Journal of Statistical Physics (2021) 182:7
https://doi.org/10.1007/s10955-020-02682-1

Quantum Statistical Learning via Quantum Wasserstein Natural Gradient

Simon Becker[1] · **Wuchen Li**[2]

Received: 15 September 2020 / Accepted: 7 December 2020 / Published online: 7 January 2021

Abstract
In this article, we introduce a new approach towards the statistical learning problem $\operatorname{argmin}_{\rho(\theta)\in\mathcal{P}_\theta} W_Q^2(\rho_\star, \rho(\theta))$ to approximate a target quantum state $\rho_\star$ by a set of parametrized quantum states $\rho(\theta)$ in a quantum L^2-Wasserstein metric. We solve this estimation problem by considering Wasserstein natural gradient flows for density operators on finite-dimensional C^* algebras. For continuous parametric models of density operators, we pull back the quantum Wasserstein metric such that the parameter space becomes a Riemannian manifold with quantum Wasserstein information matrix. Using a quantum analogue of the Benamou–Brenier formula, we derive a natural gradient flow on the parameter space. We also discuss certain continuous-variable quantum states by studying the transport of the associated Wigner probability distributions.

Keywords Quantum transport information geometry · Quantum state estimation · Quantum Wasserstein information matrix · Quantum Wasserstein natural gradient · Quantum Schrödinger bridge problem

1 Introduction

The learning problem of quantum states, i.e. positive-definite trace class operators of unit trace, is central in modern quantum theory and commonly called *quantum state tomography*. The problem of quantum state estimation is ubiquitous in quantum mechanics and has a wide range of applications: This includes the analysis of optical devices [16] as well as the reliable estimation of qubit states in quantum computing [6,24]. Until this day, there have been many recent computationally efficient approaches towards the quantum state estimation problem

Communicated by Eric A. Carlen.

✉ Simon Becker
simon.becker@damtp.cam.ac.uk

Wuchen Li
wuchen@mailbox.sc.edu

[1] DAMTP, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK

[2] Department of Mathematics, University of South Carolina, Columbia, USA

based on compressed sensing and machine learning methods such as [20,40]. For a review of the most common classical approaches towards quantum state estimation, such as Maximal likelihood estimation (MLE), we refer to [36].

However, both in physics and non-commutative geometry, many problems come as a quantum state estimation problem in disguise: Over the past years, finding suitable physical descriptors for molecular structures from data has become a vast and growing area of research, cf. the review article [39] and references therein. Recently, such quantum machine learning approaches have also been based on optimization problems in Wasserstein distances, see for example [11], where a kernel ridge regression-based model relying on the Coulomb matrix is studied. The advantage of using the Wasserstein distance is that it leads to a continuous dependence on the position of the nuclei.

In said article, it has been discovered that it is key to use a suitable parametrization of the Coulomb matrix. This parametrization is ought to be invariant under 3D translations and rotations of the molecule and therefore related to the low-dimensional parametrization problem considered in this and previous articles, cf. [10]. Also, first attempts towards quantum Wasserstein generative adversarial networks have been considered in [14]. The quantum Wasserstein distance and its generalizations considered in [8,9] have also far-reaching applications beyond quantum mechanics to the field of non-commutative probability theory which includes multivariable time series and vector-valued random variables [34]. Hence, solving the quantum state estimation problem in Wasserstein distance has become an important and widely applicable problem.

The analysis of geometric properties of the space of quantum states is called quantum information geometry and is central in the field of quantum information. The asymptotic theory of quantum state estimation and quantum information geometry has been developed in the second half of the 1980s by Nagaoka [33]. A comprehensive review of the modern field of quantum information geometry and its connection to quantum estimation can be found in [21]. In this article, we develop a new connection between these two fields based on the quantum Wasserstein metric.

It has been discovered, among others, by Otto [35], that various PDEs evolve according to the gradient flow with respect to the L^2-Wasserstein metric [23]. Later, Carlen and Maas introduced in a series of articles [12,13,15] also quantum Wasserstein metrics for open quantum systems, satisfying a detailed balance condition. In these articles, they showed, that such open quantum systems also evolve according to the L^2-Wasserstein gradient flow. Moreover, they also showed that their metric allows for a dynamical formulation extending the classical Benamou–Brenier formula [4] to the quantum setting. Here, we also mention the work by Datta and Rouzé [37,38] for additional links to a quantum version of Ricci curvature and Fisher information functional. This analysis has been complemented by articles [8,9] where different types of non-commutative multiplication operators are considered with favorable properties from a computational point of view. Besides, Carlen and Maas showed that for certain open quantum systems the gradient flow of the relative entropy with respect to an invariant state in the quantum Wasserstein metric coincides with the quantum evolution governed by the Lindblad equation. For continuous-variable states, a quantum transport framework with desirable physical features has been proposed in [17]. However, a dynamical formulation of this approach does not seem to exist, yet. Results on the entropy flow for open quantum systems have also been obtained in [32]. Another relevant definition of the Wasserstein distance is due to Golse et al. [19] and has been proposed in the study of uniform mean-field limits of quantum systems in the semiclassical parameter.

Recently, optimal transport gradient flows have been applied to estimation problems in classical probability theory. In particular, the parameter estimation problem of probability

measures by using parameterized Wasserstein gradient flows on either Kullback–Leibler (KL) divergence, also referred to as relative entropy, or L^2-Wasserstein distance has been addressed by the second author [10,28,29]. This leads to a joint study between optimal transport [42] and information geometry [1,2], namely transport information geometry [26, 30]. Here, the natural gradient induced by optimal transport is first applied for statistical learning problems. Meanwhile, this approach also introduces a new estimation theory based on Wasserstein information matrix [25]. It also develops new scientific computing algorithms by the generative adversarial network to solve classical Fokker–Planck equations, in data-poor situations [27].

In this article, we present a new approach towards quantum state estimation based on L^2-quantum Wasserstein gradients. We extend the study of the previous paragraph to quantum systems. We start by studying the problem of minimizing the distance with respect to a quantum Wasserstein metric d, for some fixed target density operator $\rho_\star$ over a parametrized manifold of states $\mathcal{P}_\theta \subset \mathscr{D}(\mathcal{H})$, i.e. we aim to identify $\operatorname{argmin}_{\rho\in\mathcal{P}_\theta} d(\rho_\star, \rho)$. We address the corresponding estimation problem for particular finite and infinite-dimensional quantum states. In the case of infinite-dimensional states, our approach towards statistical learning is based on the Wigner transform of continuous-variable quantum states. This makes this approach particularly tailored to experimental quantum state estimation in continuous-variable systems, where the Wigner distribution of the quantum state is approximately recovered [41]. A classical choice of the distance between probability measures is the Kullback–Leibler (KL) divergence. In classical probability, the metric induced by the L^2 Hessian of the KL divergence is the *Fisher–Rao metric* which provides a natural gradient descent method. The analogous concepts of relative entropy for faithful states ρ and σ

$$S(\rho\|\sigma) = -\mathrm{tr}(\rho(\log(\rho) - \log(\sigma)))$$

and Fisher information is well-established in quantum information theory, too. For finite-dimensional quantum states, our aim is then to establish low-dimensional parameterized quantum Wasserstein gradient flows based on quantum Wasserstein distances. This means we aim to find a low-dimensional representation of the minimization problem in parameter space by applying quantum Wasserstein dynamics. Our study starts by pulling back the quantum Wasserstein metric to a finite-dimensional parameter manifold, using the quantum transport (Wasserstein) information matrix. This leads to a natural gradient descent method for quantum states.

We also introduce and study a quantum analog of the Schrödinger bridge problem. As we show in this article, this problem can be solved by a quantum Benamou–Brenier's formula with quantum Fisher information functional regularization.

Summary of novel results

- We introduce the quantum transport information matrix and develop the related quantum transport/Wasserstein statistical manifold. This can be viewed as the first step of quantum transport information geometry.
- We formulate the quantum transport natural gradient flow based on quantum Wasserstein statistical manifold. We apply this flow for solving the quantum statistical learning problem.
- We also formulate the quantum Schrödinger bridge problem by controlling the quantum transport natural gradient flows.
- We study the quantum Wasserstein statistical manifold for various finite-dimensional systems such as the quantum fermionic Fokker–Planck dynamics and more general finite-dimensional open quantum systems satisfying the detailed balance condition, as well as

for continuous-variable systems with positive Wigner functions such as (mixtures of) Gaussian states.
- We illustrate our results on some simple examples and also discuss how they apply to the parameter estimation problem for quantum channels.

Outline of the article In Sect. 2 we provide a brief review of classical optimal transport theory and quantum optimal transport, i.e.

- Classical optimal transport, Sect. 2.1.
- Natural gradient flow, Sect. 2.2.
- Schrödinger bridge problem, Sect. 2.3.
- Quantum optimal transport, Sect. 2.4.
- Fermionic Fokker–Planck equation, Sect. 2.5.
- Quantum Markov semigroups satisfying detailed balance (DB), Sect. 2.6.

In Sect. 3 we then introduce the quantum Wasserstein natural gradient 3.1, the Schrödinger bridge problem for finite-dimensional quantum systems in Sect. 3.2, and the same two for certain continuous-variable systems, including Gaussian systems, in Sect. 3.3. In Sect. 4 we discuss examples of our theory. This includes the transport problem for two Gaussian states and a fully explicit case of the fermionic Fokker–Planck equation. We finish our collection of examples by illustrating how the quantum transport information matrix can also be used to perform parameter estimation for quantum channels.

Notation We denote by states $|n\rangle$, for $n \in \mathbb{N}_0$, the canonical eigenbasis of the number operator $N = a^*a$ where a is the standard annihilation operator. The continuous linear operators on a normed space X are denoted by $\mathcal{L}(X)$, the space of trace-class operators on a Hilbert space $\mathcal{H}$ by $\mathrm{TC}(\mathcal{H})$. For a set Ω we denote by $\mathrm{int}(\Omega)$ its interior. The set of quantum states (positive-definite operators of unit trace) on a Hilbert space $\mathcal{H}$ is denoted by $\mathscr{D}(\mathcal{H})$. We denote the Riemannian manifold of faithful states by $\mathscr{D}_+(\mathcal{H})$. We recall that $\partial\mathscr{D}(\mathcal{H})$ are states with zero determinant and $\mathrm{int}(\mathscr{D}(\mathcal{H})) = \mathscr{D}_+(\mathcal{H})$. We also write $\{X, Y\} = XY + YX$ for the anti-commutator and $[X, Y] = XY - YX$ for the commutator. We denote the spectrum of a linear operator T by $\mathrm{Spec}(T)$.

2 Review of Classical and Quantum Optimal Transport

Our goal is to study the problem of minimizing the distance with respect to a L^2-quantum Wasserstein distance W_Q^2, for some fixed target density operator $\rho_\star$ over a parametrized manifold of states $\mathcal{P}_\theta \subset \mathscr{D}(\mathcal{H})$, i.e. we aim to identify $\mathrm{argmin}_{\rho\in\mathcal{P}_\theta} W_Q^2(\rho_\star, \rho)$.

For this purpose, we start in this section with a review of the classical framework and highlight similarities and differences that appear in the quantum setting. In addition, we will also employ the classical framework for the study of Wigner distributions in the continuous-variable setting.

2.1 Classical Optimal Transport

The optimal transport problem dates back to 1781 when Monge asked how to find for two probability measures f_0, f_1 on $\Omega \subset \mathbb{R}^n$, with finite second moment, an optimal transport plan $T : \Omega \to \Omega$ pushing f_0 to f_1 such that the transportation cost is minimized and for all

$A \subset \Omega$ measurable

$$\inf_T \int_\Omega \|x - T(x)\|^2 f_0(x)\, dx : T_* f_0 = f_1$$

For two probability measures with densities f_0, f_1 on $\Omega \subset \mathbb{R}^n$ the square of the classical L^2-Wasserstein distance is defined as

$$W_{\text{cl}}^2(f_0, f_1) := \inf_{\pi \in \Pi(f_0, f_1)} \int_{\Omega \times \Omega} |x - y|^2 d\pi(x, y) \tag{2.1}$$

where $\Pi(f_0, f_1)$ is the set of all couplings of the two measures $f_0(x)\, dx$ and $f_1(x)\, dx$.

Equivalent to (2.1), and particularly relevant for our purposes, is a dynamical formulation, given by the Benamou–Brenier formula, which states that the Wasserstein metric is given by

$$W_{\text{cl}}^2(f_0, f_1) = \inf \int_0^1 \int_\Omega |v_t(x)|^2\, d\mu_t(x) dt \tag{2.2}$$

where the infimum is taken over all pairs (μ_t, v_t) where μ_t with $\mu_0 = f_0$ and $\mu_1 = f_1$ is a curve of measures and v_t a time-dependent vector field satisfying

$$\partial_t \mu_t + \text{div}(v_t \mu_t) = 0.$$

On a bounded domain Ω the above formulation is replaced by the corresponding Neumann problem.

The dynamical formulation above is closely connected to a Riemannian structure on the Wasserstein space. To fix ideas, we consider the space of strictly positive densities

$$\mathscr{D}_+(\Omega) = \{f \in C^\infty(\Omega, (0, \infty)) : \|f\|_{L^1} = 1\}.$$

The tangent space of $\mathscr{D}_+$ is then just given by

$$T_f \mathscr{D}_+(\Omega) = \left\{\sigma \in C^\infty(\Omega) : \int_\Omega \sigma(x)\, dx = 0\right\}.$$

For any $\Phi \in C^\infty(\Omega)$ we can then set

$$V_\Phi(x) := -\text{div}(f(x) \nabla \Phi(x)) \in T_f \mathscr{D}_+(\Omega).$$

This map provides an isomorphism, at least if Ω is compact,

$$C^\infty(\Omega)/\mathbb{R} \to T_f \mathscr{D}_+(\Omega), \text{ with } [\Phi] \mapsto V_\Phi.$$

We can therefore define the L^2-Wasserstein metric tensor by introducing:

Definition 2.1 (L^2-*Wasserstein metric tensor*) We define the metric tensor $g_f : T_f \mathscr{D}_+(\Omega) \times T_f \mathscr{D}_+(\Omega) \to \mathbb{R}$ by

$$g_f(\sigma_1, \sigma_2) := \int_\Omega \langle \nabla \Phi_1(x), \nabla \Phi_2(x) \rangle f(x)\, dx, \tag{2.3}$$

with $\sigma_i = V_{\Phi_i}$.

2.2 Natural Gradient Flow

We continue with a review of the main results of [10, Sect. 3] and explain how to minimize an objective function efficiently in parameter space.

We define the statistical parameter space as a d-dimensional Riemannian manifold Θ with connection D_θ and metric tensor $\langle \xi, \eta \rangle_\theta = \xi^T G_\theta \eta$. We then take a continuous parametrization $\Theta \ni \theta \mapsto \rho(\bullet, \theta) \in \mathscr{D}_+(\Omega)$ and introduce a natural metric tensor by pulling back (2.3) on the statistical manifold

$$g_\theta : T_\theta(\Theta)^2 \to \mathbb{R}, \text{ such that } g_\theta(\xi, \eta) = g_{\rho(\bullet,\theta)}(D_\theta \rho(\xi), D_\theta \rho(\eta)) = \langle \xi, G_W(\theta)\eta \rangle$$

where $G_W(\theta) = \left(G_\theta^* \int_\Omega \partial_{\theta_i} \rho(x,\theta)(-\mathrm{div}(\rho(x,\theta)\nabla))^{-1} \partial_{\theta_j} \rho(x,\theta)\, dx\, G_\theta\right)_{ij}$.

The Wasserstein natural gradient is then for an objective function $R(\theta)$ defined by

$$\dot{\theta}(t) = -\nabla_g R(\theta(t))$$

where ∇_g is the unique gradient vector satisfying

$$g_\theta(\nabla_g R(\theta), \xi) = \langle D_\theta R(\theta), \xi \rangle_\theta.$$

In particular, we have the identification $\nabla_g R(\theta) = G_W(\theta)^{-1} G_\theta D_\theta R(\theta)$.

The *Wasserstein gradient descent* can then be numerically implemented using a standard forward Euler method

$$\theta_{(n+1)\tau} := \theta_{n\tau} - \tau G_W(\theta_{n\tau})^{-1} G_\theta D_\theta R(\rho(\bullet, \theta_{n\tau})).$$

This gradient flow method can be interpreted as an approximate solution to the minimization problem

$$\mathrm{argmin}_{\theta \in \Theta} R(\rho(\bullet, \theta)) + \frac{W_{\mathrm{cl}}^2(\rho(\bullet, \theta_{n\tau}), \rho(\bullet, \theta))^2}{2\tau}$$

which is obvious from considering the linearized expressions

$$\begin{aligned} W_{\mathrm{cl}}^2(\rho(\bullet, \theta + \Delta\theta), \rho(\bullet, \theta))^2 &= \frac{1}{2}\langle \Delta\theta, G_W(\theta)\Delta\theta \rangle + o((\Delta\theta)^2) \\ R(\rho(\bullet, \theta + \Delta\theta)) &= R(\rho(\bullet, \theta)) + \langle D_\theta R(\rho(\bullet, \theta)), \Delta\theta \rangle_\theta + o(\Delta\theta). \end{aligned} \tag{2.4}$$

2.3 Fisher Information Regularization and Schrödinger Bridge Problem

After the works of Monge and Kantorovich, Schrödinger proposed in 1931 a similar transport problem which is nowadays referred to as the *Schrödinger bridge problem* (SBP):

Given two strictly positive densities f_0, f_1 on a domain $\Omega \subset \mathbb{R}^n$, consider

$$\inf_{m,\rho} \int_0^1 \int_\Omega \frac{m(t,x)^2}{f(t,x)}\, dx\, dt, \tag{2.5}$$

where the infimum is taken over all m and f satisfying

$$\partial_t f(t,x) + \mathrm{div}(m(t,x)) = \beta \Delta f(t,x), \quad f(0,x) = f_0(x),\ f(1,x) = f_1(x) \tag{2.6}$$

with the boundary condition

$$\langle m(t,x) - \nabla f(t,x), n(x) \rangle = 0 \quad \forall x \in \partial\Omega$$

where $n(x)$ is the normal vector of the boundary. We emphasize that the difference between the SBP and the L^2-Wasserstein metric minimization (2.2) is the presence of the diffusion term $\beta\Delta$ in the PDE (2.6). A discussion of the viscosity limit $\beta \downarrow 0$ and the convergence of the solution to the SBP can be found in [22].

The minimization problem (2.5) with PDE (2.6) is, as has been shown in [7,18] equivalent to minimizing the functional

$$\inf_{m,\rho} \int_0^1 \int_\Omega \left(\frac{m(t,x)^2}{f(t,x)} + \beta^2 (\nabla \log(f(t,x)))^2 f(t,x) \right) dx\, dt + 2\beta \mathcal{D}(f_1|f_0), \tag{2.7}$$

with a constant term representing the differences of entropies $\mathcal{D}(f_1|f_0) = \int_\Omega f_1(x)\log(f_1(x)) - f_0(x)\log(f_0(x))\, dx$ and f and m are linked by the transport equation

$$\partial_t f(t,x) + \operatorname{div}(m(t,x)) = 0, \quad f(0,x) = f_0(x),\ f(1,x) = f_1(x).$$

The advantage of studying the functional (2.7) over (2.2) is in the additional positivity and strict convexity enforced by the contribution of the Fisher information

$$\mathcal{I}(f) := \int_\Omega |\nabla \log(f(x))|^2 f(x)\, dx$$

in the objective functional. Numerical aspects of this minimization problem have been thoroughly discussed in [31].

2.4 Quantum Optimal Transport

Before introducing quantum analogues of the L^2-Wasserstein distance (2.1), we first define a notion of coupling of quantum states:

For two density operators $\rho_{\text{in}}, \rho_{\text{fi}} \in \mathscr{D}(\mathcal{H})$ the set of all couplings $\Pi(\rho_{\text{in}}, \rho_{\text{fi}})$ is defined as the set of density operator valued maps that smoothly (up to endpoints) connect the two states

$$\Pi(\rho_{\text{in}}, \rho_{\text{fi}}) := \left\{ \rho \in C([0,1], \mathscr{D}(\mathcal{H})) \cap C^\infty((0,1), \mathscr{D}(\mathcal{H})); \rho(0) = \rho_{\text{in}}, \rho(1) = \rho_{\text{fi}} \right\}. \tag{2.8}$$

To give the definition of the 2-Wasserstein distance for finite-dimensional quantum systems satisfying the detailed balance equation, we employ the differential calculus introduced in [15, Def. 4.7]. This framework allows us, in particular, to reformulate the evolution of finite-dimensional open quantum systems satisfying the detailed balance condition as a gradient flow of the relative entropy $S(\rho||\sigma)$ where σ is the invariant state, with respect to the Wasserstein metric. Before discussing this in the context of open quantum systems satisfying the detailed balance condition, we introduce the necessary differential structure:

2.4.1 Differential Calculus for Quantum Systems

Let $\mathcal{A}$ be a finite-dimensional von Neumann algebra with faithful positive tracial linear functional τ and $\mathscr{D}_+(\mathcal{A})$ the set of faithful states.

Definition 2.2 A differential structure on $\mathcal{A}$ is defined as follows:

- There exists a finite index set J and for each $j \in J$ a finite-dimensional von Neumann algebra B_j with a faithful positive tracial linear functional τ_j.
- For each $j \in J$ there exists a pair (l_j, r_j) of unital $*$-homomorphisms from $\mathcal{A}$ to B_j such that

$$\tau_j(l_j(A)) = \tau_j(r_j(A)) = \tau(A).$$

- For each $j \in J$ there is $0 \neq V_j \in B_j$ and $\bar{j}$ such that $V_j^* = V_{\bar{j}}$. Moreover, for $j \in J$ and $A_1, A_2 \in \mathcal{A}$

$$\tau_j(V_j^* l_j(A_1) V_j r_j(A_2)) = \tau_j(V_j^* r_{\bar{j}}(A_1) V_j l_{\bar{j}}(A_2)).$$

- There is a faithful state $\sigma \in \mathscr{D}_+(\mathcal{A})$ such that for each $j \in J$, V_j is an eigenvector of the modular operator $M_{l_j(\sigma), r_j(\sigma)}(V_j) := l_j(\sigma) V_j r_j(\sigma)^{-1} = e^{-\omega_j} V_j$ for some $\omega_j \in \mathbb{R}$.

Then, the derivatives $\nabla_j : \mathcal{A} \to B_j$ are defined by $\nabla_j(A) := V_j r_j(A) - l_j(A) V_j$ with gradient $\nabla A := (\nabla_1 A, ..., \nabla_{|J|} A)$ and divergence operator

$$\mathrm{div}(A) = -\sum_{j \in J} \nabla_j^* A_j$$

where $\nabla_j^* := \nabla_{\bar{j}}$ with $\bar{j}$ such that $V_{\bar{j}} = V_j^*$.

2.4.2 Wasserstein Distance

Logarithmic case The quantum L^2-Wasserstein distance, for the above differentiable structure, has been defined in [15, (9.1)], by

$$W_Q^2(\rho_{\mathrm{in}}, \rho_{\mathrm{fi}}) := \inf_{\rho \in \Pi(\rho_{\mathrm{in}}, \rho_{\mathrm{fi}})} \left\{ \int_0^1 \|\rho'(t)\|_{\rho(t)}^2 dt \right\}.$$

Here, we use the norm $\|Z\|_\rho^2 = \langle Z, L_\rho(Z) \rangle_{L^2(\tau)}$. The quantum L^2-Wasserstein distance can then be expressed as a variational problem -in analogy to the classical Brenier–Benamou formula (2.2) for the classical L^2-Wasserstein distance- by

$$W_Q^2(\rho_{\mathrm{in}}, \rho_{\mathrm{fi}}) := \inf_{\rho \in \Pi(\rho_{\mathrm{in}}, \rho_{\mathrm{fi}})} \left\{ \int_0^1 \|\nabla \Phi(t)\|_{\rho(t)}^2 \, dt \right\} \tag{2.9}$$

where Φ is coupled to ρ by the following continuity equation

$$\rho'(t) + \mathrm{div}\left(L_{\rho(t)}(\nabla \Phi(t))\right) = 0.$$

The physical interpretation of the Riemannian metric g_ρ is that for two faithful states $\rho, \sigma \in \mathscr{D}_+(\mathcal{H})$, and the quantum relative entropy, defined by

$$S_\sigma(\rho) = \tau(\rho(\log(\rho) - \log(\sigma))),$$

[15, Prop. 2.7] shows that for D denoting the derivative, the gradient $(\mathrm{grad} S_\sigma)(\rho) := (-\Delta_\rho) D S_\sigma(\rho)$, where $D S_\sigma(\rho) = \log(\rho) - \log(\sigma)$, and we have

$$(\mathscr{L}^* \rho)(\rho) = -(\mathrm{grad} S_\sigma)(\rho).$$

This implies that the gradient flow of the entropy S_σ with respect to the metric g_ρ is the dynamics of the Liouville-von Neumann equation where σ is the invariant state of the dynamics defined by $\mathscr{L}^*$.

Anti-commutator case When instead of using the the *Feynman–Kubo–Mori* integral, but rather the anti-commutator

$$L_\rho^{\mathrm{ac}}(T) := \frac{1}{2}\{T, \rho\} \tag{2.10}$$

one is lead to introduce a different L^2-Wasserstein distance [8]

$$\tilde{W}_Q^2(\rho_{\mathrm{in}}, \rho_{\mathrm{fi}}) := \inf_{\rho \in \Pi(\rho_{\mathrm{in}}, \rho_{\mathrm{fi}});} \left\{ \int_0^1 \mathrm{tr}(\rho v(t)^* v(t)) dt \right\}$$

with $v^* v = \sum_{k=1}^N v_k^* v_k$, where v and ρ are coupled by

$$\rho'(t) + \mathrm{div} L_\rho^{\mathrm{ac}}(v) = 0 :, \quad \rho(0) = \rho_{\mathrm{in}}, \rho(1) = \rho_{\mathrm{fi}}.$$

In particular, the operator $L_\rho^{\mathrm{ac}}(T)$ is invertible for $\rho > 0$ by standard results on the solvability of Lyapunov equations which imply that the inverse is explicitly given as

$$(L_\rho^{\mathrm{ac}})^{-1}(S) = -\int_0^\infty e^{-\rho s} S e^{-\rho s}\, ds.$$

2.5 Fermionic Fokker–Planck Equation

Due to its analogy to classical probability theory and classical gradient flows, we start by discussing the *quantum fermionic Fokker–Planck equation*. Instead of just stating it within the abstract differential calculus introduced in the previous section, we will provide full details to fix ideas.

The quantum fermionic Fokker–Planck equation, is the canonical gradient flow associated with the quantum Wasserstein metric and corresponds to the classical Fokker–Planck equation[1]

$$\frac{\partial \rho(x,t)}{\partial t} = \mathrm{div}(\rho(t,x)\nabla V(x)) + \beta \Delta \rho(t,x), \quad \rho(0,x) = \rho_0(x) \text{ for } x \in \mathbb{R}^d.$$

Under suitable growth conditions on V this equation has a unique invariant measure $d\mu(x) \propto e^{-\beta V(x)} dx$. Carlen and Maas introduced in [12] a Riemannian metric on density operators which extends the classical L^2-Wasserstein metric to the quantum setting and with respect to which the quantum evolution of the fermionic Fokker–Planck equation is a gradient flow. We will explain in section how to use this metric to define a natural gradient flow for parametric models of density operators.

2.5.1 Clifford Algebra

Let $\mathfrak{C}$ be the Clifford algebra on $\mathbb{R}^n$ generated by n self-adjoint operators Q_j, $j = 1, .., n$ satisfying the canonical anti-commutation relations $\{Q_i, Q_j\} = 2\delta_{ij}$. The operators Q_j are also called the *fermionic degrees of freedom*. Moreover, $\mathfrak{C}$ becomes a 2^n-dimensional Hilbert space $\mathcal{H} \sim L^2(\tau)$ with inner product $\langle A, B\rangle_{L^2(\tau)} := \tau(A^* B)$, where we introduce the normalized trace $\tau(A) = 2^{-n}\,\mathrm{tr}_{\mathbb{C}^{2^n}}(A)$.

The density operators $\mathscr{D}(\mathcal{H})$ in this setting is the closed convex set of positive operators $\rho \in \mathfrak{C}$ of unit normalized trace.

[1] Especially in statistical physics, the name *Fokker–Planck equation* is usually reserved for another equation acting on phase-space variables and the equation considered here is called the (Kramers)–Smoluchowski equation.

We can explicitly construct matrices Q_j solely from Pauli matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_y := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \text{ and } \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.11}$$

One realization of the fermionic operators Q_j, is by defining them as $Q_j := \otimes_{i=1}^n X_i$ where

$$X_i = \begin{cases} \sigma_z & \text{for } i < j, \\ \sigma_x & \text{for } i = j, \text{ and} \\ \mathrm{id}_{\mathbb{C}^2} & \text{for } i > j. \end{cases}$$

The grading operator $\Gamma : \mathfrak{C} \to \mathfrak{C}$ is the linear operator defined, for $\alpha \in \{0,1\}^n$, by $\Gamma(Q^\alpha) := (-1)^{|\alpha|} Q^\alpha$ where $Q^\alpha := \prod_{i=1}^n Q_i^{\alpha_i}$. The index set $\alpha \in \{0,1\}^n$ is called the *fermionic multi-index set*. The 2^n matrices Q^α for $\alpha \in \{0,1\}^n$ form an orthonormal system spanning $\mathfrak{C}$ which satisfies $\tau(Q^\alpha) = \delta_{0|\alpha|}$.

For two density operators $\rho_1, \rho_2 \in \mathscr{D}(\mathcal{H})$ we define the *Feynman–Kubo–Mori* operator $L_{(\rho_1,\rho_2)} : \mathcal{L}(\mathfrak{C}) \to \mathrm{TC}(\mathfrak{C})$ by

$$L_{(\rho_1,\rho_2)}(T) := \int_0^1 \rho_1^{1-s} T \rho_2^s \, ds \tag{2.12}$$

which is a contraction map into the set of trace-class operators, as Hölder's inequality shows

$$\|L_{(\rho_1,\rho_2)}(T)\|_1 \le \int_0^1 \|\rho_1^{1-s} T \rho_2^s\|_1 \, ds \le \int_0^1 \|\rho_1^{1-s}\|_{(1-s)^{-1}} \|T\|_\infty \|\rho_2^s\|_{s^{-1}} \, ds \le \|T\|_\infty.$$

Under the stronger assumption $\rho_1, \rho_2 \in \mathscr{D}_+(\mathcal{H})$, the operator $L_{(\rho_1,\rho_2)}$ becomes invertible and its inverse is given by [12, Theo. 3.4]

$$L^{-1}_{(\rho_1,\rho_2)}(T) = \int_0^\infty (\rho_1 + t)^{-1} T (\rho_2 + t)^{-1} \, dt.$$

In particular, we will just write $L_\rho := L_{(\Gamma(\rho),\rho)}$ in the sequel.

The fermionic Dirichlet form on $\mathfrak{C}$ is defined by

$$\mathcal{F}(A, A) = \tau((\nabla A)^* \nabla A) = \sum_{j=1}^n \tau((\nabla_j A)^* \nabla_j A)$$

with derivatives

$$\nabla_j(A) = \frac{1}{2}\left(Q_j A - \Gamma(A) Q_j\right) \in \mathfrak{C}, \text{ for } j \in \{1, \ldots, n\} \text{ and } A \in \mathfrak{C}. \tag{2.13}$$

The gradient $\nabla : \mathfrak{C} \to \mathfrak{C}^n$ is then defined as $\nabla(A) := (\nabla_1(A), ..., \nabla_n(A)) \in \mathfrak{C}^n$ with nullspace $\ker(\nabla) = \mathrm{span}(\mathrm{id})$. The $L^2(\tau)$-adjoint of derivatives ∇_j is just given by

$$\nabla_j^*(A) = \frac{1}{2}\left(Q_j A + \Gamma(A) Q_j\right).$$

The divergence operator is defined, for $A = (A_j)_j \in \mathfrak{C}^n$ by $\mathrm{div}(A) = -\sum_{j=1}^n \nabla_j^*(A_j)$. We define the fermionic number operator $\mathcal{N}$ as the self-adjoint operator associated to the Dirichlet form $\mathcal{F}(B, A) =: \langle B, \mathcal{N} A\rangle_{L^2(\tau)}$ where $\mathcal{N} A = -\mathrm{div}(\nabla A)$ for all $A \in \mathfrak{C}$ and $\ker(\mathcal{N}) = \mathrm{id}$. The dynamical semigroup generated by $-\mathcal{N}$ is the *quantum fermionic Fokker–Planck semigroup* defined by $P_t = e^{-t\mathcal{N}}$ which relaxes exponentially fast to its unique invariant state, the completely mixed state. In particular, $\mathcal{N}$ is the generator of an ergodic

Quantum Markov semigroup satisfying the detailed balance condition with respect to the completely mixed state.

This model can be casted in the differential calculus introduced in Sect. 2.4.1 by setting $\mathcal{A} := \mathcal{B}_j := \mathfrak{C}^n$, $V_j := Q_j$, $\omega_j := 0$, $l_j := \Gamma$ and $r_j = \mathrm{id}$ with derivatives as defined in (2.13) and a generator $\mathscr{L}A = 2\sum_{j=1}^n (Q_j A Q_j - A) = -4\mathcal{N}$.

2.6 Quantum Markov Semigroups with Detailed Balance Condition

In the rest of this section, we illustrate the ideas using the differential calculus in Sect. 2.4.1 in the case of Quantum Markov semigroups (P_t) with Lindblad generator $\mathscr{L}$, in the Heisenberg picture, acting on a finite-dimensional C^*-algebra $\mathcal{A}$ satisfying the detailed balance condition (DBC). This means, that for all times $t > 0$ the operator P_t is self-adjoint with respect to the inner product $\langle X, Y\rangle_{1,\sigma} := \tau(X^*\sigma Y)$ for some state σ. In particular, the DBC implies that σ is the unique state such that $P_t^*(\sigma) = \sigma$ for all times $t > 0$. Other possible applications of the differential calculus in Sect. 2.4.1 and thus also of the parameter estimation techniques studied in this paper are discussed in [15, Sect. 5] and include popular quantum channels such as the depolarizing channel.

The generators $\mathscr{L}$ of the quantum Markov semigroups in Heisenberg representation are characterized by [15, Theo 2.4]

$$\mathscr{L} = \sum_{j\in J} e^{-\omega_j/2}\mathscr{L}_j \text{ and } \mathscr{L}_j(A) = V_j^*[A, V_j] + [V_j^*, A]V_j \tag{2.14}$$

with J a finite set and a family of operators $(V_j)_{j\in J}$ closed under taking adjoints, as well as real numbers ω_j such that the modulation operator $M_\sigma(A) := M_{\sigma,\sigma}(A) := \sigma A \sigma^{-1}$ satisfies

$$M_\sigma(V_j) = e^{-\omega_j} V_j \text{ and } \omega_{\bar{j}} = -\omega_j.$$

We then define $\mathcal{A} = B_j = \mathcal{L}(\mathcal{H})$ where $\mathcal{H}$ is a finite-dimensional Hilbert space, write $\mathcal{B} := \prod_j B_j$, and set $l_j = r_j = \mathrm{id}_{\mathcal{A}}$. The partial derivatives are then just given by $\nabla_j A = [V_j, A]$ and $\nabla_j^* := \nabla_{\bar{j}}$ where $\bar{j}$ is such that $V_j^* = V_{\bar{j}}$. The gradient vector is thus just $\nabla = (\nabla_1, \ldots, \nabla_{|J|})$. It follows from [15, Prop. 2.5] that the Lindblad generator induces a Dirichlet form with respect to the Kubo–Martin–Schwinger inner product $\langle A, B\rangle_{\mathrm{KMS}} := \tau(X^*Y\sigma)$, i.e.

$$\langle \nabla A, \nabla B\rangle = -\langle A, \mathscr{L}B\rangle_{L^2_{\mathrm{KMS}}(\sigma)} \text{ for all } A, B \in \mathcal{A}.$$

We then define the operator

$$\widehat{\rho_j} = \int_0^1 (e^{\omega_j/2} l(\rho))^{1-s} \otimes (e^{-\omega_j/2} r(\rho))^s \, ds \in \mathcal{A}\otimes\mathcal{A}$$

with inverse operator

$$\check{\rho_j} = \int_0^\infty (t + e^{\omega_j/2} l(\rho))^{-1} \otimes (t + e^{-\omega_j/2} r(\rho))^{-1} \, dt.$$

In terms of a contraction operator # that is uniquely defined as the linear extension of the map $(A\otimes B)\#C := ACB$ for $A, B, C \in \mathcal{A}$ and *Feynman–Kubo–Mori* operator

$$L_\rho(C) := \widehat{\rho_j}\#C \tag{2.15}$$

we may then introduce a positive-definite operator $-\Delta_\rho$ on $L^2(\mathcal{A}, \tau)$

$$-\Delta_\rho(A) := \sum_{j\in J} \nabla_j^*(L_\rho(\nabla_j A)). \tag{2.16}$$

This way, the L^2-quantum Wasserstein metric becomes

$$W_Q^2(\rho_{\text{in}}, \rho_{\text{fi}}) := \inf_{\rho\in\Pi(\rho_{\text{in}},\rho_{\text{fi}})} \left\{ \int_0^1 \langle \Phi(t), -\Delta_\rho(\Phi(t))\rangle_\tau \, dt \right\} \tag{2.17}$$

where Φ is coupled to ρ by the following continuity equation

$$\rho'(t) = \Delta_{\rho(t)}\Phi(t).$$

3 Quantum Natural Gradient and Open Quantum Systems

In the following we shall impose the following condition on generators of finite-dimensional open quantum systems we consider:

Assumption 1 We assume that $\mathscr{L}$ is ergodic, i.e. $\ker(\mathscr{L}) = \text{span}\{\text{id}\}$ satisfying the detailed balance condition with invariant state σ.

3.1 Gradient Flow for Finite-Dimensional OQSs with DBC

By the ergodicity assumption, we are able to pull back the metric from the state space to the parameter space. In particular, the above assumptions are satisfied for the fermionic Fokker–Planck equation with the completely mixed state as the unique invariant state.

The statistical parameter space is as in the classical setting defined as a d-dimensional Riemannian manifold Θ with connection D_θ and metric tensor $\langle \xi, \eta\rangle_\theta = \xi^T G_\theta \eta$. We then take a continuous parametrization $\Theta \ni \theta \mapsto \rho(\theta) \in \mathscr{D}_+(\mathcal{A})$ of density operators.

We then define a norm

$$\|Z\|_\rho^2 = \left\langle Z, L_\rho(Z)\right\rangle_{L^2(\tau)}$$

where L_ρ has been defined in (2.12) for the fermionic Fokker–Planck equation and in (2.15) for general open quantum systems satisfying the DBC. In addition, we allow for L_ρ the anti-commutation operator defined in (2.10).

The associated metric tensor on $\mathscr{D}_+(\mathcal{H})$ is given by

$$g_\rho : (T_\rho\mathscr{D}_+(\mathcal{H}))^2 \to \mathbb{R}, \quad g_\rho(X, X) := \left\langle \nabla\Phi_X, L_\rho(\nabla\Phi_X)\right\rangle_{L^2(\tau)}$$

where $T_\rho\mathscr{D}_+(\mathcal{H})$ is the tangent space at ρ and $\nabla\Phi_X$ is the unique gradient field cf. [12, Theo 3.17] and [15, Lem. 7.5] satisfying

$$X = -\text{div}\left(L_\rho(\nabla\Phi_X)\right).$$

In case of L_ρ being the anti-commutator, the gradient field $\nabla\Phi_X$ can be found by solving the Lyapunov equation [8, (21)]

$$\nabla(\text{divgrad}|_{\text{span(id)}^\perp})^{-1}X = L_\rho(\nabla\Phi_X) \in \mathcal{L}(\mathcal{H}^n).$$

In particular, there exists a unique gradient $\nabla\Phi_\xi$ such that

$$\langle D_\theta\rho(\theta), \xi\rangle_\theta = -\text{div}(L_{\rho(\theta)}(\nabla\Phi_\xi)).$$

Hence, we conclude that for $\xi, \eta \in T_\theta\Theta$ there are *score functions* Φ_ξ and Φ_η[2] such that we can define the pullback metric on the parameter space

$$\begin{aligned} g_\theta(\xi,\eta) &:= g_{\rho(\theta)}(\langle D_\theta\rho(\theta),\xi\rangle_\theta, \langle D_\theta\rho(\theta),\eta\rangle_\theta) \\ &= \left\langle \nabla\Phi_\xi, L_{\rho(\theta)}(\nabla\Phi_\eta)\right\rangle_{L^2(\tau)} \\ &= -\left\langle \Phi_\xi, \mathrm{div}(L_{\rho(\theta)}(\nabla\Phi_\eta))\right\rangle_{L^2(\tau)} \\ &= \left\langle \Phi_\xi, \langle D_\theta\rho(\theta),\eta\rangle_\theta\right\rangle_{L^2(\tau)}. \end{aligned} \tag{3.1}$$

We then define the operator $-\Delta_\theta f := -\mathrm{div}(L_{\rho(\theta)}(\nabla f))$. This operator is self-adjoint with respect to $\langle\bullet,\bullet\rangle_{L^2(\tau)}$ and positive-definite with only span{id} in its nullspace by ergodicity. Using that $\langle D_\theta\rho(\theta),\eta\rangle_\theta \in \ker(\Delta_\theta)^\perp$, this implies that

$$g_\theta(\xi,\eta) := \tau\left(\langle D_\theta\rho(\theta),\xi\rangle_\theta(-\Delta_\theta)|^{-1}_{\mathrm{span}\{\mathrm{id}\}^\perp}\langle D_\theta\rho(\theta),\eta\rangle_\theta\right). \tag{3.2}$$

We can rewrite this line as a bilinear form by using the matrix G_θ, introduced above, associated with the metric on the parameter space. We can thus define the positive definite *Wasserstein information matrix*

$$G_W(\theta) = \tau\left(\widehat{e_i}^T G_\theta^T D_\theta\rho(\theta)(-\Delta_\theta)|^{-1}_{\mathrm{span}\{\mathrm{id}\}} D_\theta\rho(\theta) G_\theta \widehat{e_j}\right)_{i,j} \in \mathbb{R}^{d\times d}. \tag{3.3}$$

Thus, it follows that the metric tensor of the pullback metric on the statistical manifold is of the simple form

$$g_\theta(\xi,\eta) = \langle\xi, G_W(\theta)\eta\rangle \tag{3.4}$$

and as in Sect. 2.2 the natural Wasserstein gradient becomes for an objective function $R(\theta)$ defined by

$$\dot\theta(t) = -\nabla_g R(\theta(t)) \tag{3.5}$$

with ∇_g uniquely defined by

$$g_\theta(\nabla_g R(\theta),\xi) = \langle D_\theta R(\theta),\xi\rangle_\theta \quad \forall\xi \in T_\theta\Theta$$

such that $\nabla_g R(\theta) = G_W(\theta)^{-1} G_\theta D_\theta R(\theta)$. This is illustrated in Sect. 4.2.2 for R being the von Neumann entropy.

The gradient descent method in parameter space naturally corresponds to a gradient descent method on the parametrized manifold of states:

Proposition 3.1 *Consider an immersion $\Theta \ni \theta \mapsto \rho(\theta) \in \mathscr{D}(\mathcal{H})$ and an objective function $\mathcal{R}$ on the set of states. We can then define an objective function $R(\theta) = \mathcal{R}(\rho(\theta))$ and the gradient evolution*

$$\dot\theta(t) = -\nabla_g R(\theta),$$

induces the gradient evolution

$$\rho'(t) = -\mathrm{grad}\mathcal{R}(\rho(t))$$

on the parametrized manifold of states where $\rho(t) = \rho(\theta(t))$ and $\mathrm{grad}(\mathcal{R}(\rho(t_0))) = \langle D_\theta\rho(\theta), \nabla_g R(\theta_{t_0})\rangle_\theta$.

[2] Score functions are only defined up to elements in $\ker(\nabla)$.

Proof We always have that $\frac{d}{dt}\rho(\theta(t)) = (\rho_\theta)_*\dot{\theta}(t) = -(\rho_\theta)_*\nabla_g R(\theta(t))$. Thus it suffices to show that $(\rho_\theta)_*\nabla_g R(\theta(t)) = \mathrm{grad}\mathcal{R}(\rho(t))$.

Fix a curve $(\vartheta_\tau)_\tau$ passing through θ_{t_0} at $\tau = 0$, then it follows that

$$\begin{aligned}\frac{d}{d\tau}\big|_{\tau=0} R(\vartheta_\tau) &= g_{\theta_{t_0}}(\nabla_g R(\theta_{t_0}), \dot{\vartheta}_0) \\ &= g_{\rho(t_0)}(\langle D_\theta\rho(\theta_{t_0}), \nabla_g R(\theta_{t_0})\rangle_\theta, \langle D_\theta\rho(\theta_{t_0}), \dot{\theta}_{t_0})\rangle_\theta). \end{aligned} \tag{3.6}$$

On the other hand, we also see that

$$\begin{aligned}\frac{d}{d\tau}\big|_{\tau=0} R(\vartheta_\tau) = \frac{d}{d\tau}\big|_{\tau=0}\mathcal{R}(\rho(\tau)) &= g_{\rho(t_0)}(\mathrm{grad}(\mathcal{R}(\rho(t_0))), \dot{\rho}(\theta(0))) \\ &= g_{\rho(t_0)}(\mathrm{grad}(\mathcal{R}(\rho(t_0))), \langle D_\theta\rho(t_0), \dot{\theta}(0)\rangle_\theta). \end{aligned} \tag{3.7}$$

This shows that $\langle D_\theta\rho(\theta), \nabla_g R(\theta_{t_0})\rangle_\theta = \mathrm{grad}(\mathcal{R}(\rho(t_0)))$. □

Using (3.1) and (3.4), we thus find that the geodesics on the parameter manifold (Θ, g_θ) minimize again the square geodesic distance

$$W_Q^2(\rho(\bullet,\theta^0), \rho(\bullet,\theta^1)) = \inf_{\substack{\theta\in C^1(0,1)\cap C[0,1]\\ \theta(0)=\theta^0,\theta(1)=\theta^1}} \left\{\int_0^1 \langle\dot{\theta}(t), G_W(\theta(t))\dot{\theta}(t)\rangle\, dt\right\}. \tag{3.8}$$

The geodesics to the above Wasserstein distance are given as solutions to the following Hamiltonian system

$$\begin{aligned}&\dot{\theta} - G_W(\theta)^{-1}P = 0 \text{ and} \\ &\dot{P} + \frac{1}{2}P^T\partial_\theta G_W(\theta)^{-1}P = 0. \end{aligned} \tag{3.9}$$

Indeed, for the Lagrangian $\mathcal{L}(\theta(t), \dot{\theta}(t)) = \langle\dot{\theta}(t), G_W(\theta(t))\dot{\theta}(t)\rangle$, the associated momentum variable is $P(t) = G_W(\theta(t))\dot{\theta}(t)$ with Hamilton function $H(P(t), \theta(t)) = \frac{1}{2}\langle P(t), G_W(\theta(t))P(t)\rangle$. The system (3.9) are then precisely Hamilton's equations.

On the other hand, the geodesic equations in $\mathscr{D}_+(\mathcal{H})$ with respect to the quantum Wasserstein metric for the fermionic Fokker–Planck equation are given by [12, Theo. 5.3]

$$\begin{aligned}&\rho'(t) + \mathrm{div}(L_{\rho(t)}\nabla\Phi(t)) = 0 \\ &\Phi'(t) + \frac{1}{2}\rho(t)\flat(\nabla\Phi(t), \nabla\Phi(t)) = 0 \end{aligned} \tag{3.10}$$

where we define for $\rho \in \mathscr{D}_+(\mathcal{H})$ and $X, Y \in \mathfrak{C}^n$ the map

$$\rho\flat(X,Y) = \int_0^1\int_0^1\int_0^\alpha \frac{2\rho^{\alpha-\beta}}{(1-s)+s\rho}X^*\Gamma(\rho)^{1-\alpha}Y\frac{\rho^\beta}{(1-s)+s\rho}\, d\beta\, d\alpha\, ds. \tag{3.11}$$

The advantage of (3.9) over (3.10) lies in the low-dimensionality of the parameter space which turns (3.9) into an equation in a much lower dimensional space than the system in (3.10), in general.

3.2 Schrödinger Bridge Problem for Finite-Dimensional OQSs with DBC

We may now introduce a generalization of the quantum Brenier–Benamou formula in (2.9), to study a quantum version of the Schrödinger bridge problem, by adding a Fisher information regularizer to the dynamics. For this derivation, we shall restrict us to the scenario that the

operator L_ρ is the *Feynman–Kubo–Mori* operator as in this case, one obtains direct links to quantum entropies and quantum dynamics.

The computational advantage of the Fisher information regularization are two-fold. Firstly, it induces additional convexity to the minimization problem. Secondly, it additionally forces the density operator to remain strictly positive.

Definition 3.2 The quantum Schrödinger bridge problem (SBP) is the minimization problem for two quantum states $\rho_{\text{in}}, \rho_{\text{fi}} \in \mathscr{D}_+(\mathcal{H})$

$$\mathcal{S}(\rho_{\text{in}}, \rho_{\text{fi}}) := \inf_{\rho \in \Pi(\rho_{\text{in}}, \rho_{\text{fi}})} \inf_m \int_0^1 \|m\|^2_{\rho(t)^{-1}}\, dt. \tag{3.12}$$

where we use the inner product

$$\langle X, Y\rangle_{\rho(t)^{-1}} := \langle X, L_\rho^{-1}(Y)\rangle_{L^2(\tau)}.$$

Here m is connected to $\rho(t)$ by an inhomogeneous heat equation

$$\rho'(t) + \operatorname{div}(m(t)) = \beta T \rho(t)$$

for some fixed parameter $\beta \geq 0$ where $T = \mathscr{L}^*$ for OQS satisfying the DBC and $T = -\mathcal{N}$ in the case of the fermionic Fokker–Planck equation.

As for (3.12), in the case $\beta = 0$, the SBP reduces to the minimization of the quantum L^2-Wasserstein metric in (2.9). We now introduce the *Fisher information matrix* $I(\rho) := \|\nabla(\log(\rho) - \log(\sigma))\|^2_\rho$. We can then express the optimal transport distance problem (3.12) as an equivalent dynamical problem with Fisher information regularization:

Theorem 1 *The Schrödinger bridge problem* (3.12) *is equivalent to the following optimization problem*

$$\mathcal{S}(\rho_{\text{in}}, \rho_{\text{fi}}) := \inf_{\rho \in \Pi(\rho_{\text{in}}, \rho_{\text{fi}})} \inf_M \int_0^1 \|M(t)\|^2_{\rho(t)^{-1}} + \beta^2 I(\rho(t))\, dt + 2\beta(S_\sigma(\rho_{\text{fi}}) - S_\sigma(\rho_{\text{in}}))$$

where M satisfies the transport equation

$$\rho'(t) + \operatorname{div} M(t) = 0. \tag{3.13}$$

Proof We start by defining

$$M(t) := m(t) - \beta L_{\rho(t)}(\nabla(\log(\rho(t)) - \log(\sigma))),$$

which turns the inhomogeneous heat equation into a simple transport equation

$$\rho'(t) + \operatorname{div} M(t) = 0 \tag{3.14}$$

as

$$\mathscr{L}^* = \operatorname{div}(L_{\rho(t)}(\nabla(\log(\rho(t)) - \log(\sigma)))),$$

cf. the proof of [15, Prop. 2.7]. In case of the fermionic Fokker–Planck equation we also record that the quantum analog of the classical identity $\nabla f(x) = f(x)\nabla \log f(x)$ in the quantum setting becomes the identity [12, Lemma 3.1]

$$\nabla_i \rho = L_\rho(\nabla_i \log(\rho)).$$

Thus, we have that

$$\|m(t)\|^2_{\rho(t)^{-1}} = \left\|M(t) + \beta L_{\rho(t)}(\nabla(\log(\rho(t)) - \log(\sigma)))\right\|^2_{\rho(t)^{-1}}$$

$$= \|M(t)\|^2_{\rho(t)^{-1}} + 2\beta\langle M(t), \nabla(\log(\rho(t)) - \log(\sigma))\rangle_{L^2(\tau)} + \beta^2\|\nabla(\log(\rho(t)) - \log(\sigma))\|^2_{\rho(t)}. \tag{3.15}$$

The middle term in (3.15) is constant, and satisfies in terms of the relative von Neumann entropy $S_\sigma(\rho) = \tau(\rho(\log(\rho) - \log(\sigma)))$

$$\begin{aligned}\int_0^1 \langle M(t), \nabla(\log(\rho(t)) - \log(\sigma))\rangle_{L^2(\tau)}\, dt &= -\int_0^1 \tau(\operatorname{div}(M(t))(\log(\rho(t)) - \log(\sigma)))\, dt \\ &= \tau\left(\int_0^1 \rho'(t)(\log(\rho(t)) - \log(\sigma))\, dt\right) \\ &= S_\sigma(\rho_{\mathrm{fi}}) - S_\sigma(\rho_{\mathrm{in}}) = \text{const}\end{aligned}$$

where we integrated by parts to obtain the last line. □

3.3 Continuous-Variable Systems

As in the theory of classical probability theory, there exists a close analogue of quantum *Gaussian states* $\mathcal{G}(\mathcal{H}_m)$ on $\mathcal{H}_m := L^2(\mathbb{R}^m)$ defined as follows (cf. [5] and references therein for more details):

Gaussian states are states $\rho \in \mathscr{D}(\mathcal{H}_m)$ such that their characteristic function $\chi_\rho : \mathbb{C}^m \to \mathbb{C}$

$$\chi_\rho(z) := \operatorname{tr}(\rho D(z)) \tag{3.16}$$

is the characteristic function of a Gaussian random variable over $\mathbb{C}^m$, i.e. $\chi(\xi) = \exp\left(-\frac{1}{4}\langle\xi, \gamma\xi\rangle + i\langle d, \xi\rangle\right)$ where $\gamma > 0$ is a positive definite matrix satisfying $\gamma + i\nu \geq 0$, for $\nu := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^{\oplus_{i=1}^m}$, and $d \in \mathbb{R}^{2m}$. Here, $D(z)$ is the displacement operator

$$D(z) = \exp\left(\sum_{j=1}^m (z_j a_j^* - \bar{z}_j a_j)\right).$$

Conversely, the density operator $\rho \in \mathscr{D}(\mathcal{H}_m)$ can be recovered from its characteristic function by

$$\rho = \int_{\mathbb{C}^m} \chi_\rho(z) D(-z) \frac{dz}{\pi^m}.$$

We can associate a canonical random variable to any Gaussian state in terms of their Wigner function

$$P_\rho(z) := \int_{\mathbb{C}^m} \chi_\rho(w) e^{z^T w^* - z^\dagger w} \frac{dw}{\pi^{2m}} \geq 0 \tag{3.17}$$

which is of unit L^1 norm and a Gaussian distribution on $\mathbb{R}^{2m}$ as well.

A particularly simple and relevant example of a Gaussian state are thermal states with mean photon number $N \in [0, \infty)$

$$\rho_N := \frac{1}{N+1} \sum_{n=0}^\infty \left(\frac{N}{N+1}\right)^n |n\rangle\langle n|$$

as their characteristic functions and Wigner distributions

$$\chi_{\rho_N}(z) := e^{-(2N+1)|z|^2/2} \text{ and } P_{\rho_N}(z) := \frac{2}{\pi(2N+1)} e^{-\frac{2}{2N+1}|z|^2}. \tag{3.18}$$

are centered and uncorrelated.

Thermal states have the special property that they are the maximum entropy states for a fixed average energy

$$\rho_N = \text{argmax}_{\rho;\text{tr}(\rho a^* a)\leq N} - \text{tr}(\rho \log(\rho)).$$

We finally mention that although Wigner distributions functions are positive as operators on $L^2(\mathbb{R}^{2m})$, they are not pointwise positive in general and therefore also not always genuine probability distributions (cf.the Wigner distribution function associated to $|1\rangle\langle 1|$).

In addition, the Wigner distribution function of a state ρ satisfies the energy identity

$$\int_{\mathbb{R}^{2n}} |z|^2 P_\rho(z)\, dz = \int_{\mathbb{R}^{2n}} |z|^2 \rho(z)\, dz = \text{tr}(\rho x^2) + \text{tr}(\rho p^2) = \text{tr}((2a^* a + 1)\rho)$$

where x and p are the position and momentum operator.

Thus, the classical L^2-Wasserstein distance, corresponds in this formalism to an energy penalization and we define the optimal transport functional with phase-space variable square penalization

$$\inf_{m,\rho} \int_0^1 \int_{\mathbb{R}^{2n}} \frac{|m(t,z)|^2}{\rho(t,z)}\, dz\, dt$$

where ρ satisfies the Fokker–Planck equation

$$\partial_t \rho(t,z) + \text{div}(m(t,z)) = \beta \Delta \rho(t,z), \quad \rho(0,z) = \rho_0(z),\ \rho(1,z) = \rho_1(z)$$

with parameter $\beta \geq 0$, where $\beta = 0$ corresponds to the optimal transport in L^2-Wasserstein distance [10] and $\beta > 0$ to the Schrödinger bridge problem [31].

Proposition 3.3 (Separability) *Let $\rho_\theta^{(i)}$ be a family of Gaussian states on Hilbert spaces $L^2(\mathbb{R}^{2n(i)})$, and $\rho_\theta := \bigotimes_{i=1}^N \rho_\theta^{(i)}$, then the Wasserstein information matrix satisfies*

$$G_W(\rho) = \sum_{i=1}^N G_W(\rho^{(i)}).$$

Proof It follows directly from (3.16) that the characteristic function of a tensor product is the product of the individual characteristic functions. Using the Fourier transform and (3.17), this immediately translates into the Wigner functions being a product of Wigner functions (3.17). The result then follows from [25, Prop. 5]. □

4 Examples

In this section, we demonstrate the quantum transport information matrix and its related gradient and Hamiltonian flows in some well-known probability models.

4.1 Examples for the Quantum Wigner Distribution

4.1.1 Gaussian Mixture Model

For Gaussian states ρ_i we consider the Gaussian Wigner probability distributions P_{ρ_i} associated to them.

Let $X_i \sim \mathcal{N}(\mu_i, \Sigma_i)$ be normal random variables, then it follows that

$$\mathbb{E}(X_i) = m_i, \mathrm{Var}(X_i) = \mathbb{E}(X_i X_i^*) - \mathbb{E}(X_i)\mathbb{E}(X_i)^* = \Sigma.$$

Let $X = \sum_{i=1}^N \lambda_i X_i$ be a Gaussian mixture with $\lambda_i \geq 0$ summing up to one, then clearly $\mu_X := \mathbb{E}(X) = \sum_{i=1}^N \lambda_i \mu_i$ and also for the second moment $m_{X_i} := \mathbb{E}(X_i X_i^*)$ we find

$$m_X := E_X(xx^*) = \sum_{i=1}^N \lambda_i \mu_{X_i}.$$

Thus, the covariance matrix is given by

$$\mathrm{Var}(X) = \sum_{i=1}^N \lambda_i \Sigma_i + \sum_{i=1}^N \lambda_i \mu_i \mu_i^* - \mathbb{E}(X)\mathbb{E}(X)^*$$

where $\sum_{i=1}^N \lambda_i \mu_i \mu_i^* - \mathbb{E}(X)\mathbb{E}(X)^* \geq 0$ by Jensen's inequality. Thus, since the variance of a mixture is increasing, the condition $\Sigma_i + i\nu \geq 0$ is satisfied for the extremal states and clearly the state associated with the mixture X is

$$\rho = \sum_{i=1}^N \lambda_i \rho_i.$$

To parametrize multivariate Gaussian distributions $\mathcal{N}(\mu, \Sigma)$ that are Wigner functions of Gaussian states, it is natural to consider the parameter space $\theta = (\mu, \Sigma) \in \Theta := \mathbb{R}^{2m} \times \{\gamma \in \mathbb{R}^{2m\times 2m}; \gamma > 0 \text{ and } \gamma + i\nu > 0\}$. The Wasserstein metric tensor for the multivariate Gaussian model is

$$g_\theta(\xi, \eta) = \langle \mu_\xi, \mu_\eta \rangle + \mathrm{tr}(S_\xi \Sigma S_\eta)$$

for $\xi = (\mu_\xi, \Sigma_\xi)$ and $\eta = (\mu_\eta, \Sigma_\eta)$ and S_ξ and S_η solving the Lyapunov equations

$$\Sigma_\xi = \{S_\xi, \Sigma\} \text{ and } \Sigma_\eta = \{S_\eta, \Sigma\}.$$

In fact, for $Q = Q^*$, we can define the map $L_\Sigma(Q) := \int_0^\infty e^{-\Sigma t} Q e^{-\Sigma t}\, dt$, solving Lyapunov equation $Q = \{L_\Sigma(Q), \Sigma\}$, then $L_\Sigma(\Sigma_\eta) = S_\eta$ and $L_\Sigma(\Sigma_\xi) = S_\xi$.

This way, setting $G_W := \mathbb{1}_{\mathbb{R}^{2n}} \oplus (L_{\Sigma_\eta} \Sigma L_{\Sigma_\eta})$ we find that

$$g_\theta(\xi, \eta) = \langle (\mu_\xi, \Sigma_\xi), G_W(\mu_\eta, \Sigma_\eta) \rangle.$$

Example 1 (Gaussian states; Numerical solution) We consider two Gaussian states with associated Wigner distributions and parameters $\theta^0 := (\Sigma_0, \mu_0)$ and $\theta^1 := (\Sigma_1, \mu_1)$

$$\Sigma_0 := \begin{pmatrix} 26 & 1 \\ 1 & 1 \end{pmatrix} \text{ and } \Sigma_1 := \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}. \tag{4.1}$$

which are easily shown to satisfy $\Sigma + i\nu \geq 0$ and expectation values

$$\mu_0 := (-1, -1)^t, \text{ and } \mu_1 := (2, 7)^t. \tag{4.2}$$

For Wigner functions

$$W(\Sigma, \mu)(x, \xi) = \frac{e^{-\frac{1}{2}\langle (x,\xi)^t - \mu, \Sigma^{-1}((x,\xi)^t - \mu)\rangle}}{2\pi\sqrt{|\Sigma|}}$$

we then want to analyze the optimal transport plan between $W(\theta^0)$ and $W(\theta^1)$ (Figs. 1, 2).

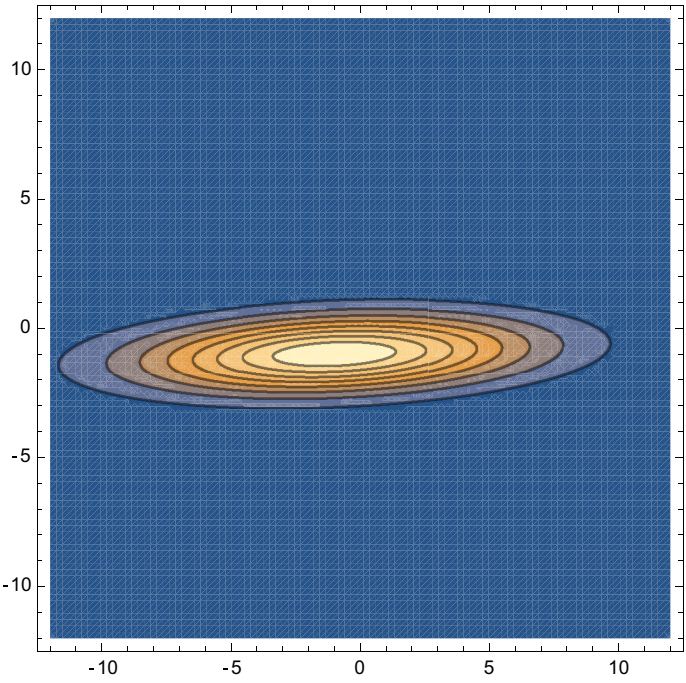

Fig. 1 The Wigner function $W(\Sigma_0, \mu_0)$

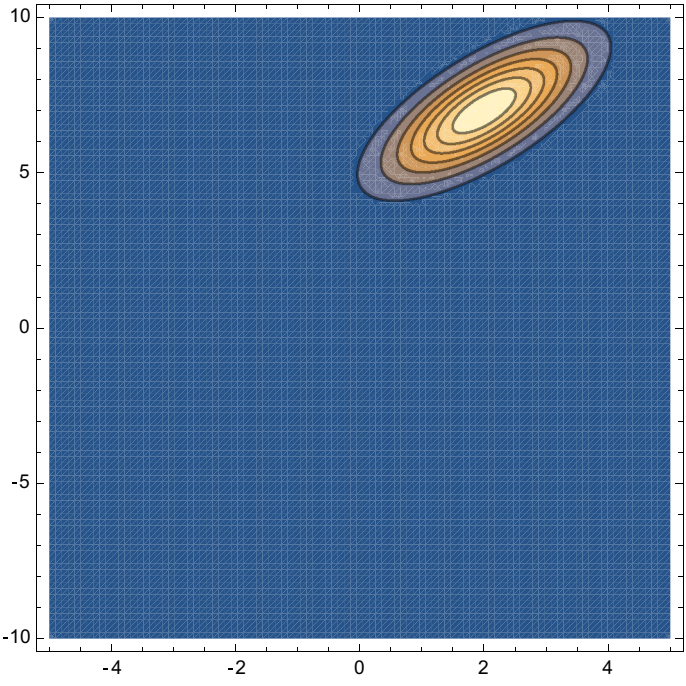

Fig. 2 The Wigner function $W(\Sigma_1, \mu_1)$

Recall that our objective is to find geodesics on the parameter manifold (Θ, g_θ) that minimize the square geodesic distance

$$W_Q^2(\rho(\bullet,\theta^0),\rho(\bullet,\theta^1)) = \inf_{\substack{\theta\in C^1(0,1)\cap C[0,1]\\ \theta(0)=\theta^0,\theta(1)=\theta^1}} \left\{ \int_0^1 \langle \dot\theta(t), G_W(\theta(t))\dot\theta(t)\rangle \, dt \right\}. \tag{4.3}$$

We then discretize the integral of the optimal control problem as

$$\begin{aligned} &\min_{\theta_i;1\le i\le N-1} N^{-1}\sum_{i=1}^{N-1}\left\langle \left(\frac{\theta_{i+1}-\theta_i}{N}\right), G_W(\theta_i)\left(\frac{\theta_{i+1}-\theta_i}{N}\right)\right\rangle \\ &\quad = \min N^{-3}\sum_{i=1}^{N-1}\left(\|\mu_{i+1}-\mu_i\|^2 + \operatorname{tr}\left((S_{\theta_{i+1}}-S_{\theta_i})\Sigma_i(S_{\theta_{i+1}}-S_{\theta_i})\right)\right) \end{aligned} \tag{4.4}$$

with boundary conditions $\theta_0 = \theta^0$ and $\theta_N = \theta^1$.

This minimization problem can be easily solved using a simple Monte-Carlo algorithm minimizing (2.5) that only accepts transitions to states that satisfy the two constraints

$$\Sigma_i \ge 0 \text{ and } \Sigma_i + i\nu \ge 0.$$

The numerical solution to the quantum transport problem of the two parametrized Gaussian states is illustrate in Fig. 3.

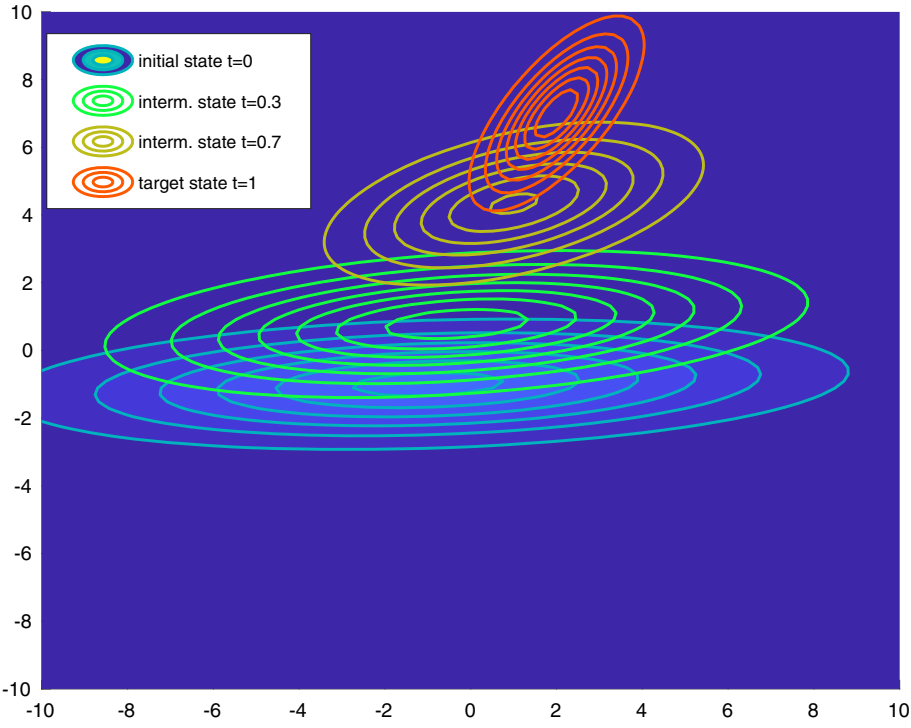

Fig. 3 Optimal quantum transport map from quantum state with Wigner function $W(\theta^0)$ to quantum state with Wigner function $W(\theta^1)$

4.2 Examples Involving the Quantum Fermionic Fokker–Planck Equation

Example 2 (Fermionic Fokker–Planck equation; Analytic solution) We consider the fermionic Fokker–Planck equation as introduced in Sect. 2.5 for simplest case $n = 1$, i.e. $\mathfrak{C}$ can be identified with the two-dimensional Hilbert space $\mathrm{span}\{\mathrm{id}_{\mathbb{C}^{2\times 2}}, \sigma_x\}$ in which case we can solve the problem analytically.

The grading operator is defined by

$$\Gamma(\mathrm{id}) = \mathrm{id} \text{ and } \Gamma(\sigma_x) = -\sigma_x.$$

The faithful states in $\mathfrak{C}$ are then parametrized by

$$(-1, 1) \ni \theta \mapsto \rho(\theta) := \mathrm{id} + \theta\sigma_x.$$

We can diagonalize this density operator using the unitary map $U = 2^{-1/2}(\sigma_x - \sigma_z)$. This way, $U\rho(\theta)U = \mathrm{diag}(1 + \theta, 1 - \theta)$. The derivative is given by

$$\nabla(\alpha\mathrm{id} + \beta\sigma_x) = \beta\mathrm{id}.$$

The operator $L_{\rho(\theta),\Gamma(\rho(\theta))}(\mathrm{id}) = \int_0^1 (\rho(\theta))^{1-s}(\rho(-\theta))^s \, ds$ becomes therefore after conjugating with U

$$UL_{\rho(\theta),\Gamma(\rho(\theta))}(\mathrm{id})U = \int_0^1 (1-\theta)^{1-s}(1+\theta)^s \, ds \mathrm{id}_{\mathbb{C}^{2\times 2}} = \frac{\theta}{\mathrm{artanh}(\theta)}\mathrm{id}_{\mathbb{C}^{2\times 2}}.$$

This implies that $-\Delta_{\rho(\theta)}|_{\mathrm{span}(\sigma_x)} = \frac{\theta}{\mathrm{artanh}(\theta)}\mathrm{id}$. Using that $D_\theta\rho(\theta) = \sigma_x$ and that $G_\theta = \mathrm{id}$, we find from (3.3) that

$$G_W(\theta) = \frac{\mathrm{artanh}(\theta)}{\theta}.$$

As before, our objective is to find geodesics on the parameter manifold (Θ, g_θ) that minimize the square geodesic distance

$$W_Q^2(\rho(\bullet, \theta^0), \rho(\bullet, \theta^1)) = \inf_{\substack{\theta\in C^1(0,1)\cap C[0,1] \\ \theta(0)=\theta^0, \theta(1)=\theta^1}} \left\{ \int_0^1 \mathcal{L}(\theta(t), \dot\theta(t)) \, dt \right\} \tag{4.5}$$

where $\mathcal{L}(\theta(t), \dot\theta(t)) := \dot\theta(t)^2 G_W(\theta(t))$ is the Lagrangian. The associated Euler–Lagrange equation

$$\partial_1\mathcal{L}(\theta(t), \dot\theta(t)) - \frac{d}{dt}\partial_2\mathcal{L}(\theta(t), \dot\theta(t)) = 0$$

becomes

$$\begin{aligned} &\dot\theta(t)^2 G_W'(\theta(t)) - 2\frac{d}{dt}(\dot\theta(t)G_W(\theta(t))) \\ &\quad = -\dot\theta(t)^2 G_W'(\theta(t)) - 2\ddot\theta(t)G_W(\theta(t)) = 0. \end{aligned} \tag{4.6}$$

Using that $G_W(\theta) > 0$, we find the identities for $\pm\dot\theta(t) > 0$

$$\frac{d}{dt}\log(G_W(\theta(t))) = \frac{G_W'(\theta(t))\dot\theta(t)}{G_W(\theta(t))} \text{ and } \frac{d}{dt}\log(\pm\dot\theta(t)) = \frac{\ddot\theta(t)}{\dot\theta(t)}.$$

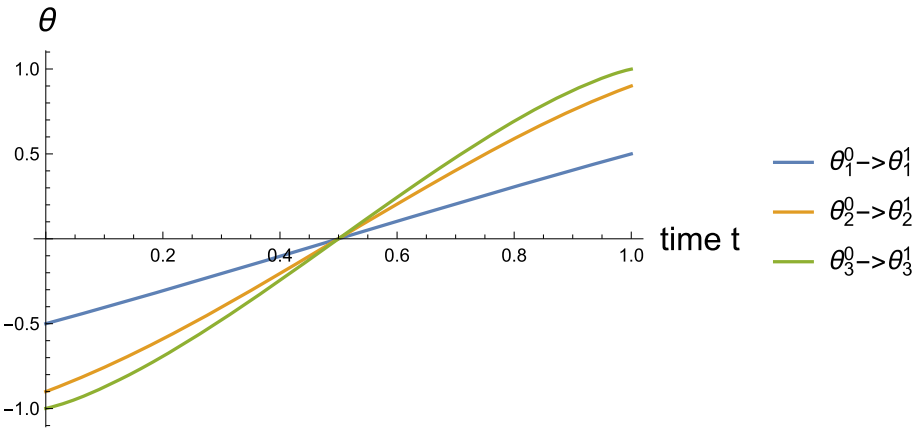

Fig. 4 Illustration of solution (4.7) for three different boundary parameters (4.8)

Assuming that $\theta^1 > \theta^0$ in the sequel and thus dropping $\pm$ for simplicity, the Euler–Lagrange equation is then equivalent, for some constant $k \in \mathbb{R}$, to the ODE

$$-\log(G_W(\theta(t))) + k = \log(\pm\dot{\theta}(t)).$$

Introducing then the function $\zeta(x) := \frac{\mathrm{Li}_2(x)-\mathrm{Li}_2(-x)}{2}$ in terms of the dilogarithm, Li_2, we can then specify the constant e^k by

$$e^k := \int_0^1 \frac{\dot{\theta}(t)}{\theta(t)} \operatorname{artanh}(\theta(t))\, dt = \int_{\theta^0}^{\theta^1} \frac{\operatorname{artanh}(x)}{x}\, dx = \zeta(\theta^1) - \zeta(\theta^0).$$

In particular, this allows us to explicitly state the solution to the optimal transport problem

$$\theta(t) = \zeta^{-1}\left(t\zeta(\theta^1) + (1-t)\zeta(\theta^0)\right). \tag{4.7}$$

We illustrate this by choosing three different pairs of parameters (Fig. 4)

$$\begin{aligned} \theta_1^0 &= -\tfrac{1}{2},\ \theta_1^1 = \tfrac{1}{2}, \\ \theta_2^0 &= -\tfrac{9}{10},\ \theta_2^1 = \tfrac{9}{10}, \\ \theta_3^0 &= -\tfrac{999}{1000},\ \theta_3^1 = \tfrac{999}{1000}. \end{aligned} \tag{4.8}$$

4.2.1 Anti-commutator Case

We can repeat the previous analysis by considering instead of the Feynman–Kubo–Mori multiplication operator $L_{\rho(\theta),\Gamma(\rho(\theta))}$ the anti-commutator (2.10) which satisfies

$$L^{\mathrm{ac}}_{\rho(\theta)}(\mathrm{id}) = \rho(\theta). \tag{4.9}$$

Thus, using that $\nabla^*(\rho(\theta)) = 1$, we find that $-\Delta_{\rho(\theta)}|_{\mathrm{span}(\sigma_x)} = 1$ and therefore also

$$G_W(\theta) \equiv 1.$$

In particular, since the Lagrangian is just $\mathcal{L}(\dot{\theta}(t)) = \dot{\theta}(t)^2$, the geodesics in parameter space are just straight lines as the Euler–Lagrange equation $\ddot{\theta}(t) = 0$ immediately shows.

4.2.2 Wasserstein Natural Gradient

We shall now also illustrate the Wasserstein natural gradient for the quantum Fokker–Planck equation as in Example 2 by minimizing the von Neumann entropy as objective function

$$R(\theta) = \tau(\rho(\theta)\log(\rho(\theta))).$$

From the matrix logarithm

$$\log(\rho(\theta)) = \frac{1}{2}\begin{pmatrix} \log(1-\theta^2) & \log\left(\frac{1+\theta}{1-\theta}\right) \\ \log\left(\frac{1+\theta}{1-\theta}\right) & \log(1-\theta^2) \end{pmatrix} \tag{4.10}$$

we then immediately see that

$$R(\theta) = \frac{1}{2}\left(\log(1-\theta^2) + \theta\log\left(\frac{1+\theta}{1-\theta}\right)\right)$$

and hence $D_\theta R(\theta) = \operatorname{artanh}(\theta)$. Therefore, the Wasserstein gradient (3.5) becomes $\nabla_g R(\theta) = G_W(\theta)^{-1} D_\theta R(\theta) = -\theta$. The gradient descent equation therefore becomes in parameter space

$$\theta'(t) = -\theta(t)$$

which implies that we will converge exponentially fast to the unique minimizer the completely mixed state that corresponds to $\theta = 0$.

4.3 Channel Parameter Estimation-Pushforward of Quantum States

The idea of parameter estimation of probability densities constructed from the pushforward of possibly nonlinear activation functions, relevant for neural networks, has been investigated by the second author in [25].

In quantum theory the framework is somewhat different, since quantum operations on a physical system are described by linear (super)-operators, so-called quantum channels rather than non-linear one-dimensional functions. A quantum channel is a completely positive and trace preserving (CPTP) map. Thus, it is natural to consider the situation where a state is parametrized by the output of a quantum channel Φ_θ depending on some parameter θ which is the quantum analogue of the pushforward of probability measures by parametrized functions.

We shall illustrate how such problems can be studied in our framework by considering the quantum depolarizing channel (4.13) with the quantum fermionic Fokker–Planck equation, introduced in Sect. 2.5, for $n = 2$.

Example 3 (Depolarizing channel and quantum Fokker–Planck dynamics) Consider the fermionic Fokker–Planck equation with $n = 2$, the fermionic operators are then given by

$$Q_1 := \sigma_x \otimes \mathrm{id}_{\mathbb{C}^2} \text{ and } Q_2 := \sigma_z \otimes \sigma_x$$

and thus

$$Q^{(0,0)} = \mathrm{id}_{\mathbb{C}^4},\ Q^{(1,0)} = \sigma_x \otimes \mathrm{id}_{\mathbb{C}^2},\ Q^{(0,1)} = \sigma_z \otimes \sigma_x,\ Q^{(1,1)} = -i\sigma_y \otimes \sigma_x. \tag{4.11}$$

Thus, we find for the gradients

$$\nabla(Q^{0,0}) = \mathbf{0}, \quad \nabla(Q^{1,0}) = (Q^{(0,0)}, 0)$$

$$\nabla(Q^{0,1}) = (0,\, Q^{(0,0)}),\ \text{and}\ \nabla(Q^{1,1}) = (Q^{(0,1)}, -Q^{(1,0)}). \tag{4.12}$$

We consider the depolarizing channel for some density operator $\rho = \frac{1}{2}(Q^{(1,0)} + Q^{(0,0)})$ and limiting state $\frac{1}{2}(Q^{(0,1)} + Q^{(0,0)})$

$$\Phi_\theta(\rho) = \frac{1}{2}\left(e^{-\theta}Q^{(1,0)} + (1-e^{-\theta})Q^{(0,1)} + Q^{(0,0)}\right). \tag{4.13}$$

Then, after applying the anti-commutation operator (2.10)

$$\begin{aligned}
L^{\mathrm{ac}}_{\Phi_\theta(\rho)}(\nabla(Q^{1,0})) &= \frac{1}{2}(e^{-\theta}Q^{(1,0)} + (1-e^{-\theta})Q^{(0,1)} + Q^{(0,0)}, 0)\\
L^{\mathrm{ac}}_{\Phi_\theta(\rho)}(\nabla(Q^{0,1})) &= \frac{1}{2}(0, e^{-\theta}Q^{(1,0)} + (1-e^{-\theta})Q^{(0,1)} + Q^{(0,0)})\\
L^{\mathrm{ac}}_{\Phi_\theta(\rho)}(\nabla(Q^{1,1})) &= \frac{1}{2}((1-e^{-\theta})Q^{(0,0)} + Q^{(0,1)}, -e^{-\theta}Q^{(0,0)} - Q^{(1,0)})
\end{aligned} \tag{4.14}$$

we find for the Laplacian

$$\begin{aligned}
-\Delta_{\Phi_\theta}(Q^{(1,0)}) &= \frac{1}{2}\left((1-e^{-\theta})Q^{(1,1)} + Q^{(1,0)}\right)\\
-\Delta_{\Phi_\theta}(Q^{(0,1)}) &= \frac{1}{2}\left(-e^{-\theta}Q^{(1,1)} + Q^{(0,1)}\right)\\
-\Delta_{\Phi_\theta}(Q^{(1,1)}) &= \frac{1}{2}\left((1-e^{-\theta})Q^{(1,0)} - e^{-\theta}Q^{(0,1)}\right) + Q^{(1,1)}.
\end{aligned} \tag{4.15}$$

This means that the Laplacian has a matrix representation

$$-\Delta_{\Phi_\theta}|_{\mathrm{span}\{\mathrm{id}\}^\perp} = \frac{1}{2}\begin{pmatrix} 1 & 0 & 1-e^{-\theta} \\ 0 & 1 & -e^{-\theta} \\ 1-e^{-\theta} & -e^{-\theta} & 2 \end{pmatrix} \tag{4.16}$$

with inverse

$$(-\Delta_{\Phi_\theta}|_{\mathrm{span}\{\mathrm{id}\}^\perp})^{-1} = \frac{2}{2e^{\theta} + e^{2\theta} - 2}\begin{pmatrix} 2e^{2\theta}-1 & 1-e^{\theta} & e^{\theta}\left(1-e^{\theta}\right) \\ 1-e^{\theta} & 2e^{\theta}+e^{2\theta}-1 & e^{\theta} \\ e^{\theta}\left(1-e^{\theta}\right) & e^{\theta} & e^{2\theta} \end{pmatrix}. \tag{4.17}$$

Since, $D_\theta\Phi_\theta(\rho) = \frac{e^{-\theta}}{2}\left(Q^{(0,1)} - Q^{(1,0)}\right)$ we thus find

$$\begin{aligned}
G_W(\theta) &= \tau(D_\theta\Phi_\theta(\rho)(-\Delta_{\Phi_\theta}|_{\mathrm{span}\{\mathrm{id}\}^\perp})^{-1}D_\theta\Phi_\theta(\rho))\\
&= \frac{e^{-2\theta}\left((2e^{2\theta}-1) - 2(1-e^{\theta}) + (2e^{\theta}+e^{2\theta}-1)\right)}{4e^{\theta}+2e^{2\theta}-2} = \frac{3-4e^{-2\theta}+6e^{-\theta}}{4e^{\theta}+2(e^{2\theta}-1)}.
\end{aligned} \tag{4.18}$$

5 Discussion

In this paper, we pull back the quantum Wasserstein-2 metric into a parameterized quantum statistical models. This allows us to develop a quantum Wasserstein/transport information matrix. Using this matrix, we develop the quantum transport natural gradient methods and apply them to the quantum statistical learning problems. Besides, we also consider the optimal control problem of quantum transport natural gradient flows, which leads to the derivation of quantum Schrödinger bridge problem. Several analytical examples, such as the transport

of Gaussian states on the statistical manifold in Example 1, the transport of states for the gradient induced by quantum fermionic Fokker–Planck equation in Sect. 4.2 on the statistical manifold, and the parameter estimation problem for channels in Sect. 4.3, are provided.

Our results initialize the joint study among quantum information geometry and quantum optimal transport. We pull back the quantum system dynamics into a finite-dimensional parameter space generated by statistical and machine learning models. We call this area *quantum transport information geometry*. Here the interaction study between quantum Fisher and quantum Wasserstein information matrices becomes essential. We expect that this joint study would be useful in developing transport estimation theory of quantum information theory, and designing AI-driven quantum computing algorithms for quantum systems. In the future, we will continue this line of study following transport information geometry [26,30].

Acknowledgements This work was supported by the EPSRC Grant EP/L016516/1 for the University of Cambridge CDT, the CCA and generous funding from IPAM in Los Angeles (SB). Wuchen Li was supported by University of South Carolina, start up funding.

References

1. Alicki, R.: On the detailed balance condition for non-Hamiltonian systems. Rep. Math. Phys. **10**(2), 249–258 (1976)
2. Amari, S.: Information Geometry and Its Applications, 1st edn. Springer Publishing Company, Incorporated (2016)
3. Ay, N., Jost, J., Lê, H.V., Schwachhöfer, L.: Information Geometry, vol. 64. Springer, Cham (2017)
4. Becker, S., Datta, N., Lami, L., Rouzé, C.: Convergence rates for the quantum central limit theorem (2019). arXiv:1912.06129
5. Benamou, J.-D., Brenier, Y.: A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. Numer. Math. **84**(3), 375–393 (2000)
6. Blume-Kohout, R.: Optimal, reliable estimation of quantum states. New J. Phys. **12**(4), 043034 (2010)
7. Carlen, E., Maas, J.: An analog of the 2-Wasserstein metric in non-commutative probability under which the fermionic Fokker–Planck equation is gradient flow for the entropy. Commun. Math. Phys. **331**, 887–926 (2014)
8. Carlen, E., Maas, J.: Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance. Commun. Math. Phys. **331**, 887–926 (2018)
9. Carlen, E., Maas, J.: Non-commutative calculus, optimal transport and functional inequalities in dissipative quantum systems. J. Stat. Phys. **2020**(178), 319–378 (2020)
10. Çaylak, O., Anatole von Lilienfeld, O., Baumeier, B.: Wasserstein metric for improved QML with adjacency matrix representations (2020). arXiv:2001.11005
11. Chakrabarti, S., Huang, Y., Tongyang, L., Feizi, S., Wu, X.: Quantum Wasserstein GANs. In: 33rd Conference on Neural Information Processing Systems (NeurIPS 2019)
12. Chen, Y., Li, W.: Optimal transport natural gradient for statistical manifolds with continuous sample space. Inf. Geom. **3**, 1–32 (2020)
13. Chen, Y., Georgiou, T., Pavon, M.: On the relation between optimal transport and Schrödinger bridges: a stochastic control viewpoint. J. Optim. Theory Appl. **169**(2), 671–691 (2016)
14. Chen, Y., Georgiou, T.T., Tannenbaum, A.: Wasserstein geometry of quantum states and optimal transport of matrix-valued measures. In: Emerging Applications of Control and Systems Theory, pp. 139–150. Springer, Berlin (2018)

15. Chen, Yongxin, Georgiou, Tryphon T., Tannenbaum, Allen: Matrix optimal mass transport: a quantum mechanical approach. IEEE Trans. Autom. Control **63**(8), 2612–2619 (2018)
16. D'Ariano, G.M., Laurentis, M., Paris, M.G.A., Porzio, A., Solimeno, S.: Quantum tomography as a tool for the characterization of optical devices. J. Optics B **4**(3), S127 (2002)
17. De Palma, G., Trevisan, D.: Quantum optimal transport with quantum channels (2019). arXiv:1911.00803
18. Evans, L.C., Gangbo, W.: Differential equations methods for the Monge–Kantorovich mass transfer problem. Mem. AMS **137**(653), 166 (1999)
19. Golse, F., Mouhot, C., Thierry, Paul: On the mean field and classical limits of quantum mechanics. Commun. Math. Phys. **343**(1), 165–205 (2016). 22
20. Gross, D., Liu, Y.K., Flammia, S., Becker, S., Eisert, J.: Quantum state tomography via compressed sensing. Phys. Rev. Lett. **105**(15), 150401 (2010)
21. Hayashi, M.: Quantum Information Theory. Mathematical Foundation. Graduate Texts in Physics. Springer, Berlin (2017)
22. Lafferty, J.: The density manifold and configuration space quantization. Trans. Am. Math. Soc. **305**(2), 699–741 (1988)
23. Léonard, C.: A survey of the Schrödinger problem and some of its connections with optimal transport. arXiv:1308.0215 (2013)
24. Li, W.C.: Transport information geometry I: Riemannian calculus on probability simplex (2018). arXiv:1803.06360
25. Li, W.C.: Hessian metric via transport information geometry (2020). arXiv:2003.10526
26. Li, W.C., Montufar, G.: Natural gradient via optimal transport. Information Geometry (2018)
27. Li, W.C., Montufar, G.: Ricci curvature for parameter statistics via optimal transport, Information Geometry (2020)
28. Li, W.C., Zhao, J.: Wasserstein information matrix (2020). arXiv:1910.11248
29. Li, W.C., Yin, P., Osher, S.: Computations of optimal transport distance with Fisher information regularization (2018). arXiv:1704.04605
30. Li, W.C., Liu, S., Zha, H., Zhou, H.: Parametric Fokker–Planck equation, GSI (2019)
31. Lvovsky, A.I., Raymer, M.G.: Continuous-variable optical quantum state tomography. Rev. Mod. Phys. **81**(1), 299–332 (2009)
32. Mittnenzweig, M., Mielke, A.: An entropic gradient structure for Lindblad equations and couplings of quantum systems to macroscopic models. J. Stat. Phys. **167**(2), 205–233 (2017)
33. Nagaoka, H.: Differential geometrical aspects of quantum state estimation and relative entropy. In: Belavkin, V.P., Hirota, O., Hudson, R.L. (eds.) Quantum Communications and Measurement, pp. 449–452. Plenum, New York (1995)
34. Ning, L., Georgiou, T., Tannenbaum, A.: On matrix-valued Monge–Kantorovich optimal mass transport. IEEE Trans. Autom. Control **60**(2), 373–382 (2015)
35. Otto, F.: The geometry of dissipative evolution equations: the porous medium equation. Commun. Partial Differ. Equ. **26**(2001), 101–174 (2001)
36. Paris, M., Rehacek, J.: Quantum State Estimation. Lecture Notes in Physics, vol. 649. Springer, Berlin (2004)
37. Rouzé, C., Datta, N.: Concentration of quantum states from quantum functional and transportation cost inequalities. J. Math. Phys. **60**(1), 012202 (2019)
38. Rouzé, Datta N: Relating relative entropy, optimal transport and Fisher information: a quantum HWI inequality. Ann. Henri Poincaré **21**, 2115–2150 (2020)
39. Schmidt, J., Marques, M.R.G., Botti, S., et al.: Recent advances and applications of machine learning in solid-state materials science. Comput. Mater. **5**, 83 (2019)
40. Torlai, G., Mazzola, G., Carrasquilla, J., et al.: Neural-network quantum state tomography. Nat. Phys. **14**, 447–450 (2018)
41. Villani, C.: Optimal Transport: Old and New. Springer, Berlin (2009)
42. Vogel, K., Risken, H.: Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase. Phys. Rev. A **40**(5), 2847–2849 (1989)