \tag{6.8}$$

where the integral may be infinite. This is of course consistent with what one would obtain by differentiating in time $E(G) = E(G^\Lambda)$ along solutions of (6.6), even if G was not absolutely continuous. This definition is also consistent with the usual representation

$$F(G) = \|\nabla_{FR}E(G)\|^2_{T_G\mathbb{P}}$$

in terms of squared Riemannian norms. Indeed, the above computation shows that the tangent vector $\nabla_{FR}E(G)$ is represented in $T_G\mathbb{P}$ by the matrix function $\left(\frac{\mathrm{d}G^\Lambda}{\mathrm{d}\Lambda}\right)^{-1}(x)$, and plugging $U = \left(\frac{\mathrm{d}G^\Lambda}{\mathrm{d}\Lambda}\right)^{-1}$ into (6.1) gives exactly (6.8).

Denote the semigroup generated by the gradient flow (6.6) by

$$\mathsf{S}_t(G) := \Lambda I + e^{-t}(G - \Lambda I). \tag{6.9}$$

Its absolutely-continuous part is

$$[\mathsf{S}_t(G)]^\Lambda = \Lambda I + e^{-t}(G^\Lambda - \Lambda I) = \mathsf{S}_t(G^\Lambda), \tag{6.10}$$

and the singular part is

$$[\mathsf{S}_t(G)]^\perp = e^{-t}G^\perp. \tag{6.11}$$

The main regularization property of the heat flow that will serve our purpose is the time decay of the driving entropy, expressed here in the slightly stronger sense (in every fiber):

Lemma 6.2 (Exponential decay)**.** *For any G in $\mathbb{P}$ and $t \geq s \geq 0$, we have*

$$-\log\det\left(\frac{\mathrm{d}\mathsf{S}_t(G^\Lambda)}{\mathrm{d}\Lambda}(x)\right) \leq e^{-(t-s)}\left\{-\log\det\left(\frac{\mathrm{d}\mathsf{S}_s(G^\Lambda)}{\mathrm{d}\Lambda}(x)\right)\right\}$$

for Λ-a.e. $x \in \Omega$. Consequently,

$$E(\mathsf{S}_t(G)) \leq e^{-(t-s)}E(\mathsf{S}_s(G)).$$

Remark 6.3. For abstract gradient flows in Riemannian manifolds $\dot{x} = -\nabla\Phi(x)$ the exponential decay $\Phi(x_t) \leq e^{-2\lambda(t-s)}\Phi(x_s)$ is often related to the λ-convexity of Φ along Riemannian geodesics.
Lemma 6.2 thus suggests that our entropy functional might be $\frac{1}{2}$-geodesically convex w.r.t. to our Fisher-Rao distance d_{FR} on $\mathbb{P}$. This will be proved completely rigorously later on, see Theorem 8.

Proof: Writing for simplicity $g = \frac{\mathrm{d}G^\Lambda}{\mathrm{d}\Lambda}$ and employing the concavity of the function $\log\det$, we deduce

$$\begin{aligned}
\log\det\left(\frac{\mathrm{d}\mathsf{S}_t(G^\Lambda)}{\mathrm{d}\Lambda}\right) &= \log\det\Big(I + e^{-t}(g-I)\Big) \\
&= \log\det\Big((1-e^{s-t})I + e^{s-t}(I + e^{-s}(g-I))\Big) \\
&\geq e^{s-t}\log\det(I + e^{-s}(g-I)) = e^{s-t}\log\det\left(\frac{\mathrm{d}\mathsf{S}_s(G^\Lambda)}{\mathrm{d}\Lambda}\right).
\end{aligned}$$

The second part of the statement immediately follows by integrating w.r.t Λ over Ω. ■

6.3. The Schrödinger problem. With the Fisher information now properly defined, the Schrödinger problem on $(\mathbb{P}, d_{FR})$ rather classically reads, cf. [47, 24, 18, 23, 70, 71]:

Definition 6.4 (Schrödinger problem). *Given a fixed $\epsilon > 0$ and $G_0, G_1 \in \mathbb{P}$ we define*

$$\mathfrak{S}_\epsilon(G_0, G_1) := \inf_G \left\{ \frac{1}{2}\int_0^1 |\dot{G}_t|^2_{FR} \mathrm{d}t + \frac{\varepsilon^2}{2}\int_0^1 F(G_t)\,\mathrm{d}t, \right.$$
$$\left. \text{s.t. } t \mapsto G_t \in \mathbb{P} \text{ has endpoints } G_0, G_1 \right\}. \tag{6.12}$$

The metric speed in the first term is computed relatively to the distance d_{FR} on $\mathbb{P}$, and $F(G_t)$ is the Fisher information just defined in (6.8). Owing to our characterization of AC^2 curves (Lemma 4.6) and our explicit formula for the Fisher information we see that (6.13) also reads

$$\mathfrak{S}_\epsilon(G_0, G_1) = \inf_{\mathcal{A}_1(G_0,G_1)} \left\{ \frac{1}{2}\int_0^1 \left(\int_\Omega \mathrm{d}G_t U_t : U_t \right) \mathrm{d}t \right.$$
$$\left. + \frac{\epsilon^2}{2} \operatorname{tr} \int_0^1 \left(\int_\Omega \mathrm{d}\Lambda \left[\left(\frac{\mathrm{d}G_t^\Lambda}{\mathrm{d}\Lambda} \right)^{-1} - I \right] \right) \mathrm{d}t \right\}, \tag{6.13}$$

where the admissible pairs $(G_t, U_t)_{t\in[0,1]} \in \mathcal{A}_1(G_0, G_1)$ are as in Definition 5.2.

We will prove shortly that this problem has a unique minimizer, which we call the *ε-geodesic* or *Schrödinger bridge* between G_0, G_1. In order to study this dynamical Schrödinger problem we shall need the following technical result:

Proposition 6.5. *The kinetic action functional*

$$G \mapsto \mathcal{K}(G) := \frac{1}{2}\int_0^1 |\dot{G}_t|^2_{FR} \mathrm{d}t$$

is convex for the linear interpolation on $C([0,1]; \mathbb{P}_{FR})$, and l.s.c. for the pointwise weak-$*$ convergence (i-e $G_t^n \overset{*}{\rightharpoonup} G_t$ for all $t \in [0,1]$).

Proof: We claim that we have the dual formulation

$$\mathcal{K}(G) = \sup_{\varphi \in C_b^{1,0}([0,1]\times\Omega;\mathcal{H})} \left\{ \int_\Omega \mathrm{d}G_1 : \varphi_1 - \int_\Omega \mathrm{d}G_0 : \varphi_0 \right.$$

$$-\frac{1}{2}\int_0^1\int_\Omega \mathrm{d}G_t\varphi_t:\varphi_t\,\mathrm{d}t-\int_0^1\int_\Omega \mathrm{d}G_t:\partial_t\varphi_t\,\mathrm{d}t\Bigg\} \quad (6.14)$$

where we denoted $\varphi_t(x)=\varphi(t,x)$ with an obvious abuse of notations. To see this, observe from our characterization of AC^2 curves (Lemma 4.6 in $\mathbb{P}$ rather than in $\mathbb{H}^+$, see Remark 5.3) that

$$\frac{1}{2}\int_0^1|\dot{G}_t|^2_{FR}\mathrm{d}t=\int_0^1\int_\Omega \mathrm{d}G_tU_t:U_t\,\mathrm{d}t=\int_0^1\int_\Omega(g_tU_t:U_t)\,\mathrm{d}\gamma_t\mathrm{d}t,$$

where we denoted as before $g_t(x)=\frac{\mathrm{d}G_t}{\mathrm{d}\gamma_t}(x)$ with $\gamma_t=\operatorname{tr}G_t$. The potential $U_t(x)$ is such that

$$\partial_tG_t=(G_tU_t)^{Sym},$$

and we recall from Lemma 4.2 that $t\mapsto G_t$ is AC^2 in time with values in TV. In particular the ODE holds for a.e. t in the Banach space $\mathbb{H}_{TV}$, and since $(G_tU_t)^{Sym}\ll\gamma_t$ we see that $\partial_tG_t\ll\gamma_t$ as well. We denote

$$\xi_t(x):=\frac{\mathrm{d}(\partial_tG_t)}{\mathrm{d}\gamma_t}\quad\text{and}\quad (g_t(x)U_t(x))^{Sym}=\frac{\mathrm{d}\left((G_tU_t)^{Sym}\right)}{\mathrm{d}\gamma_t}$$

the corresponding Radon-Nikodym densities, and we have thus the constraint

$$\xi_t(x)=(g_t(x)U_t(x))^{Sym}\qquad\text{for }\mathrm{d}t\otimes\mathrm{d}\gamma_t\text{-a.e. }(t,x).$$

Now it is a simple (pointwise and finite-dimensional) exercise to check that, given $g\in\mathcal{P}$ (with unit trace) and $\xi,U\in\mathcal{H}$ with $\xi=(gU)^{Sym}=\frac{1}{2}(gU+Ug)$, we have the characterization

$$\frac{1}{2}gU:U=\max_{\phi\in\mathcal{H}}\left\{-\frac{1}{2}g\phi:\phi+\xi:\phi\right\}. \quad (6.15)$$

The maximizer is of course $\phi=U$. For any fixed test-function $\varphi=\varphi_t(x)\in C^{1,0}_b$ this gives

$$\begin{aligned}
\frac{1}{2}\int_0^1|\dot{G}_t|^2\mathrm{d}t&=\int_0^1\int_\Omega(g_tU_t:U_t\mathrm{d}t)\mathrm{d}\lambda_t\\
&=\int_0^1\int_\Omega\left(\max_{\phi\in\mathcal{H}}\left\{-\frac{1}{2}g_t(x)\phi:\phi+\xi_t(x):\phi\right\}\right)\mathrm{d}\lambda_t(x)\mathrm{d}t\\
&\geq\int_0^1\int_\Omega\left\{-\frac{1}{2}g_t(x)\varphi_t(x):\varphi_t(x)+\xi_t(x):\varphi_t(x)\right\}\mathrm{d}\lambda_t(x)\mathrm{d}t
\end{aligned}$$

$$= -\frac{1}{2}\int_0^1 \int_\Omega \mathrm{d}G_t\varphi_t : \varphi_t \,\mathrm{d}t + \int_0^1 \int_\Omega \mathrm{d}(\partial_t G_t) : \varphi_t \,\mathrm{d}t$$
$$= \int_\Omega \mathrm{d}G_1 : \varphi_1 - \int_\Omega \mathrm{d}G_0 : \varphi_0$$
$$-\frac{1}{2}\int_0^1 \int_\Omega \mathrm{d}G_t\varphi_t : \varphi_t \mathrm{d}t - \int_0^1 \int_\Omega \mathrm{d}G_t : \partial_t\varphi_t \mathrm{d}t$$

In order to check that there is no duality gap in the above inequality, we simply approximate the optimizer in (6.15): More precisely, a density argument [36, Thm. 2.11] allows to pick a sequence $\varphi^n \in C_b^{1,0}$ such that $\varphi^n \to U$ in $L^2(0,1;L^2(\mathrm{d}G_t;\mathcal{H}))$ and $\varphi^n \to U$ in $L^1(0,1;L^1(\mathrm{d}(\partial_t G_t;\mathcal{H})))$, hence

$$-\frac{1}{2}\int_0^1 \int_\Omega \mathrm{d}G_t\varphi_t^n : \varphi_t^n \,\mathrm{d}t + \int_0^1 \int_\Omega \mathrm{d}(\partial_t G_t) : \varphi_t^n \,\mathrm{d}t$$
$$\xrightarrow[n\to\infty]{} -\frac{1}{2}\int_0^1 \int_\Omega \mathrm{d}G_t U_t : U_t \,\mathrm{d}t + \int_0^1 \int_\Omega \mathrm{d}(\partial_t G_t) : U_t \,\mathrm{d}t$$

as needed and (6.14) follows.

Returning now to our main statement, we see from the dual representation (6.15) that the kinetic action can be written as a supremum of convex (linear) functionals, hence the convexity. Likewise, observe that, for fixed $\varphi \in C_b^{1,0}$, the functional

$$G \mapsto \int_\Omega \mathrm{d}G_1 : \varphi_1 - \int_\Omega \mathrm{d}G_0 : \varphi_0 - \frac{1}{2}\int_0^1 \int_\Omega \mathrm{d}G_t\varphi_t : \varphi_t \mathrm{d}t$$
$$- \int_0^1 \int_\Omega \mathrm{d}G_t : \partial_t\varphi_t \mathrm{d}t$$

is continuous for the pointwise weak-* convergence (the first two boundary terms are immediate, and for the time-space integrals one can simply apply Lebesgue's dominated convergence with $m_t = \operatorname{tr} G_t(\Omega) \equiv 1$). As a consequence the action functional $\mathcal{K}$ is l.s.c. as a supremum of continuous functionals. ∎

Theorem 6. *For fixed $G_0, G_1 \in \mathbb{P}$ with $E(G_0), E(G_1) < +\infty$ the infimum in* (6.12)(6.13) *is always a minimum.*

Moreover this minimum is attained for a unique curve $G^\epsilon \in AC^2([0,1];\mathbb{P}_{FR})$ and a unique potential $U^\epsilon \in L^2(0,1;L^2(\mathrm{d}G_t;\mathcal{H}))$.

For the usual Schrödinger problem it is known that $E(G_0), E(G_1) < +\infty$ is also a necessary condition for the well-posedness of (6.13) but for the sake of simplicity we omit the details, see e.g. [56].

Proof: Pick from Corollary 5.5 a Fisher-Rao geodesic $(G_t)_{t\in[0,1]}$ between G_0, G_1 (for $\varepsilon = 0$). Slightly anticipating on independent results from the next section, let $G^\varepsilon = (G^\varepsilon_t)_{t\in[0,1]}$ be the curve obtained by perturbing G as in Lemma 7.6 (here $\varepsilon > 0$ is the fixed temperature parameter in the Schrödinger functional). Then (7.28) guarantees that G^ε has finite ε-cost, and therefore the infimum in (6.13) is finite.

Choose now any minimizing sequence $\{G^n\}_n$, and let us implement the direct method in the Calculus of Variations. From (6.12) we see that the kinetic action $\mathcal{K}(G^n) = \frac{1}{2}\int_0^1 |\dot{G}^n_t|^2_{FR}\mathrm{d}t$ is bounded, thus $\{G^n\}_n$ is uniformly AC^2 and therefore equicontinuous w.r.t. the metric space $(\mathbb{P}, d_{FR})$. From a variant of the Arzelà-Ascoli theorem (Lemma A.3 in the Appendix) and the weak-$*$ lower semicontinuity from Lemma 5.11 we see that there is a d_{FR}-continuous curve such that (up to extraction of a subsequence if needed)

$$G^n_t \overset{*}{\rightharpoonup} G_t \quad \text{for all} \quad t \in [0,1].$$

The lower semicontinuity in Proposition 6.5 immediately gives

$$\int_0^1 |\dot{G}_t|^2 \mathrm{d}t \leq \liminf \int_0^1 |\dot{G}^n_t|^2 \mathrm{d}t.$$

As for the Fisher information $\mathcal{F}(G) = \int_0^1 F(G_t)\mathrm{d}t$, note that $g \mapsto Tr\, g^{-1}$ is (linearly) strictly convex and l.s.c. on $\mathcal{H}^+$. As a consequence we infer from [4, thm. 2.34] that F is l.s.c. for the weak-$*$ convergence on $\mathbb{H}^+$. By Fatou's lemma we see that

$$\int_0^1 F(G^n_t)\,\mathrm{d}t \leq \int_0^1 \liminf F(G^n_t)\,\mathrm{d}t \leq \liminf \int_0^1 F(G^n_t)\,\mathrm{d}t$$

and therefore $G = \lim G^n$ is a minimizer.

Finally, the uniqueness follows from the convexity of the AC^2 energy in Proposition 6.5 combined with the strict convexity of the Fisher information, and the proof is complete. ■

7. Γ-convergence and geodesic convexity

In this section we aim at proving the Γ-convergence of the ε-Schrödinger problem (6.4) towards the geodesic problem (5.1) as $\varepsilon \to 0$, for fixed endpoints $G_0, G_1 \in \mathbb{P}$. As it is often the case, the $\Gamma - \liminf$ part will not be too difficult and will rely on some suitable lower semicontinuity. The construction of recovery sequences in the $\Gamma - \limsup$ will be technically more involved, and requires the construction of a suitable ε-perturbation $(\tilde{G}_t)_{t\in[0,1]}$ of any fixed AC^2 curve $(G_t)_{t\in[0,1]}$. Our construction will involve the heat flow (6.6) as a quantitative regularizing tool, and a careful examination of the defect of optimality (at order one in ε) will yield as a byproduct the $\frac{1}{2}$-geodesic convexity of the entropy. This strategy was already exploited in [8], and will be extended to abstract metric spaces in our subsequent work [56].

We will first need a significant number of technical preliminaries. For brevity, and given the Lebesgue decomposition $G = G^\Lambda + G^\perp$ of an arbitrary $G \in \mathbb{P}$ with respect to the fixed reference measure Λ, we will denote below

$$g(x) = \frac{\mathrm{d}G^\Lambda}{\mathrm{d}\Lambda}(x).$$

We use similar lowercase notations $\tilde{g}, g_0$ for the Radon-Nikodym densities (w.r.t. Λ) of other corresponding objects $\tilde{G}, G_0$, etc.

In the next several lemmas, $(G_t)_{t\in[0,1]}$ is a given curve in

$$AC^2(0,1;\mathbb{P}_{FR}),$$

and we consider a fixed Lipschitz-continuous, non-negative function $h(t)$ with

$$h(t) > 0 \quad \text{for all} \quad t \in (0,1).$$

This will allow to control a change of time scale $s = h(t)$, and h will be carefully chosen later on. Then we define the perturbed curve

$$\tilde{G}_t := \mathsf{S}_{h(t)} G_t \qquad \text{for any fixed } t \in [0,1] \tag{7.1}$$

constructed by solving the heat-flow for a time $s = h(t)$ starting from G_t.

Lemma 7.1. *For $\tilde{G}$ defined by* (7.1) *we have $F(\tilde{G}) \in L^\infty_{loc}(0,1)$.*

Proof: For every $t \in (0,1)$ we can estimate

$$F(\tilde{G}_t) = \operatorname{tr}\int_\Omega \mathrm{d}\Lambda\left[\left(I + e^{-h(t)}(g_t - I)\right)^{-1} - I\right]$$
$$\leq \operatorname{tr}\int_\Omega \mathrm{d}\Lambda\left[\left((1 - e^{-h(t)})I)\right)^{-1} - I\right] = \frac{1}{1 - e^{-h(t)}} - 1,$$

and we conclude observing that with our assumptions $h(t) > 0$ is locally bounded away from zero. ■

Lemma 7.2. *We have $\tilde{G} \in AC^2_{loc}(0,1;\mathbb{P}_{FR})$ and, moreover, $\tilde{G} \in C([0,1];\mathbb{P})$.*

Proof: Differentiating (7.1) in time at a.a. $t \in (0,1)$ with (6.9) gives

$$\partial_t\tilde{G}_t = e^{-h(t)}\partial_t G_t - h'(t)e^{-h(t)}(G_t - \Lambda I),$$

whence

$$\partial_t\tilde{G} - h'(t)\Lambda e^{-h(t)}(I - g_t) + h'(t)e^{-h(t)}G_t^\perp = e^{-h(t)}\partial_t G_t. \tag{7.2}$$

By Lemma 4.6 (characterization of AC^2 curves), there exists $U_t \in L^2(0,1;L^2(\mathrm{d}G_t;\mathcal{H}))$ such that

$$\partial_t G_t = (G_t U_t)^{Sym} = \Lambda\,(g_t U_t)^{Sym} + (G_t^\perp U_t)^{Sym} \tag{7.3}$$

in the weak sense. Consequently, by linearity of the Lebesgue decomposition,

$$\partial_t G_t^\Lambda = \Lambda\,(g_t U_t)^{Sym}, \tag{7.4}$$

$$\partial_t G_t^\perp = (G_t^\perp U_t)^{Sym}. \tag{7.5}$$

Moreover, $|\dot{G}_t|^2 = \int_\Omega \mathrm{d}G_t U_t : U_t = \int_\Omega \mathrm{d}\Lambda g_t U_t : U_t + \int_\Omega \mathrm{d}G_t^\perp U_t : U_t$. Now we identify in (7.2)

$$\Lambda e^{-h(t)}(I - g_t) = \Lambda\left[(I + e^{-h(t)}(g_t - I))\tilde{V}_t\right]^{Sym}, \tag{7.6}$$

where

$$\tilde{V}_t := \left(I + e^{-h(t)}(g_t - I)\right)^{-1} - I.$$

Indeed,

$$\left[\left(I + e^{-h(t)}(g_t - I)\right)\tilde{V}_t\right]^{Sym} = I - I - e^{-h(t)}(g_t - I) = e^{-h(t)}(I - g_t).$$

Furthermore, for Λ-a.a. x and a.a. t,

$$(I + e^{-h(t)}(g_t(x) - I))\,\Xi_1 : \Xi_2$$

defines a real scalar product on the space of matrices $\Xi \in \mathbb{C}^{d\times d}$. Let Π denote the corresponding orthogonal projection onto the subspace of Hermitian matrices (we omit the indexes and simply write $\Pi = \Pi_{t,x}$ for simplicity). Then (7.4) implies

$$\partial_t G_t^\Lambda = \Lambda\Big[(I + e^{-h(t)}(g_t - I))\tilde{W}_t\Big]^{Sym}, \tag{7.7}$$

where

$$\tilde{W}_t := \Pi\Big(\big(I + e^{-h(t)}(g_t - I)\big)^{-1} g_t U_t\Big).$$

Now set

$$\tilde{U}_t := \big(1 - \mathbf{1}_{\operatorname{supp} G_t^\perp}\big)\big(h'(t)\tilde{V}_t + e^{-h(t)}\tilde{W}_t\big) + \mathbf{1}_{\operatorname{supp} G_t^\perp}(-h'(t)I + U_t). \tag{7.8}$$

From (7.2) we see that

$$\partial_t \tilde{G}_t^\Lambda - \Lambda\Big[(I + e^{-h(t)}(g_t - I))(h'(t)\tilde{V}_t)\Big]^{Sym} = \Lambda\Big[(I + e^{-h(t)}(g_t - I))(e^{-h(t)}\tilde{W}_t)\Big]^{Sym}. \tag{7.9}$$

From (7.8) and (7.9) we deduce that

$$\partial_t \tilde{G}_t^\Lambda = \Lambda\Big[(I + e^{-h(t)}(g_t - I))\tilde{U}_t\Big]^{Sym} \tag{7.10}$$

For the singular part, (7.2), (6.11) and (7.5) yield

$$\partial_t \tilde{G}_t^\perp + h'(t)\tilde{G}_t^\perp = (\tilde{G}_t^\perp U_t)^{Sym}, \tag{7.11}$$

and using (7.8) we conclude that

$$\partial_t \tilde{G}_t^\perp = (\tilde{G}_t^\perp \tilde{U}_t)^{Sym}. \tag{7.12}$$

Now (7.8) gives

$$\begin{aligned}
\frac{1}{2}\int_\Omega \mathrm{d}\tilde{G}_t \tilde{U}_t : \tilde{U}_t
&= \frac{1}{2}\int_\Omega \mathrm{d}\Lambda (I + e^{-h(t)}(g_t - I))\tilde{U}_t : \tilde{U}_t + \frac{1}{2}\int_\Omega \mathrm{d}\tilde{G}_t^\perp \tilde{U}_t : \tilde{U}_t \\
&\le |h'(t)|^2 \int_\Omega \mathrm{d}\Lambda (I + e^{-h(t)}(g_t - I))\tilde{V}_t : \tilde{V}_t \\
&\quad + e^{-2h(t)} \int_\Omega \mathrm{d}\Lambda (I + e^{-h(t)}(g_t - I))\tilde{W}_t : \tilde{W}_t
\end{aligned}$$

$$+|h'(t)|^2\int_\Omega \mathrm{d}\tilde{G}_t^\perp I:I+\int_\Omega \mathrm{d}\tilde{G}_t^\perp U_t:U_t \quad (7.13)$$

for a.a. t. An explicit computation shows moreover that

$$\int_\Omega \mathrm{d}\Lambda(I+e^{-h(t)}(g_t-I))\tilde{V}_t:\tilde{V}_t+\int_\Omega \mathrm{d}\tilde{G}_t^\perp I:I$$
$$=\int_\Omega \mathrm{d}\Lambda\Big(I-(I+e^{-h(t)}(g_t-I))\Big):\Big(\big(I+e^{-h(t)}(g_t-I)\big)^{-1}-I\Big)+\mathrm{tr}\int_\Omega \mathrm{d}\tilde{G}_t^\perp$$
$$=\int_\Omega \mathrm{d}\Lambda\,\mathrm{tr}\Big(\tilde{g}_t^{-1}+\tilde{g}_t-2I\Big)+\mathrm{tr}\int_\Omega \mathrm{d}\tilde{G}_t^\perp$$
$$=\int_\Omega \mathrm{d}\Lambda\,\mathrm{tr}\Big(\tilde{g}_t^{-1}-I\Big)=F(\tilde{G}_t) \quad (7.14)$$

since $\Lambda\tilde{g}_t+\tilde{G}_t^\perp=\tilde{G}_t$ has unit mass for a.a. t. Hence, the sum of the first and the third terms in the right-hand side of (7.13) is $L^\infty_{loc}(0,1)$ by Lemma 7.1.

The second and the fourth term in the right-hand side of (7.13) can can be estimated in the following way:

$$e^{-2h(t)}\int_\Omega \mathrm{d}\Lambda(I+e^{-h(t)}(g_t-I))\tilde{W}_t:\tilde{W}_t+\int_\Omega \mathrm{d}\tilde{G}_t^\perp U:U$$
$$=e^{-2h(t)}\int_\Omega \mathrm{d}\Lambda(I+e^{-h(t)}(g_t-I))\Big[\Pi\Big(\big(I+e^{-h(t)}(g_t-I)\big)^{-1}g_tU_t\Big)\Big]$$
$$:\Big[\Pi\Big(\big(I+e^{-h(t)}(g_t-I)\big)^{-1}g_tU_t\Big)\Big]$$
$$+e^{-h(t)}\int_\Omega \mathrm{d}G_t^\perp U_t:U_t$$
$$\le e^{-2h(t)}\int_\Omega \mathrm{d}\Lambda(I+e^{-h(t)}(g_t-I))\Big[\big(I+e^{-h(t)}(g_t-I)\big)^{-1}g_tU_t\Big]$$
$$:\Big[\big(I+e^{-h(t)}(g_t-I)\big)^{-1}g_tU_t\Big]$$
$$+e^{-h(t)}\int_\Omega \mathrm{d}G_t^\perp U_t:U_t$$
$$=e^{-2h(t)}\int_\Omega \mathrm{d}\Lambda g_t(I+e^{-h(t)}(g_t-I))^{-1}g_tU_t:U_t+e^{-h(t)}\int_\Omega \mathrm{d}G_t^\perp U_t:U_t$$

$$\leq e^{-h(t)}\int_\Omega \mathrm{d}\Lambda g_t U_t : U_t + e^{-h(t)}\int_\Omega \mathrm{d}G_t^\perp U_t : U_t$$
$$= e^{-h(t)}\int_\Omega \mathrm{d}G_t U_t : U_t = e^{-h(t)}|\dot{G}_t|^2. \quad (7.15)$$

In the last inequality we have employed the purely algebraic inequality

$$A(I+\theta(A-I))^{-1}A \leq \theta^{-1}A \quad (7.16)$$

for any positive-semidefinite matrix A and $\theta \in (0,1)$. Since G is an AC^2 curve we have $|\dot{G}|^2 \in L^2(0,1)$ in (7.15), the above calculations show that the right-hand side in (7.13) is L^2_{loc} in time, hence

$$\tilde{U}_t \in L^2_{loc}(0,1;L^2(\mathrm{d}\tilde{G}_t;\mathcal{H})).$$

Given that by construction $\partial_t\tilde{G}_t = (\tilde{G}_t\tilde{U}_t)^{Sym}$ we conclude from Lemma 4.6 that $\tilde{G} \in AC^2_{loc}(0,1;\mathbb{P}_{FR})$ and $\left|\dot{\tilde{G}}\right|^2 = \int_\Omega \mathrm{d}\tilde{G}\tilde{U} : \tilde{U}$ for a.a. $t \in (0,1)$.

In view of Corollary 5.8 and in order to finally show the strong continuity of $\tilde{G}$ at the boundary (for definiteness, at $t=0$), we estimate

$$\begin{aligned}
\|\tilde{G}_t - \tilde{G}_0\|_{TV} &= \int_\Omega \mathrm{d}\Lambda|\tilde{g}_t - \tilde{g}_0|_2 + \|e^{-h(t)}G_t^\perp - e^{-h(0)}G_0^\perp\|_{TV}\\
&= \int_\Omega \mathrm{d}\Lambda\left|\left(I + e^{-h(t)}(g_t - I)\right) - \left(I + e^{-h(0)}(g_0 - I)\right)\right|_2\\
&\quad + \left\|e^{-h(t)}(G_t^\perp - G_0^\perp) + (e^{-h(t)} - e^{-h(0)})G_0^\perp\right\|_{TV}\\
&\leq |e^{-h(t)} - e^{-h(0)}|\int_\Omega \mathrm{d}\Lambda\,|g_0 - I|_2 + e^{-h(t)}\int_\Omega \mathrm{d}\Lambda|g_t - g_0|_2\\
&\quad + e^{-h(t)}\|G_t^\perp - G_0^\perp\|_{TV} + |e^{-h(t)} - e^{-h(0)}| \times \|G_0^\perp\|_{TV} \to 0
\end{aligned}$$

as $t \searrow 0$ and the proof is achieved. ■

Lemma 7.3 (Chain rule). *For a.a.t $\in (0,1)$ we have* $\operatorname{tr}\tilde{U}_t \in L^1(\Omega;\mathrm{d}\Lambda)$, *where $\tilde{U}_t$ is defined by* (7.8) *and represents the dynamics of $\tilde{G}$ through $\partial_t\tilde{G}_t = (\tilde{G}_t\tilde{U}_t)^{Sym}$. Moreover $E(\tilde{G}_t) \in AC^2_{loc}(0,1)$, and*

$$\frac{d}{dt}\left(E(\tilde{G}_t)\right) = -\int_\Omega \operatorname{tr}\tilde{U}_t\mathrm{d}\Lambda \quad (7.17)$$

a.e. in $(0,1)$.

Proof: Recalling that $\Lambda I \in \mathbb{P}$ has unit mass and $\operatorname{tr}\tilde{U}$ is real, we observe that the Cauchy-Schwarz inequality gives for a.e. $t \in (0,1)$

$$
\begin{aligned}
\left(\int_\Omega \mathrm{d}\Lambda \operatorname{tr}\tilde{U}_t\right)^2 &= \left(\int_\Omega \mathrm{d}\Lambda \tilde{g}_t \tilde{U}_t : \tilde{g}_t^{-1}\right)^2 \\
&\le \left(\int_\Omega \mathrm{d}\Lambda \tilde{g}_t \tilde{U}_t : \tilde{U}_t\right)\left(\int_\Omega \mathrm{d}\Lambda \tilde{g}_t \tilde{g}_t^{-1} : \tilde{g}_t^{-1}\right) \\
&= \left(\int_\Omega \mathrm{d}\tilde{G}_t \tilde{U}_t : \tilde{U}_t\right)\left(\int_\Omega \mathrm{d}\Lambda \operatorname{tr}\left(\left[I + e^{-h(t)}(g_t - I)\right]^{-1}\right)\right) \\
&\le \left(\int_\Omega \mathrm{d}\tilde{G}_t \tilde{U}_t : \tilde{U}_t\right)\left(\int_\Omega \mathrm{d}\Lambda \operatorname{tr}\left(\left[(1 - e^{-h(t)})I)\right]^{-1}\right)\right) \\
&= \frac{1}{1 - e^{-h(t)}} \int_\Omega \mathrm{d}\tilde{G}_t \tilde{U}_t : \tilde{U}_t \quad (7.18)
\end{aligned}
$$

Hence, $\operatorname{tr}\tilde{U}_t \in L^1(\Omega; \mathrm{d}\Lambda)$ for a.a. t.

It is well-known that the first variation of $g \mapsto \int_\Omega \mathrm{d}\Lambda \log\det g$ is $x \mapsto g^{-1}(x)$, as soon as g is invertible Λ-a.e. in Ω. Since $\partial_t \tilde{G}_t^\Lambda = \left(\tilde{G}_t^\Lambda \tilde{U}_t\right)^{Sym}$ in the weak sense, see (7.10), for every smooth test function $\psi : (0,1) \to \mathbb{R}$ we have, in the sense of distributions $\mathcal{D}'(0,1)$,

$$
\begin{aligned}
\left\langle \frac{d}{dt} E(\tilde{G}), \psi \right\rangle &= \left\langle \int_\Omega \mathrm{d}\Lambda \log\det\left(I + e^{-h(t)}(g - I)\right), \psi' \right\rangle \\
&= -\left\langle \int_\Omega \left(\mathrm{d}\tilde{G}^\Lambda \tilde{U}\right)^{Sym} : (I + e^{-h(t)}(g - I))^{-1}, \psi \right\rangle \\
&= -\left\langle \int_\Omega \mathrm{d}\Lambda \left((I + e^{-h(t)}(g - I))\tilde{U}\right) : (I + e^{-h(t)}(g - I))^{-1}, \psi \right\rangle \\
&= -\left\langle \int_\Omega \mathrm{d}\Lambda \operatorname{tr}\tilde{U}, \psi \right\rangle.
\end{aligned}
$$

This shows that (7.17) holds in the sense of distributions, but since the right-hand side is locally L^2 in time (due to (7.18) with $h(t) > 0$ locally bounded away from zero), it also holds almost everywhere in $(0,1)$. ■

The main ingredient in our construction of recovery sequences later on will be

Lemma 7.4. *For a.a. $t \in (0,1)$ we have*

$$\frac{1}{2}\left|\dot{\tilde{G}}_t\right|^2 + \frac{1}{2}|h'(t)|^2 F(\tilde{G}_t) + h'(t)\frac{d}{dt}\left(E(\tilde{G}_t)\right) \leq \frac{1}{2}e^{-h(t)}\left|\dot{G}_t\right|^2 \leq \frac{1}{2}\left|\dot{G}_t\right|^2. \quad (7.19)$$

Let us slightly anticipate at this stage that our construction of recovery sequences will consist in keeping the second Fisher information term and throwing away part of the third term on the left, while the proof of the geodesic convexity of the entropy will on the contrary be based on discarding $F \geq 0$ and integrating by parts $h'\frac{dE}{dt}$.

Proof: We use the same framework as in the proof of Lemma 7.2 and keep the the same notations (for $\tilde{U}, \tilde{V}, \tilde{W}$). From (7.8) we compute

$$\begin{aligned}
&\frac{1}{2}\int_\Omega \mathrm{d}\Lambda\left[I + e^{-h(t)}(g_t - I)\right]\tilde{U}_t : \tilde{U}_t \\
&\qquad + \frac{1}{2}|h'(t)|^2\int_\Omega \mathrm{d}\Lambda\left[I + e^{-h(t)}(g_t - I)\right]\tilde{V}_t : \tilde{V}_t \\
&\qquad - h'(t)\int_\Omega \mathrm{d}\Lambda\left[I + e^{-h(t)}(g_t - I)\right]\tilde{U}_t : \tilde{V}_t \\
&\qquad\qquad = \frac{1}{2}e^{-2h(t)}\int_\Omega \mathrm{d}\Lambda\left[I + e^{-h(t)}(g_t - I)\right]\tilde{W}_t : \tilde{W}_t, \quad (7.20)
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{2}\int_\Omega \mathrm{d}\tilde{G}_t^\perp \tilde{U}_t : \tilde{U}_t + \frac{1}{2}|h'(t)|^2\int_\Omega \mathrm{d}\tilde{G}_t^\perp I : I \\
&\qquad + h'(t)\int_\Omega \mathrm{d}\tilde{G}_t^\perp \tilde{U}_t : I \\
&\qquad\qquad = \frac{1}{2}\int_\Omega \mathrm{d}\tilde{G}_t^\perp U_t : U_t. \quad (7.21)
\end{aligned}$$

The sum of the left-hand sides of (7.20) and (7.21) is equal to the left-hand side of (7.19). Indeed, the sum of the integrals in the first terms of (7.20) and (7.21) is exactly the squared metric derivative $\left|\dot{\tilde{G}}_t\right|^2 = \int_\Omega \mathrm{d}\tilde{G}_t \tilde{U}_t : \tilde{U}_t$ of $\tilde{G}$. The sum of the second terms of (7.20) and (7.21) matches the second term of (7.19) due to (7.14). Moreover,

$$-\int_\Omega \mathrm{d}\Lambda\left[I + e^{-h(t)}(g_t - I)\right]\tilde{U}_t : \tilde{V}_t + \int_\Omega \mathrm{d}\tilde{G}_t^\perp \tilde{U}_t : I$$

$$= -\int_\Omega \mathrm{d}\Lambda \Big[I + e^{-h(t)}(g_t - I)\Big]\tilde{U}_t :$$
$$\Big(\big(I + e^{-h(t)}(g_t - I)\big)^{-1} - I\Big) + \operatorname{tr}\int_\Omega \mathrm{d}\tilde{G}_t^\perp \tilde{U}_t$$
$$= -\int \operatorname{tr}\tilde{U}_t \mathrm{d}\Lambda + \frac{d}{dt}\int \operatorname{tr}\tilde{G}_t^\Lambda + \frac{d}{dt}\int \operatorname{tr}\tilde{G}_t^\perp = \frac{d}{dt}\Big(E(\tilde{G}_t)\Big).$$

We have used Lemma 7.3, formulas (7.10) and (7.12), and the fact that the mass of $\tilde{G}_t \in \mathbb{P}$ is conserved. We conclude by estimating the sum of the right-hand sides of (7.20) and (7.21) exactly as as in (7.15). ■

The next result will essentially allow to integrate (7.19) in time all the way to $t = 0, 1$, which will be crucial later on but is a priori not legitimate so far since Lemma 7.2 and Lemma 7.3 only give regularity *locally* in $(0, 1)$ at this stage.

Lemma 7.5. *Assume that* $h(0) = h(1) = 0$ *and that* $h'(t)$ *is bounded away from* 0 *near* $t = 0$ *and* $t = 1$. *Then* $t \mapsto E(\tilde{G}_t)$ *is continuous at the endpoints* 0 *and* 1.

Proof: Since $t = 0$ and $t = 1$ are completely symmetric we only prove the statement at 0. Observe from Lemma 7.2 that $\tilde{G}_t \to \tilde{G}_0 = G_0$ in d_{FR} as $t \to 0^+$, or equivalently in total variation. Since TV is stronger than the weak-$*$ convergence and due to the lower semi-continuity of the entropy E for the weak-$*$ convergence [4, thm. 2.34], we conclude immediately that $t \mapsto E(\tilde{G}_t)$ is lower semicontinuous at $t = 0^+$ and therefore it suffices to prove the upper semicontinuity. We only consider the case when $E(G_0) = E(\tilde{G}_0)$ is finite, otherwise there is nothing to prove.

Step 1: Assume first that $\det g_0(x) \geq C > 0$ is bounded away from 0. The curve G is AC^2 with values in $\mathbb{P}$, thus by (4.7) and Lemma 3.3 it is also AC^2 with values in the Banach space $\mathbb{H}_{TV}$. By Morrey's embedding [6] we see that $G : [0, 1] \to \mathbb{H}_{TV}$ is $\frac{1}{2}$-Hölder continuous. Thus,

$$\int_\Omega \mathrm{d}\Lambda |g_0(x) - g_t(x)|_2 = \|G_0^\Lambda - G_t^\Lambda\|_{TV} \leq \|G_0 - G_t\|_{TV} \leq Ct^{1/2}.$$

Consequently $\int_{[|g_0 - g_t|_2 > t^{1/4}]} \mathrm{d}\Lambda \leq Ct^{1/4}$, whence

$$\int_{[g_0 - g_t > \frac{t^{1/4}}{\sqrt{d}} I]} \mathrm{d}\Lambda \leq Ct^{1/4}. \tag{7.22}$$

Then

$$\begin{aligned}
E(\tilde{G}_t) &= -\int_\Omega \mathrm{d}\Lambda \log\det\left[I + e^{-h(t)}(g_t - I)\right] \\
&= -\int_{[g_0 - g_t > \frac{t^{1/4}}{\sqrt{d}} I]} \mathrm{d}\Lambda \log\det\left[I + e^{-h(t)}(g_t - I)\right] \\
&\quad - \int_{[g_0 - g_t \le \frac{t^{1/4}}{\sqrt{d}} I]} \mathrm{d}\Lambda \log\det\left[I + e^{-h(t)}(g_t - I)\right] \\
&\le -\int_{[g_0 - g_t > \frac{t^{1/4}}{\sqrt{d}} I]} \mathrm{d}\Lambda \log\det\left[(1 - e^{-h(t)})I\right] \\
&\quad - \int_{[g_0 - g_t \le \frac{t^{1/4}}{\sqrt{d}} I]} \mathrm{d}\Lambda \log\det\left(e^{-h(t)} g_t\right) \\
&\le -d \log\left(1 - e^{-h(t)}\right) \int_{[g_0 - g_t > \frac{t^{1/4}}{\sqrt{d}} I]} \mathrm{d}\Lambda \\
&\quad - \int_{[g_0 - g_t \le \frac{t^{1/4}}{\sqrt{d}} I]} \mathrm{d}\Lambda \log\det\left[e^{-h(t)}\left(g_0 - \frac{t^{1/4}}{\sqrt{d}} I\right)\right] \\
&\le -Ct^{1/4} \log\left(1 - e^{-h(t)}\right) + \left\{ \int_{[g_0 - g_t > \frac{t^{1/4}}{\sqrt{d}} I]} \mathrm{d}\Lambda \log\det\left[e^{-h(t)}\left(g_0 - \frac{t^{1/4}}{\sqrt{d}} I\right)\right] \right. \\
&\quad \left. - \int_\Omega \mathrm{d}\Lambda \log\det\left[e^{-h(t)}\left(g_0 - \frac{t^{1/4}}{\sqrt{d}} I\right)\right] \right\} \\
&\le -Ct^{1/4} \log\left(1 - e^{-h(t)}\right) - \int_\Omega \mathrm{d}\Lambda \log\det\left[e^{-h(t)}\left(g_0 - \frac{t^{1/4}}{\sqrt{d}} I\right)\right] \\
&\quad + \int_{[g_0 - g_t > \frac{t^{1/4}}{\sqrt{d}} I]} \mathrm{d}\Lambda \log\det g_0. \qquad (7.23)
\end{aligned}$$

The last inequality simply follows from $e^{-h(t)}(g_0 - \frac{t^{1/4}}{\sqrt{d}} I) \le e^{-h(t)} g_0 \le g_0$, and we exploited on several occasions that $\det g_0$ is bounded away from zero to guarantee that $g_0 - \frac{t^{1/4}}{\sqrt{d}} I$ remains positive definite at least for small times. The first term in the r.h.s. of (7.23) behaves as $t^{\frac{1}{4}} |\log h(t)|$, which tends to zero as $t \to 0^+$ due to our current assumptions on h. The second term tends to $-\int_\Omega \mathrm{d}\Lambda \log\det g_0 = E(\tilde{G}_0)$. (Our temporary assumption that $\det g_0$ is bounded away from zero allows to apply Lebesgue's dominated convergence theorem and we omit the details.) Finally, the third member tends to zero due to (7.22) and $\log\det g_0 \in L^1(\mathrm{d}\Lambda)$, by absolute continuity of the integral.

Step 2: for general g_0 we argue now by approximation. Consider the curve

$$(G^n)_t := \mathsf{S}_{\frac{1}{n}}(G_t),$$

and let $(\tilde{G}^n)_t = \mathsf{S}_{h(t)}(G^n_t)$ be the perturbed curve constructed as before but starting now from G^n instead of G. By (7.19) with $h(t) \equiv \frac{1}{n}$,

$$|\dot{G}^n_t|^2 \leq e^{-\frac{1}{n}}|\dot{G}_t|^2 \tag{7.24}$$

for a.a. $t \in (0,1)$ and therefore $G^n \in AC^2(0,1;\mathbb{P})$. Obviously

$$\det g^n_0(x) = \det\left[I + e^{-\frac{1}{n}}(g_0(x) - I)\right] \geq \det\left[(1 - e^{-\frac{1}{n}})I\right]$$

is bounded away from 0 (for fixed n), hence from step 1

$$E(\tilde{G}^n_0) \geq \limsup_{t\to 0^+} E(\tilde{G}^n_t). \tag{7.25}$$

Note that by Lemma 6.2 the integrand in the entropy $E(\tilde{G}^n_t)$ is nondecreasing in n for any $x \in \Omega$ (t is fixed here), thus by Beppo Levi's monotone convergence theorem

$$\lim_{n\to\infty} E(\tilde{G}^n_t) = E(\tilde{G}_t) \quad \text{for all} \quad t \in [0,1]. \tag{7.26}$$

Set

$$E^n_+(t) := \int_0^t \left[\frac{d}{ds}E(\tilde{G}^n_s)\right]_+ ds, \qquad E^n_-(t) := E(\tilde{G}^n_t) - E^n_+(t).$$

That $\left[\frac{d}{ds}E(\tilde{G}^n_s)\right]_+ \geq 0$ is indeed integrable in this definition follows from (7.19) and our assumption that $h'(t)$ is bounded away from zero (and positive) near $t = 0$. Recall from Lemma 7.3 that $t \mapsto E(\tilde{G}_t)$ is locally absolutely continuous, so the only scenario possibly contradicting the upper semicontinuity would be an initial upwards jump. In the limit $n \to +\infty$ the "good part" E^n_+ accordingly contains the nondecreasing absolutely continuous part of $E(\tilde{G}^n_t)$, while the "bad part" E^n_- contains the nonincreasing absolutely continuous part together with any possible (asymptotically) "bad" jump at $t = 0^+$. Dropping the first two non-negative terms in (7.19) and integrating in time from 0 to small $t > 0$ implies that

$$0 \le E_+^n(t) \le \frac{1}{2\inf_{[0,t]} h'} \int_0^t |\dot{G}_s^n|^2 \, \mathrm{d}s \le C \int_0^t |\dot{G}_s|^2 \, \mathrm{d}s,$$

where the last inequality follows from (7.24). Note that we have employed the fact that h' is positive and bounded away from 0 near $t = 0$. Since G is AC^2 in the right-hand side, we infer that

$$\limsup_{t\to 0^+} \limsup_{n\to\infty} E_+^n(t) = 0. \tag{7.27}$$

It is clear that $E_-^n(t)$ is a non-increasing function at least for $t > 0$. Using the elementary fact that $\lim_{n\to\infty}(a_n + b_n) = \liminf_{n\to\infty} a_n + \limsup_{n\to\infty} b_n$ for any real sequences such that the first limit exists, we estimate

$$\begin{aligned}
\limsup_{t\to 0^+} E(\tilde{G}_t) &= \limsup_{t\to 0^+} \lim_{n\to\infty} E(\tilde{G}_t^n) = \limsup_{t\to 0^+} \lim_{n\to\infty} (E_-^n(t) + E_+^n(t)) \\
&= \limsup_{t\to 0^+} \left(\liminf_{n\to\infty} E_-^n(t) + \limsup_{n\to\infty} E_+^n(t) \right) \\
&\le \limsup_{t\to 0^+} \liminf_{n\to\infty} E_-^n(t) + \limsup_{t\to 0^+} \limsup_{n\to\infty} E_+^n(t) \\
&\overset{(7.27)}{=} \limsup_{t\to 0^+} \liminf_{n\to\infty} E_-^n(t) \le \sup_{t>0} \liminf_{n\to\infty} E_-^n(t).
\end{aligned}$$

Leveraging the time monotonicity $E_-^n \downarrow$ for $t > 0$ we continue as

$$\begin{aligned}
\sup_{t>0} \liminf_{n\to\infty} E_-^n(t) &\le \liminf_{n\to\infty} \sup_{t>0} E_-^n(t) \\
&= \liminf_{n\to\infty} \limsup_{t\to 0^+} E_-^n(t) \le \liminf_{n\to\infty} \limsup_{t\to 0^+} (E_-^n(t) + E_+^n(t)) \\
&= \liminf_{n\to\infty} \limsup_{t\to 0^+} E(\tilde{G}_t^n) \overset{(7.25)}{=} \liminf_{n\to\infty} E(\tilde{G}_0^n) \overset{(7.26)}{=} E(\tilde{G}_0).
\end{aligned}$$

■

We start now carefully choosing some specific $h(t)$ in order to retrieve quantitative information

Lemma 7.6. *For $\varepsilon > 0$ and a given path $G \in AC^2(0,1;\mathbb{P}_{FR})$ with $E(G_0), E(G_1) < +\infty$, let G^ε be the path obtained as*

$$G_t^\varepsilon := \mathsf{S}_{h(t)} G_t \quad \textit{with} \quad h(t) := \varepsilon \min(t, 1-t).$$

Then $G^\varepsilon \in AC^2(0,1;\mathbb{P}_{FR})$*,* $F(G^\varepsilon) \in L^1(0,1)$*, and*

$$\frac{1}{2}\int_0^1 \left|\dot{G}_t^\varepsilon\right|^2 \mathrm{d}t + \frac{\varepsilon^2}{2}\int_0^1 F(G_t^\varepsilon)\,\mathrm{d}t \le \frac{1}{2}\int_0^1 \left|\dot{G}_t\right|^2 + \varepsilon(E(G_0)+E(G_1)). \tag{7.28}$$

Proof: Take (7.19) and integrate by parts separately on the time intervals $(\delta, 1/2)$ and $(1/2, 1-\delta)$ for some small $\delta > 0$ to obtain

$$\begin{aligned} \frac{1}{2}\int_\delta^{1-\delta} \left|\dot{G}_t^\varepsilon\right|^2 \mathrm{d}t + \frac{\varepsilon^2}{2}\int_\delta^{1-\delta} F(G_t^\varepsilon)\,\mathrm{d}t + 2\varepsilon E(G_{1/2}^\varepsilon) & \\ \le \frac{1}{2}\int_\delta^{1-\delta} \left|\dot{G}_t\right|^2 + \varepsilon(E(G_\delta^\varepsilon)+E(G_{1-\delta}^\varepsilon)). & \end{aligned} \tag{7.29}$$

(Note that $h''(t) \equiv 0$ separately inside both intervals.) The odd-looking term $2\varepsilon E(G_{1/2}^\varepsilon)$ arises from the two boundary terms at $t = 1/2^\pm$ in the two integrations by parts, and can be safely discarded since it is non-negative. Let us point out that although this term is simply ignored here, it will be crucial later on in our proof of the geodesic convexity, Theorem 8.

Our assumption that $G \in AC^2$ together with Lemma 7.5 give a uniform bound for the right-hand side

$$\begin{aligned} \limsup_{\delta\to 0} \frac{1}{2}\int_\delta^{1-\delta} \left|\dot{G}_t\right|^2 + \varepsilon(E(G_\delta^\varepsilon)+E(G_{1-\delta}^\varepsilon)) & \\ = \frac{1}{2}\int_0^1 \left|\dot{G}_t\right|^2 + \varepsilon(E(G_0^\varepsilon)+E(G_1^\varepsilon)) & \\ = \frac{1}{2}\int_0^1 \left|\dot{G}_t\right|^2 + \varepsilon(E(G_0)+E(G_1)) < +\infty & \end{aligned}$$

(since we assumed that the endpoints have finite entropy). As a consequence (7.29) holds with $\delta = 0$, which is exactly our claim. ∎

We now have enough technical tools to prove the Γ-convergence. For $G_0, G_1 \in \mathbb{P}$ and $G \in C([0,1];\mathbb{P})$, let

$$\iota_{G_0,G_1}(G) = \begin{cases} 0 & \text{if } G|_{t=0} = G_0 \text{ and } G|_{t=1} = G_1 \\ +\infty & \text{otherwise} \end{cases}$$

be the convex indicator of the endpoint constraints. For any $G \in C([0,1];\mathbb{P})$ we define the kinetic action

$$\mathcal{K}(G) := \frac{1}{2}\int_0^1 |\dot{G}_t|^2_{FR}\,\mathrm{d}t,$$

with the usual convention that
$\mathcal{K}(G) = +\infty$ whenever $G \notin AC^2(0,1;\mathbb{P}_{FR})$. We also set

$$\mathcal{F}(G) := \int_0^1 F(G_t)\,\mathrm{d}t$$

where the integral may be infinite.

Theorem 7. *Let $G_0, G_1 \in \mathbb{P}$ with $E(G_0), E(G_1) < +\infty$. Then*

$$\Gamma - \lim_{\epsilon\to 0}\left[\mathcal{K} + \frac{\epsilon^2}{2}\mathcal{F} + \iota_{G_0,G_1}\right] = \mathcal{K} + \iota_{G_0,G_1} \tag{7.30}$$

both in the strong uniform topology of $C([0,1];\mathbb{P})$ and pointwise-in-time weak-$*$ topology of $\mathbb{P} \subset (C_0(\Omega;\mathcal{H}))^*$.

Proof: Since the strong uniform topology (w.r.t. to TV or d_{FR}, equivalently) is stronger than the pointwise weak-$*$ topology, it suffices to prove the $\Gamma-\liminf$ for the latter topology and the $\Gamma-\limsup$ for the former one.

The $\Gamma-\liminf$ is quite obvious since $\mathcal{K}$ is lower-semicontinuous w.r.t. pointwise-in-time weak-$*$ topology (Proposition 6.5), the convex indicator ι_{G_0,G_1} is also l.s.c., and $\mathcal{K} + \iota_{G_0,G_1} \leq \mathcal{K} + \epsilon^2\mathcal{F} + \iota_{G_0,G_1}$.

To prove the $\Gamma-\limsup$, for any sequence $\varepsilon_n \to 0$ it suffices to construct a recovery sequence for (7.30). Moreover, we can restrict to curves G such that the right-hand side of (7.30) is finite, that is, $G \in AC^2(0,1;\mathbb{P}_{FR})$ with $G(0) = G_0$, $G(1) = G_1$ (otherwise there is noting to prove). Thus, the desired recovery sequence $G^n \in C([0,1];\mathbb{P})$ should satisfy

$$\limsup_{n\to\infty}\left\{\frac{1}{2}\int_0^1 |\dot{G}^n_t|^2\mathrm{d}t + \frac{\varepsilon_n^2}{2}\int_0^1 F(G^n_t)\mathrm{d}t\right\} \leq \frac{1}{2}\int_0^1 |\dot{G}_t|^2 \tag{7.31}$$

with $G^n(0) = G_0$ and $G^n_1 = G_1$. By (7.28), the sequence G^{ε_n} from Lemma 7.6 does the job. Indeed, the time-continuity of G^{ε_n} follows from Lemma 7.2. It remains to check that $G^{\varepsilon_n} \to G$ in the uniform topology of $C([0,1];\mathbb{P})$. From the representation formula (6.9) for solutions of the heat flow we compute explicitly

$$\|G_t^{\varepsilon_n} - G_t\|_{TV} = \left\| \left(\Lambda I + e^{-\varepsilon_n \min(t,1-t)}(G_t - \Lambda I)\right) - G_t \right\|_{TV}$$
$$= \left(1 - e^{-\varepsilon_n \min(t,1-t)}\right)\|G_t - \Lambda I\|_{TV} \leq \left(1 - e^{-\varepsilon_n \min(t,1-t)}\right)(\|G_t\|_{TV} + \|\Lambda I\|_{TV})$$
$$= 2\left(1 - e^{-\varepsilon_n \min(t,1-t)}\right) \to 0$$

uniformly in t as $\varepsilon_n \to 0$. By the lower bound in (4.7) and Lemma 3.3 we conclude that

$$d_{FR}^2(G_t^{\varepsilon_n}, G_t) \leq C d_H^2(G_t^{\varepsilon_n}, G_t) \leq C\|G_t^{\varepsilon_n} - G_t\|_{TV} \to 0$$

uniformly in t as $\varepsilon_n \to 0$, and the proof is complete. ∎

As an immediate and natural consequence we have that the ε-geodesics (minimizers of $\mathcal{K} + \varepsilon^2\mathcal{F}$) converge to FR-geodesics (minimizers of $\mathcal{K}$):

Corollary 7.7. *Let $\varepsilon_k \searrow 0$ and G^k be the corresponding solution of the Schrödinger problem* (6.13) *with $\varepsilon = \varepsilon_k$ and $E(G_0), E(G_1) < \infty$. Then there exists $G \in C([0,1];\mathbb{P})$ such that, up to a subsequence,*

$$G_t^k \xrightarrow[k\to\infty]{} G_t \qquad \textit{weakly-}*$$

for every $t \in [0,1]$, and G solves the geodesic problem (5.1). *Moreover,*

$$2\mathfrak{S}_{\varepsilon_k}(G_0, G_1) \to d_{FR}^2(G_0, G_1).$$

Proof: Recall that Γ-convergence precisely guarantees that limits of minimizers are minimizers, [14, Theorem 1.21], thus in view of Theorem 7 it suffices to prove that the set of minimizers $\{G^k\}$ is relatively compact in the pointwise-in-time weak-∗ topology. Indeed, the AC^2-energies of the curves G^k are uniformly bounded since

$$\mathcal{K}(G^k) \leq \mathcal{K}(G^k) + \frac{\epsilon_k^2}{2}\mathcal{F}(G^k) \leq \mathcal{K}(G^1) + \frac{\epsilon_k^2}{2}\mathcal{F}(G^1) \leq \mathcal{K}(G^1) + \frac{\epsilon_1^2}{2}\mathcal{F}(G^1) < +\infty.$$

By the fundamental estimate (4.4) with $m_t^k = M = 1$ on $\mathbb{P}$ we get

$$\forall\, t,s \in [0,1],\ \forall k \in \mathbb{N}: \qquad \|G_s^k - G_t^k\|_{TV} \leq C|t-s|^{1/2}.$$

Arguing as in the proof of Lemma 5.11, we deduce that there exists a TV-continuous curve $(G_t)_{t\in[0,1]}$ connecting G_0 and G_1 such that

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

along some subsequence $k \to \infty$. ∎

Remark 7.8. If $E(G_0)$ or $E(G_1)$ are infinite, a careful regularization (involving of course the heat flow) allows to prove that the ε_n-geodesics with suitably regularized endpoints $G_0^n, G_1^n \to G_0, G_1$ still converge to the geodesic with endpoints G_0, G_1. The statement and proof both become slightly more involved (essentially one should make sure to regularize enough so that $\varepsilon_n[E(G_0^n) + E(G_1^n)] \to 0$), and we omit the details for the sake of brevity.

We finish this section with a slightly different and perhaps unexpected consequence of our previous construction of the regularized curves:

Theorem 8. *The entropy $E : \mathbb{P} \to \mathbb{R}^+$ is $\frac{1}{2}$-geodesically convex for the Fisher-Rao distance, namely for any $G_0, G_1 \in \mathbb{P}$ and any FR geodesic $(G_\theta)_{\theta\in[0,1]}$ joining G_0, G_1 there holds*

$$E(G_\theta) \le (1-\theta)E(G_0) + \theta E(G_1) - \frac{1}{4}\theta(1-\theta)d_{FR}^2(G_0, G_1), \forall \theta \in (0,1). \tag{7.32}$$

Proof: Observe that our statement is vacuous if either of the endpoints G_0, G_1 has infinite entropy, thus we need only consider

$$E(G_0), E(G_1) < +\infty.$$

Pick any geodesic $(G_t)_{t\in[0,1]}$, fix $\theta \in (0,1)$, and let

$$H_\theta(t) := \begin{cases} \dfrac{1}{\theta}t & \text{if } t \in [0,\theta], \\ -\dfrac{1}{1-\theta}(t-1) & \text{if } t \in [\theta, 1] \end{cases}$$

be the hat function centered at $t = \theta$ with height 1 and vanishing at the boundaries. Setting $h(t) := \varepsilon H_\theta(t)$ for small $\varepsilon > 0$, let

$$G_t^\varepsilon := \mathsf{S}_{h(t)} G_t$$

be the perturbed curve constucted as before. Integrating (7.19) in time separately on $[0,\theta]$ and $[\theta,1]$, discarding the nonnegative Fisher information term, and leveraging the continuity of the entropy from Lemma 7.5 (with $G_t^\varepsilon = G_t$ at the endpoints $t = 0,1$), we get

$$\frac{1}{2}\int_0^1 \left|\dot{G}_t^\varepsilon\right|^2 \mathrm{d}t + 0 + \frac{\varepsilon}{\theta}\left[E(G_\theta^\varepsilon) - E(G_0)\right] + \frac{\varepsilon}{1-\theta}\left[E(G_\theta^\varepsilon) - E(G_1)\right]$$
$$\le \frac{1}{2}\int_0^1 e^{-\varepsilon H_\theta(t)} \left|\dot{G}_t\right|^2 \mathrm{d}t.$$

Because G^ε is an admissible curve connecting G_0, G_1 the first term in the left-hand side is larger than the AC^2 energy of the minimizing geodesic G: Multiplying by $\theta(1-\theta)/\varepsilon > 0$ and rearranging gives

$$E(G^\varepsilon_\theta) \leq (1-\theta)E(G_0) + \theta E(G_1) + \frac{\theta(1-\theta)}{2}\int_0^1 \frac{e^{-\varepsilon H_\theta(t)}-1}{\varepsilon}|\dot{G}_t|^2 \mathrm{d}t.$$

Since $G^\varepsilon \to G$ weakly-$*$ for any fixed time, the lower semicontinuity of the entropy allows to take the $\liminf$ as $\varepsilon \to 0$ in the left-hand side. For the right-hand side, observe that because G is a minimizing geodesic it has constant speed, $|\dot{G}_t|^2 = cst = d^2_{FR}(G_0, G_1)$. As a consequence we get

$$\begin{aligned} E(G_\theta) &\leq \liminf_{\varepsilon\to 0} E(G^\varepsilon_\theta) \\ &\leq (1-\theta)E(G_0) + \theta E(G_1) + \lim_{\varepsilon \to 0}\frac{\theta(1-\theta)}{2} d^2_{FR}(G_0,G_1)\int_0^1 \frac{e^{-\varepsilon H_\theta(t)}-1}{\varepsilon}\mathrm{d}t \\ &= (1-\theta)E(G_0) + \theta E(G_1) - \frac{\theta(1-\theta)}{2} d^2_{FR}(G_0,G_1)\int_0^1 H_\theta(t)\mathrm{d}t. \end{aligned}$$

Our statement finally follows from $\int_0^1 H_\theta(t)\mathrm{d}t = \frac{1}{2}$ for all θ. ■

Appendix A.Some technical lemmas

Lemma A.1 (Constant-speed reparametrization)**.** *Let (G_t, U_t) be a curve connecting G_0, G_1 with finite energy*

$$E[G,U] = \int_0^1 \int_\Omega \mathrm{d}G_t U_t : U_t \mathrm{d}t < +\infty.$$

Then there exists a curve $(\check{G}_t, \check{U}_t)_{t\in[0,1]}$ connecting $\check{G}_0 = G_0$ to $\check{G}_1 = G_1$ with

$$\|\check{U}_t\|_{L^2(\mathrm{d}\check{G}_t)} \equiv cst \quad \text{and} \quad E[\check{G},\check{U}] = \left(\int_0^1 \|U_t\|_{L^2(\mathrm{d}G_t)}\mathrm{d}t\right)^2 \leq E[G,U]$$

with strict inequality unless $\|U_t\|_{L^2(\mathrm{d}G_t)}$ is constant in time.

Proof: The argument is fairly standard (see e.g. [45, lemma 5.3] or [5, lemma 1.1.4]) hence we only sketch the idea. Consider the change of time variable

$$\mathrm{s}(t) := \int_0^t \|U_\tau\|_{L^2(\mathrm{d}G_\tau)}\mathrm{d}\tau, \quad \text{and} \quad \mathrm{t}(s) := \min\{t \in [0,1] : \quad \mathrm{s}(t) = s\}$$

(t is the left inverse of s). Setting

$$\widetilde{G}_s := G_{\mathrm{t}(s)}, \qquad \widetilde{U}_s := \frac{U_{\mathrm{t}(s)}}{\|U_{\mathrm{t}(s)}\|_{L^2(\widetilde{G}_s)}}$$

gives an admissible path connecting $\widetilde{G}_0 = G_0$ to $\widetilde{G}_L = G_1$ in time $s \in [0, L]$ with

$$L = \mathrm{s}(1) = \int_0^1 \|U_\tau\|_{L^2(\mathrm{d}G_\tau)} \mathrm{d}\tau$$

and unit speed $\|\widetilde{U}_s\|_{L^2(\mathrm{d}\widetilde{G}_s)} \equiv 1$. This is clear at least formally from the chain-rule

$$\partial_s \widetilde{G}_s = \frac{\mathrm{dt}}{\mathrm{d}s}(s) \partial_t G_t(\mathrm{t}(s)) = \frac{1}{\frac{\mathrm{ds}}{\mathrm{d}t}(\mathrm{t}(s))} \left(G_{\mathrm{t}(s)} U_{\mathrm{t}(s)} \right)^{Sym}$$
$$= \frac{1}{\|U_{\mathrm{t}(s)}\|_{L^2(\mathrm{d}G_{\mathrm{t}(s)})}} \left(G_{\mathrm{t}(s)} U_{\mathrm{t}(s)} \right)^{Sym} = (\widetilde{G}_s \widetilde{U}_s)^{Sym}$$

and can be made rigorous since the denominator $\|U_{\mathrm{t}(s)}\|_{L^2(\mathrm{d}G_{\mathrm{t}(s)})}$ only vanishes at the discontinuity points of $\mathrm{t}(s)$, which are countable (being t monotone nondecreasing) and therefore ds-negligible. Scaling $t = Ls$

$$(\check{G}_t, \check{U}_t) := (\widetilde{G}_{tL}, L\widetilde{U}_{tL}) \qquad \text{for } t \in [0,1]$$

back to the unit interval and noticing that $\|\check{U}_t\|_{L^2(\mathrm{d}\check{G}_t)} \equiv L$ is constant, we get

$$E[\check{G}, \check{U}] = \int_0^1 \|\check{U}_t\|^2_{L^2(\mathrm{d}\check{G}_t)} \mathrm{d}t = \int_0^1 L^2 \mathrm{d}t = L^2$$
$$= \left(\int_0^1 \|U_\tau\|_{L^2(\mathrm{d}G_\tau)} \mathrm{d}\tau \right)^2 \leq \int_0^1 \|U_\tau\|^2_{L^2(\mathrm{d}G_\tau)} \mathrm{d}\tau = E[G, U]$$

as desired. (The inequality is strict unless (G, U) has constant speed as in our statement). ∎

Lemma A.2 (Refined Banach-Alaoglu [45]). *Let $(X, \|\cdot\|)$ be a separable normed vector space. Assume that there exists a sequence of seminorms $\{\|\cdot\|_k\}$ ($k = 0, 1, 2, \dots$) on X such that for every $x \in X$ one has*

$$\|x\|_k \leq C\|x\|$$

with a constant C independent of k,x, and

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

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

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

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

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

Lemma A.3 (Refined Arzelà-Ascoli [5, 15])**.** *Let (X,ρ) be a metric space. Assume that there exists a Hausdorff topology σ on X such that ρ is sequentially lower semicontinuous with respect to σ. Let $(x^k)_t$, $t \in [0,1]$, be a sequence of curves lying in a common σ-sequentially compact set $K \subset X$. Let it be equicontinuous in the sense that there exists a symmetric continuous function $\omega : [0,1]\times[0,1] \to \mathbb{R}_+$, $\omega(t,t) = 0$, such that*

$$\rho((x^k)_t, (x^k)_{\bar t}) \leq \omega(t, \bar t). \tag{A.2}$$

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

$$\rho(x_t, x_{\bar t}) \leq \omega(t, \bar t), \tag{A.3}$$

and (up to a not relabelled subsequence)

$$(x^k)_t \to x_t \tag{A.4}$$

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

Proof of Proposition 3.4: First of all, let us show that the right-hand side is always finite and that the minimum is always attained. For any fixed $A_1 \in \mathcal{H}^+$ it is easy to check that $(A_t, U_t) := (t^2 A_1, \frac{2}{t} I)$ gives an admissible path connecting $A_0 = 0$ to A_1 with finite energy. In particular any two matrices $A_0, A_1 \in \mathcal{H}^+$ can be connected through zero as $A_0 \rightsquigarrow 0 \rightsquigarrow A_1$ with finite cost, thus the problem is proper. For fixed A_0, A_1 consider now a minimizing sequence $(A_t^n, U_t^n)_{t\in[0,1]}$. Note that our Lemma 4.2 applies in particular when $\Omega = \{x\}$ is a one-point space, which gives here equicontinuity and pointwise relative compactness in the form

$$|A_t^n - A_s^n|_2 \leq C|t-s|^{\frac{1}{2}}, \qquad m_t^n = \operatorname{tr} A_t^n \leq M, \qquad s,t \in [0,1]$$

uniformly in n. By the classical Arzelá-Ascoli theorem we get, up to extraction of a subsequence if needed,

$$A^n \to A \qquad \text{uniformly in } C([0,1];\mathcal{H}^+).$$

This immediately shows that the matrix-valued measure

$$\mu^n := A^n_t \mathrm{d}t \to A_t dt =: \mu$$

at least weakly-$*$ on $\mathbb{H}^+(0,1)$. Because

$$\|U^n\|^2_{L^2(\mathrm{d}\mu^n)} = \int_0^1 A^n_t U^n_t : U^n_t \,\mathrm{d}t \leq C,$$

an easy application of our Banach-Alaoglu variant (lemma A.2) in varying $L^2(\mathrm{d}\mu^n)$ spaces gives a limit $U \in L^2(\mathrm{d}\mu)$ with

$$\int_0^1 A_t U_t : U_t \mathrm{d}t \leq \liminf_{n\to\infty} \int_0^1 A^n_t U^n_t : U^n_t \,\mathrm{d}t$$

and such that $\int_0^1 A^n U^n : V\,\mathrm{d}t \to \int_0^1 AU : V\,\mathrm{d}t$ for any reasonably smooth test function V. This shows that this limit $(A_t, U_t)_{t\in[0,1]}$ is an admissible curve joining A_0, A_1 with energy

$$E[A,U] \leq \liminf E[A^n, U^n]$$

and this pair is therefore a minimizer.

In order to identify now the left-hand side and the right-hand side of (3.5) we proceed in two steps.

Step 1: assume that $A_0, A_1 \in \mathcal{H}^{++}(d)$ are positive definite, and let $\check{A}_{0,1} := \mathfrak{r}(A_{0,1})$ be the corresponding real extensions as defined in (3.8). From Proposition 3.6 there holds

$$d^2_B(A_0, A_1) = \frac{1}{2} W^2_2(\mathcal{N}(\check{A}_0), \mathcal{N}(\check{A}_1)).$$

In this real setting it is known [65, Prop. A] that

$$W^2_2(\mathcal{N}(\check{A}_0), \mathcal{N}(\check{A}_1)) = \frac{1}{4} \min_{\check{A}\in\mathcal{A}(\check{A}_0,\check{A}_1)} \int_0^1 \check{A}_t \check{U}_t : \check{U}_t \,\mathrm{d}t,$$

where the infimum runs of course over real pairs $\frac{d\check{A}_t}{dt} = (\check{A}_t \check{U}_t)^{Sym}$ with $\check{U}_t \in \mathcal{S}(2d)$. Complexifying back $(\check{A}, \check{U}) \rightsquigarrow (A, U)$ gives the result.

Step 2: if now either A_0 or A_1 are only *semi*-definite we approximate $A^n_0 :=$

$A_0 + \frac{1}{n}I \to A_0$ and $A_1^n := A_1 + \frac{1}{n}I \to A_1$. Clearly $A_{0,1}^n \in \mathcal{H}^{++}$ are positive-definite, hence step 1 applies.

Note from (3.4) that the left-hand side of (3.5) is of course continuous in A_0, A_1, hence it suffices to show that the optimal value in the right-hand side is continuous for this particular choice of $A_{0,1}^n \to A_{0,1}$. Observe that

$$\mathcal{J}^*(A_0, A_1) := \frac{1}{4} \min_{\mathcal{A}(A_0,A_1)} \int_0^1 A_t U_t : U_t \mathrm{d}t$$

is a well-defined function of $A_0, A_1 \in \mathcal{H}^+$, and we proved earlier that there always exists a minimizer. Arguing precisely as for the existence of the minimizers, cf. also Lemma 5.11, it is easy to prove that $\mathcal{J}^*$ is lower semi-continuous in both arguments. Indeed, up to a subsequence, any sequence $(A_t^n, U_t^n)_{t\in[0,1]}$ of minimizers in $\mathcal{J}^*(A_0^n, A_1^n)$ converges to an admissible candidate $(A_t, U_t)_{t\in[0,1]}$ connecting A_0, A_1, hence $\mathcal{J}^*(A_0, A_1) \le E[A, U] \le \liminf E[A^n, U^n] = \liminf \mathcal{J}^*(A_0^n, A_1^n)$ (regardless of the particular form of A_0^n, A_1^n).

In order to establish the upper continuity, let $A_0 = RD_0R^*$ be a spectral decomposition of A_0 and note that obviously $A_0^n = A_0 + \frac{1}{n}I = R(D_0 + \frac{1}{n}I)R^*$. Since A_0^n and A_0 commute it is easy to check that

$$A_{0t}^n := R\left[(1-t)\sqrt{D_0} + t\sqrt{D_0 + \frac{1}{n}I}\right]^2 R^*,$$

$$U_{0t}^n := 2R\left(\sqrt{D_0 + \frac{1}{n}I} - \sqrt{D_0}\right)R^*\sqrt{A_{0t}^{-1}}$$

defines an admissible path $(A_{0t}^n)_{t\in[0,1]}$ between A_0 and A_0^n. Moreover, a straightforward computation shows that the corresponding energy is

$$\frac{1}{4}\int_0^1 A_{0t}^n U_{0t}^n : U_{0t}^n \mathrm{d}t = \left|\sqrt{D_0} - \sqrt{D_0 + I/n}\right|_2^2 \xrightarrow[n\to\infty]{} 0. \tag{A.5}$$

(Actually this path is exactly the Bures geodesic between A_0, A_0^n.) A similar construction yields a path $(A_{1t}^n)_{t\in[0,1]}$ connecting A_1^n to A_1 with cost

$$\frac{1}{4}\int_0^1 A_{1t}^n U_{1t}^n : U_{1t}^n \mathrm{d}t = \left|\sqrt{D_1} - \sqrt{D_1 + I/n}\right|_2^2 \xrightarrow[n\to\infty]{} 0, \tag{A.6}$$

where D_1 is the spectral decomposition of $A_1 = QD_1Q^*$. Now pick a minimizer $(\tilde{A}_t, \tilde{U}_t)_{t\in[0,1]}$ in the definition of $\mathcal{J}^*(A_0, A_1)$ and fix a small $\theta_n \in (0,1)$

to be determined shortly. Rescaling in time and concatenating the paths $A_0^n \rightsquigarrow A_0 \rightsquigarrow A_1 \rightsquigarrow A_1^n$ in the intervals $t \in [0, \theta_n]$, $t \in [\theta_n, 1-\theta_n]$, and $t \in [1-\theta_n, 1]$, respectively, we obtain an admissible path $(\hat{A}^n, \hat{U}^n)$ from A_0^n to A_1^n whose energy is bounded as

$$\begin{aligned} 4\mathcal{J}^*(A_0^n, A_1^n) &\leq \int_0^1 \hat{A}_t^n \hat{U}_t^n : \hat{U}_t^n \,\mathrm{d}t \\ &= \frac{1}{\theta_n}\int_0^1 A_{0t}^n U_{0t}^n : U_{0t}^n \,\mathrm{d}t + \frac{1}{1-2\theta_n}\int_0^1 \tilde{A}_t \tilde{U}_t : \tilde{U}_t \,\mathrm{d}t \\ &\qquad + \frac{1}{\theta_n}\int_0^1 A_{1t}^n U_{1t}^n : U_{1t}^n \,\mathrm{d}t. \end{aligned}$$

By (A.5), (A.6) we see that the first and third integrals in the right-hand side tend to zero, while the second integral is exactly

$$4\mathcal{J}^*(A_0, A_1)$$

by definition of $(\tilde{A}, \tilde{U})$. Choosing $\theta_n \to 0$ sufficiently slowly and taking lim sup in the previous inequality gives

$$\limsup_{n\to\infty} 4\mathcal{J}^*(A_0^n, A_1^n) \leq \lim_{n\to\infty} \frac{1}{1-2\theta_n}\int_0^1 \tilde{A}_t \tilde{U}_t : \tilde{U}_t \,\mathrm{d}t + 0 = 4\mathcal{J}^*(A_0, A_1)$$

and the proof is complete. ∎

Proof of Proposition 3.8: Let $\rho_0 = \mathcal{N}(A_0)$ and $\rho_1 = \mathcal{N}(A_1)$. By [49, Thm. 3.3 and Thm. 3.4], if one could write the (f, g) transform

$$\begin{cases} \rho_0 = f_0 g_0 \\ \rho_1 = f_1 g_1 \end{cases} \tag{A.7}$$

for $(f_t)_{t\in[0,1]}$ a *forward* solution of the heat equation with initial datum f_0 and $(g_t)_{t\in[0,1]}$ a *backward* solution with terminal datum g_1, then the solution of (3.16) would be given by

$$\rho_t = f_t g_t.$$

Since the product of Gaussian distributions is Gaussian, it is legitimate to try and solve for $f_0 = \mathcal{N}(B_0)$ and $g_1 = \mathcal{N}(C_1)$ as Gaussians. Since the heat flow for Gaussians is explicitly given by (3.14) we see that the corresponding forward and backward solutions read

$$\begin{cases} f_t = \mathcal{N}(B_t) & \text{with } B_t = B_0 + 2tI \\ g_t = \mathcal{N}(C_t) & \text{with } C_t = C_1 + 2(1-t)I \end{cases}$$

Exploiting the algebraic product rule

$$\mathcal{N}(B) \times \mathcal{N}(C) = \mathcal{N}\left([B^{-1} + C^{-1}]^{-1}\right),$$

we see that, given A_0, A_1 the Schrödinger system (A.7) is equivalent to solving

$$\begin{cases} \mathcal{N}(A_0) = \mathcal{N}(B_0)\mathcal{N}(C_0) \\ \mathcal{N}(A_1) = \mathcal{N}(B_1)\mathcal{N}(C_1) \end{cases} \Leftrightarrow \begin{cases} A_0^{-1} = B_0^{-1} + [C_1 + 2I]^{-1} \\ A_1^{-1} = [B_0 + 2I]^{-1} + C_1^{-1} \end{cases}$$

in $B_0, C_1 \in \mathcal{S}^{++}$. It is then a simple exercise to check that this system has a unique solution, and in particular the covariance A_t of $\rho_t = f_t g_t$ is fully determined by

$$A_t^{-1} = [B_0 + 2tI]^{-1} + [C_1 + 2(1-t)I]^{-1}.$$

∎

References

[1] N. A. Ahmed and D. Gokhale. Entropy expressions and their estimators for multivariate distributions. *IEEE Transactions on Information Theory*, 35(3):688–692, 1989.

[2] S. Albeverio, K. Yasue, and J. Zambrini. Euclidean quantum mechanics: analytical approach. In *Annales de l'IHP Physique théorique*, volume 50, pages 259–308, 1989.

[3] P. C. Alvarez-Esteban, E. del Barrio, J. A. Cuesta-Albertos, C. Matrán, et al. Wide consensus aggregation in the wasserstein space. application to location-scatter families. *Bernoulli*, 24(4A):3147–3179, 2018.

[4] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems*, volume 254. Clarendon Press Oxford, 2000.

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

[6] W. Arendt and M. Kreuter. Mapping theorems for Sobolev spaces of vector-valued functions. *Studia Math.*, 240(3):275–299, 2018.

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

[8] A. Baradat and L. Monsaingeon. Small noise limit and convexity for generalized incompressible flows, schrödinger problems, and optimal transport. *Archive for Rational Mechanics and Analysis*, 235(2):1357–1403, 2020.

[9] F. Barbaresco. Information geometry of covariance matrix: Cartan-siegel homogeneous bounded domains, Mostow/Berger fibration and Fréchet median. In *Matrix information geometry*, pages 199–255. Springer, 2013.

[10] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge - Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.
[11] I. Bengtsson and K. Życzkowski. *Geometry of quantum states: an introduction to quantum entanglement*. Cambridge University Press, Cambridge, 2017.
[12] R. Bhatia, T. Jain, and Y. Lim. On the bures–wasserstein distance between positive definite matrices. *Expositiones Mathematicae*, 37(2):165–191, 2019.
[13] P. Biane and D. Voiculescu. A free probability analogue of the wasserstein metric on the trace–state space. *Geometric & Functional Analysis GAFA*, 11(6):1125–1138, 2001.
[14] A. Braides. *Γ-convergence for beginners*, volume 22 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2002.
[15] Y. Brenier and D. Vorotnikov. On optimal transport of matrix-valued measures. *SIAM Journal on Mathematical Analysis*, 52(3):2849–2873, 2020.
[16] 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.
[17] D. Burago, Y. Burago, and S. Ivanov. *A course in metric geometry*. AMS, 2001.
[18] E. Carlen. Stochastic mechanics: a look back and a look ahead. *Diffusion, quantum theory and radically elementary mathematics*, 47:117–139, 2014.
[19] 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.
[20] E. A. Carlen and J. Maas. Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance. *J. Funct. Anal.*, 273(5):1810–1869, 2017.
[21] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer. Convergence of entropic schemes for optimal transport and gradient flows. *SIAM Journal on Mathematical Analysis*, 49(2):1385–1418, 2017.
[22] Y. Chen, W. Gangbo, T. T. Georgiou, and A. Tannenbaum. On the matrix Monge-Kantorovich problem. *European Journal of Applied Mathematics, to appear*.
[23] Y. Chen, T. T. Georgiou, and M. Pavon. On the relation between optimal transport and schrödinger bridges: A stochastic control viewpoint. *Journal of Optimization Theory and Applications*, 169(2):671–691, 2016.
[24] Y. Chen, T. T. Georgiou, and M. Pavon. Stochastic control liasons: Richard sinkhorn meets gaspard monge on a schroedinger bridge. *arXiv preprint arXiv:2005.10963*, 2020.
[25] Y. Chen, T. T. Georgiou, and A. Tannenbaum. Matrix optimal mass transport: a quantum mechanical approach. *IEEE Trans. Automat. Control*, 63(8):2612–2619, 2018.
[26] 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.
[27] Y. Chen, T. T. Georgiou, and A. Tannenbaum. Interpolation of matrices and matrix-valued densities: the unbalanced case. *European J. Appl. Math.*, 30(3):458–480, 2019.
[28] 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.
[29] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. Scaling algorithms for unbalanced optimal transport problems. *Mathematics of Computation*, 87(314):2563–2609, 2018.
[30] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems*, pages 2292–2300, 2013.

[31] F. Demengel and R. Temam. Convex functions of a measure and applications. *Indiana University Mathematics Journal*, 33(5):673–709, 1984.
[32] J. Dittmann. Explicit formulae for the Bures metric. *Journal of Physics A: Mathematical and General*, 32(14):2663, 1999.
[33] 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.
[34] A. Ferrante, C. Masiero, and M. Pavon. Time and spectral domain relative entropy: A new approach to multivariate spectral estimation. *IEEE Transactions on Automatic Control*, 57(10):2561–2575, 2012.
[35] A. Ferrante, M. Pavon, and F. Ramponi. Hellinger versus Kullback–Leibler multivariable spectrum approximation. *IEEE Transactions on Automatic Control*, 53(4):954–967, 2008.
[36] I. Fonseca and G. Leoni. *Modern Methods in the Calculus of Variations: L^p Spaces*. Springer Science & Business Media, 2007.
[37] C. Goffman and J. Serrin. Sublinear functions of measures and variational integrals. *Duke Math. J.*, 31:159–178, 1964.
[38] F. Golse, C. Mouhot, and T. Paul. On the mean field and classical limits of quantum mechanics. *Comm. Math. Phys.*, 343(1):165–205, 2016.
[39] F. Itakura. Analysis synthesis telephony based on the maximum likelihood method. In *The 6th international congress on acoustics, 1968*, pages 280–292, 1968.
[40] H. Janati, B. Muzellec, G. Peyré, and M. Cuturi. Entropic optimal transport between (unbalanced) gaussian measures has a closed form, 2020.
[41] X. Jiang, L. Ning, and T. T. Georgiou. Distances and riemannian metrics for multivariate spectral densities. *IEEE Transactions on Automatic Control*, 57(7):1723–1735, 2012.
[42] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker–Planck equation. *SIAM journal on mathematical analysis*, 29(1):1–17, 1998.
[43] B. Khesin, J. Lenells, G. Misiołek, and S. C. Preston. Geometry of diffeomorphism groups, complete integrability and geometric statistics. *Geom. Funct. Anal.*, 23(1):334–366, 2013.
[44] B. Khesin, G. Misiolek, and K. Modin. Geometric hydrodynamics of compressible fluids. *arXiv preprint arXiv:2001.01143*, 2020.
[45] 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.
[46] V. Laschos and A. Mielke. Geometric properties of cones with applications on the Hellinger-Kantorovich space, and a new distance on the space of probability measures. *J. Funct. Anal.*, 276(11):3529–3576, 2019.
[47] F. Léger. A geometric perspective on regularized optimal transport. *J. Dynam. Differential Equations*, 31(4):1777–1791, 2019.
[48] C. Léonard. From the Schrödinger problem to the Monge–Kantorovich problem. *Journal of Functional Analysis*, 262(4):1879–1920, 2012.
[49] C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *arXiv preprint arXiv:1308.0215*, 2013.
[50] 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.
[51] 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.
[52] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Annals of Mathematics*, pages 903–991, 2009.

[53] L. Malagò, L. Montrucchio, and G. Pistone. Wasserstein Riemannian geometry of Gaussian densities. *Inf. Geom.*, 1(2):137–179, 2018.
[54] 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.
[55] K. Modin. Geometry of matrix decompositions seen through optimal transport and information geometry. *Journal of Geometric Mechanics*, 9(3):335–390, 2017.
[56] L. Monsaingeon, L. Tamanini, and D. Vorotnikov. The dynamical schrödinger problem in abstract metric spaces: Γ-convergence and convexity. *In preparation*, 2020.
[57] B. Musicus and R. Johnson. Multichannel relative-entropy spectrum analysis. *IEEE transactions on acoustics, speech, and signal processing*, 34(3):554–564, 1986.
[58] L. Ning, T. T. Georgiou, and A. Tannenbaum. On matrix-valued monge–kantorovich optimal mass transport. *IEEE transactions on automatic control*, 60(2):373–382, 2014.
[59] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.
[60] G. Peyré, L. Chizat, F.-X. Vialard, and J. Solomon. Quantum entropic regularization of matrix-valued optimal transport. *European Journal of Applied Mathematics*, 30(6):1079–1102, 2019.
[61] G. Peyré, M. Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.
[62] R. T. Powers and E. Størmer. Free states of the canonical anticommutation relations. *Communications in Mathematical Physics*, 16(1):1–33, 1970.
[63] F. Santambrogio. *Optimal transport for applied mathematicians*. Birkhäuser/Springer, 2015.
[64] T. Séjourné, J. Feydy, F.-X. Vialard, A. Trouvé, and G. Peyré. Sinkhorn divergences for unbalanced optimal transport. *arXiv preprint arXiv:1910.12958*, 2019.
[65] A. Takatsu. Wasserstein geometry of gaussian measures. *Osaka Journal of Mathematics*, 48(4):1005–1026, 2011.
[66] T. Tao. *Topics in random matrix theory*. American Mathematical Soc., 2012.
[67] A. Uhlmann. The metric of Bures and the geometric phase. In *Groups and related Topics, Gielerak et al. (eds.)*, pages 267–274. Springer, 1992.
[68] C. Villani. *Topics in optimal transportation*. American Mathematical Soc., 2003.
[69] C. Villani. *Optimal transport: old and new*. Springer Science & Business Media, 2008.
[70] K. Yasue. Stochastic calculus of variations. *Journal of functional Analysis*, 41(3):327–340, 1981.
[71] J.-C. Zambrini. Variational processes and stochastic versions of mechanics. *Journal of Mathematical Physics*, 27(9):2307–2330, 1986.

Léonard Monsaingeon
GFM Universidade de Lisboa,
Campo Grande, Edifício C6, 1749-016 Lisboa, Portugal
& IECL Université de Lorraine, F-54506 Vandoeuvre-lès-Nancy Cedex, France
E-mail address: leonard.monsaingeon@univ-lorraine.fr

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