A Dual Formula for the Noncommutative Transport Distance

Melchior Wirth[1] (ID)

Abstract
In this article we study the noncommutative transport distance introduced by Carlen and Maas and its entropic regularization defined by Becker and Li. We prove a duality formula that can be understood as a quantum version of the dual Benamou–Brenier formulation of the Wasserstein distance in terms of subsolutions of a Hamilton–Jacobi–Bellmann equation.

Keywords Quantum Markov semigroup · Duality · Quantum optimal transport

1 Introduction

The theory of optimal transport [28, 29] has experienced rapid growth in recent years with applications in diverse fields across pure and applied mathematics. Along with this growth came a lot of interest in extending the methods of optimal transport beyond the scope of its original formulation as an optimization problem for the transport cost between two probability measures.

One such extension deals with "quantum spaces", where the probability measures are replaced by density matrices or density operators. Most of the work on quantum optimal transport in this sense can be grouped into one of the following two categories. The first approach (see e.g. [6–9, 11, 22, 27]) takes a quantum Markov semigroup (QMS) as input datum and relies on a noncommutative analog of the Benamou–Brenier formulation [4] of the Wasserstein distance for probability measures on Euclidean space

$$W_2^2(\mu, \nu) = \inf \left\{ \int_0^1 \int_{\mathbb{R}^n} |v_t|^2 \, d\rho_t \, dt : \rho_0 = \mu, \rho_1 = \nu, \dot{\rho}_t + \nabla \cdot (\rho_t v_t) = 0 \right\}.$$

In the simple case when the generator $\mathscr{L}$ of the QMS is of the form

$$\mathscr{L}A = \sum_{j \in \mathcal{J}} [V_j, [V_j, A]]$$

Communicated by Eric A. Carlen.

✉ Melchior Wirth
melchior.wirth@ist.ac.at

[1] Institute of Science and Technology Austria (ISTA), Am Campus 1, 3400 Klosterneuburg, Austria

with self-adjoint matrices V_j, the associated noncommutative transport distance $\mathcal{W}$ on the set of density matrices is given by

$$\mathcal{W}^2(\rho_0, \rho_1) = \inf \left\{ \int_0^1 \sum_{j \in \mathcal{J}} \tau(\mathbf{W}_j^*(t)[\rho(t)]_0^{-1}(\mathbf{W}_j)) \, dt : \dot\rho(t) = \sum_{j \in \mathcal{J}} [V_j, \mathbf{W}_j(t)] \right\},$$

where the infimum is taken over curves ρ that satisfy $\rho(0) = \rho_0$, $\rho(1) = \rho_1$, and where

$$[X]_0(A) = \int_0^1 X^\alpha A X^{1-\alpha} \, d\alpha.$$

For the definition of the metric $\mathcal{W}$ in the more general case of a QMS satisfying the detailed balance condition (DBC), we refer to the next section.

This approach has proven fruitful in applications to noncommutative functional inequalities, similar in spirit to the heuristics known as Otto calculus [8, 9, 12, 31].

The second approach (see e.g. [13, 14, 17, 23, 25, 26]) seeks to find a suitable noncommutative analog of the Monge–Kantorovich formulation [20] of the Wasserstein distance via couplings (or transport plans):

$$W_p^p(\mu, \nu) = \inf \left\{ \int_{X \times X} d^p(x, y) \, d\pi(x, y) : (\text{pr}_1)_\# \pi = \mu, (\text{pr}_2)_\# \pi = \nu \right\}.$$

This approach also allows to consider a quantum version of Monge–Kantorovich problem for arbitrary cost functions. So far, possible connections between these two approaches in the quantum world stay elusive.

The focus of this article lies on the noncommutative transport distance $\mathcal{W}$ introduced in the first approach. More precisely, we prove a dual formula that is a noncommutative analog of the expression of the classical L^2-Wasserstein distance in terms of subsolutions of the Hamilton–Jacobi equation [5, 24]

$$W_2^2(\mu, \nu) = \frac{1}{2} \inf \left\{ \int_{\mathbb{R}^n} u_1 \, d\mu - \int_{\mathbb{R}^n} u_0 \, d\nu : \dot u_t + \frac{1}{2} |\nabla u_t|^2 \leq 0 \right\}.$$

This result yields a noncommutative version of the dual formula obtained independently by Erbar et al. [15] and Gangb et al. [16] for the Wasserstein-like transport distance on graphs. In fact, we prove a dual formula that is not only valid for the metric $\mathcal{W}$, but also for the entropic regularization recently introduced by Becker–Li [3]. When the generator $\mathscr{L}$ is again of the simple form discussed above, the entropic regularization $\mathcal{W}_\varepsilon$ is a metric obtained when replacing the constraint

$$\dot\rho(t) = \sum_{j \in \mathcal{J}} [V_j, \mathbf{W}_j(t)]$$

in the definition of $\mathcal{W}$ by

$$\dot\rho(t) = \sum_{j \in \mathcal{J}} [V_j, \mathbf{W}_j(t)] + \varepsilon \mathscr{L}^\dagger \rho(t).$$

With the notation introduced in the next section, the main result of this article reads as follows.

Theorem *Let* $\sigma \in M_n(\mathbb{C})$ *be an invertible density matrix and* (P_t) *an ergodic QMS on* $M_n(\mathbb{C})$ *that satisfies the* σ*-DBC. The entropic regularization* $\mathcal{W}_\varepsilon$ *of noncommutative transport*

distance induced by (P_t) *satisfies the following dual formula*:

$$\frac{1}{2}\mathcal{W}_{\varepsilon}^2(\rho_0, \rho_1) = \sup\{\tau(A(1)\rho_1 - A(0)\rho_0) \mid A \in HJB_{\varepsilon}^1\}.$$

Here a QMS (P_t) is said to satisfy the σ-DBC if

$$\tau((P_t A)^* B\sigma) = \tau(A^*(P_t B)\sigma)$$

for all $A, B \in M_n(\mathbb{C})$ and $t \geq 0$. If σ is the identity matrix, this is the case exactly when the generator is of the form $\mathscr{L}A = \sum_j [V_j, [V_j, A]]$ with self-adjoint matrices V_j.

Moreover, HJB_{ε}^1 stands for the set of all Hamilton–Jacobi–Bellmann subsolutions, a suitable noncommutative variant of solutions of the differential inequality

$$\dot{u}(t) + \frac{1}{2}|\nabla u(t)|^2 - \varepsilon \Delta u(t) \leq 0.$$

Other metrics similar to $\mathcal{W}$ also occur in the literature, most notably the one called the "anticommutator case" in [3, 10, 11]. In [9, 30], a class of such metrics was studied in a systematic way, and our main theorem applies in fact to this wider class of metrics. For the anticommutator case, this duality formula was obtained before in [10].

There are still some very natural questions left open. For one, we do not discuss the existence of optimizers. While for the primal problem this follows from a standard compactness argument, this question is more delicate for the dual problem, even when dealing with probability densities on discrete spaces instead of density matrices, and one has to relax the problem to obtain maximizers (see [16, Sects. 6–7]).

Another interesting direction would be to extend the duality result from matrix algebras to infinite-dimensional systems. While a definition of the metric $\mathcal{W}$ for QMSs on semi-finite von Neumann algebras is available [19, 30], the problem of duality seems to be much harder to address. Even for abstract diffusion semigroups, the best known result only shows that the primal distance is the upper length distance associated with the dual distance and leaves the question of equality open [2, Proposition 10.11].

2 Setting and Basic Definitions

In this section we introduce basic facts and definitions about QMSs that will be used later on. In particular, we review the definition of the noncommutative transport distance from [8] and its entropic regularization introduced in [3]. Our notation mostly follows [8, 9]. For a list of symbols we refer the reader to the end of this article.

Let $M_n(\mathbb{C})$ denote the complex $n \times n$ matrices and let $\mathcal{A}$ be a unital $*$-subalgebra of $M_n(\mathbb{C})$. Let $\mathcal{A}_h$ denote the self-adjoint part of $\mathcal{A}$, $\mathcal{A}_+$ the cone of positive elements of $\mathcal{A}$ and $\mathcal{A}_{++}$ the subset of invertible positive elements. We write τ for the normalized trace on $M_n(\mathbb{C})$, that is,

$$\tau(A) = \frac{1}{n}\sum_{k=1}^n A_{kk},$$

and $\mathfrak{H}_{\mathcal{A}}$ for the Hilbert space formed by equipping $\mathcal{A}$ with the GNS inner product

$$\langle \cdot, \cdot \rangle_{\mathfrak{H}_A} : \mathcal{A} \times \mathcal{A} \to \mathbb{C}, \quad (A, B) \mapsto \tau(A^* B).$$

The adjoint of a linear operator $\mathscr{K} : \mathfrak{H}_{\mathcal{A}} \to \mathfrak{H}_{\mathcal{A}}$ is denoted by $\mathscr{K}^{\dagger}$.

We write $\mathfrak{S}(\mathcal{A})$ for the set of all density matrices on $\mathcal{A}$, that is, all positive elements $\rho \in \mathcal{A}$ with $\tau(\rho) = 1$. The subset of invertible density matrices is denoted by $\mathfrak{S}_+(\mathcal{A})$.

A QMS on $\mathcal{A}$ is a family $(P_t)_{t\geq 0}$ of linear operators on $\mathcal{A}$ that satisfy the following conditions:

- P_t is unital and completely positive for every $t \geq 0$,
- $P_0 = \mathrm{id}_{\mathcal{A}}$, $P_{s+t} = P_s P_t$ for all $s, t \geq 0$,
- $t \mapsto P_t$ is continuous.

We consider a QMS (P_t) on $\mathcal{A}$ which extends to a QMS on $M_n(\mathbb{C})$ satisfying the σ-detailed balance condition (σ-DBC) for some density matrix $\sigma \in \mathfrak{S}_+(\mathcal{A})$, that is,

$$\tau((P_t A)^* B \sigma) = \tau(A^*(P_t B)\sigma)$$

for $A, B \in \mathcal{A}$ and $t \geq 0$. For $\sigma = \mathrm{id}_{\mathcal{A}}$, this reduces to the symmetry condition $P_t^\dagger = P_t$.

Let $\mathscr{L}$ denote the generator of (P_t), that is, the linear operator on $\mathcal{A}$ given by

$$\mathscr{L}(A) = \lim_{t \searrow 0} \frac{P_t A - A}{t}.$$

We further assume that (P_t) is *ergodic* (or primitive), that is, the kernel of $\mathscr{L}$ is one-dimensional. This assumption is natural in this context as it ensures that the metric $\mathcal{W}_{\Lambda, \varepsilon}$ defined below is the geodesic distance induced by a Riemannian metric on $\mathfrak{S}_+(\mathcal{A})$ and in particular that it is finite.

Generators of QMSs are often described by their Lindblad form, but here we will rely on the additional structure coming from the σ-DBC and use a presentation of $\mathscr{L}$ provided by Alicki's theorem [1,Theorem 3], [8,Theorem 3.1] instead: There exists a finite set $\mathcal{J}$, real numbers ω_j for $j \in \mathcal{J}$ and $V_j \in M_n(\mathbb{C})$ for $j \in \mathcal{J}$ with the following properties:

- $\tau(V_j^* V_k) = \delta_{jk}$ for $j, k \in \mathcal{J}$,
- $\tau(V_j) = 0$ for $j \in \mathcal{J}$,
- for every $j \in \mathcal{J}$ there exists a unique $j^* \in \mathcal{J}$ with $V_{j^*} = V_j^*$,
- $\sigma V_j \sigma^{-1} = e^{-\omega_j} V_j$ for $j \in \mathcal{J}$

such that

$$\mathscr{L}(A) = \sum_{j \in \mathcal{J}} \left(e^{-\omega_j/2} V_j^* [A, V_j] - e^{\omega_j/2} [A, V_j] V_j^* \right)$$

for $A \in \mathcal{A}$.

The numbers ω_j are called Bohr frequencies of $\mathscr{L}$ and are uniquely determined by (P_t). The matrices V_j are not uniquely determined by (P_t) and σ, but in the following we will fix a set $\{V_j \mid j \in \mathcal{J}\}$ that satisfies the preceding conditions.

Next we will discuss how the data from Alicki's theorem give rise to a differential structure associated with $\mathscr{L}$.

Let

$$\mathfrak{H}_{\mathcal{A},\mathcal{J}} = \bigoplus_{j \in \mathcal{J}} \mathfrak{H}_{\mathcal{A}}^{(j)},$$

where $\mathfrak{H}_{\mathcal{A}}^{(j)}$ is a copy of $\mathfrak{H}_{\mathcal{A}}$ for $j \in \mathcal{J}$. This is the quantum analog of the space of tangent vector fields in our setting.

We write ∂_j for $[V_j, \cdot]$ and

$$\nabla \colon \mathfrak{H}_{\mathcal{A}} \to \mathfrak{H}_{\mathcal{A},\mathcal{J}}, \quad \nabla(A) = (\partial_j(A))_{j \in \mathcal{J}},$$

which provide analogs of the partial derivatives and the usual gradient operator, respectively. The commutator ∂_j satisfies the product rule

$$\partial_j(AB) = A\partial_j(B) + \partial_j(A)B. \tag{1}$$

Note that in contrast too the usual partial derivatives, the order of the factors plays a role here. This is one central reason for many of the differences and intricacies of the quantum optimal transport distance compared to the classical Wasserstein distance.

Continuing with the analogy with calculus, we write div for the adjoint of $-\nabla$, that is,

$$\mathrm{div} = -\sum_{j \in \mathcal{J}} \partial_j^{\dagger}.$$

The crucial ingredient in the definition of $\mathcal{W}$, which allows to deal with the noncommutativity of the product rule, is the operator $[\rho]_{\omega}$, whose definition we recall next. For $X \in \mathcal{A}_+$ and $\alpha \in \mathbb{R}$ define

$$[X]_{\alpha} : \mathfrak{H}_{\mathcal{A}} \to \mathfrak{H}_{\mathcal{A}}, \quad [X]_{\alpha}(A) = \int_0^1 e^{\alpha(s-1/2)} X^s A X^{1-s} \, ds.$$

The motivation for this definition is a chain rule identity [8, Eq. (5.7)], which can best be illustrated in the case $\alpha = 0$:

$$[X]_0(\partial_j(\log X)) = \partial_j(X).$$

Given $\boldsymbol{\alpha} = (\alpha_j)_{j \in \mathcal{J}}$, we define

$$[X]_{\boldsymbol{\alpha}} : \mathfrak{H}_{\mathcal{A}, \mathcal{J}} \to \mathfrak{H}_{\mathcal{A}, \mathcal{J}}, \quad (\mathbf{V}_j)_{j \in \mathcal{J}} \mapsto ([X]_{\alpha_j} \mathbf{V}_j)_{j \in \mathcal{J}}.$$

For $\varepsilon \geq 0$ we write $\mathsf{CE}_{\varepsilon}(\rho_0, \rho_1)$ for the set of all pairs $(\rho, \mathbf{V})$ such that $\rho \in H^1([0, 1]; \mathfrak{S}_+(\mathcal{A}))$ with $\rho(0) = \rho_0$, $\rho(1) = \rho_1$, $\mathbf{V} \in L^2([0, 1]; \mathfrak{H}_{\mathcal{A}, \mathcal{J}})$ and

$$\dot{\rho}(t) + \mathrm{div}\, \mathbf{V}(t) = \varepsilon \mathscr{L}^{\dagger} \rho(t) \tag{2}$$

for a.e. $t \in [0, 1]$.

Here and in the following we write $H^1([0, 1]; \mathfrak{S}_+(\mathcal{A}))$ for the space of all maps $\rho : [0, 1] \to \mathfrak{S}_+(\mathcal{A})$ such that $(t \mapsto \tau(A\rho(t))) \in H^1([0, 1])$ for all $A \in \mathcal{A}$. The space $L^2([0, 1]; \mathfrak{H}_{\mathcal{A}, \mathcal{J}})$ and other vector-valued functions spaces occurring later are defined similarly.

We define a metric $\mathcal{W}_{\varepsilon}$ on $\mathfrak{S}_+(\mathcal{A})$ by

$$\mathcal{W}_{\varepsilon}^2(\rho_0, \rho_1) = \inf_{(\rho, \mathbf{V}) \in \mathsf{CE}_{\varepsilon}(\rho_0, \rho_1)} \int_0^1 \langle \mathbf{V}(t), [\rho(t)]_{\omega}^{-1} \mathbf{V}(t) \rangle \, dt,$$

where $\boldsymbol{\omega} = (\omega_j)_{j \in \mathcal{J}}$ with the Bohr frequencies ω_j of $\mathscr{L}$.

For $\varepsilon = 0$, this is the noncommutative transport distance $\mathcal{W}$ introduced in [8] (as distance function associated with a Riemannian metric on $\mathfrak{S}(\mathcal{A})_+$), and for $\varepsilon > 0$, this is the entropic regularization of $\mathcal{W}$ introduced in [3].

A standard mollification argument shows that the infimum in the definition of $\mathcal{W}_{\varepsilon}$ can equivalently be taken over $(\rho, \mathbf{V}) \in \mathsf{CE}_{\varepsilon}(\rho_0, \rho_1)$ with $\rho \in C^{\infty}([0, 1]; \mathfrak{S}_+(\mathcal{A}))$. More precisely, if $(\rho, \mathbf{V}) \in \mathsf{CE}_{\varepsilon}(\rho_0, \rho_1)$ and $(\eta_{\delta})_{\delta > 0}$ is a mollifying kernel, then $(\rho * \eta_{\delta}, \mathbf{V} * \eta_{\delta})$ satisfies (2). A suitable reparametrization of the time parameter gives a pair $(\rho^{\delta}, \mathbf{V}^{\delta}) \in \mathsf{CE}_{\varepsilon}(\rho_0, \rho_1)$ such that ρ^{δ} is smooth and

$$\lim_{\delta \searrow 0} \int_0^1 \langle \mathbf{V}^{\delta}(t), [\rho^{\delta}(t)]_{\omega}^{-1} \mathbf{V}^{\delta}(t) \rangle \, dt = \int_0^1 \langle \mathbf{V}(t), [\rho(t)]_{\omega}^{-1} \mathbf{V}(t) \rangle \, dt$$

By a substitution one can reformulate the minimization problem for $\mathcal{W}_\varepsilon$ in such a way that the constraint becomes independent from ε. For that purpose define the relative entropy of $\rho \in \mathfrak{S}_+(\mathcal{A})$ with respect to σ by

$$D(\rho \| \sigma) = \tau(\rho(\log \rho - \log \sigma))$$

and the Fisher information of $\rho \in \mathfrak{S}_+(\mathcal{A})$ by

$$\mathcal{I}(\rho) = \langle [\rho]_{\boldsymbol{\omega}} \nabla(\log \rho - \log \sigma), \nabla(\log \rho - \log \sigma) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}}.$$

According to [3, Theorem 1], one has

$$\mathcal{W}_\varepsilon^2(\rho_0, \rho_1) = \inf_{(\rho, \mathbf{W}) \in \mathsf{CE}_0(\rho_0, \rho_1)} \int_0^1 (\langle \mathbf{W}(t), [\rho(t)]_{\boldsymbol{\omega}}^{-1} \mathbf{W}(t) \rangle + \varepsilon^2 \mathcal{I}(\rho(t))) \, dt$$
$$+ 2\varepsilon(D(\rho_1 \| \sigma) - D(\rho_0 \| \sigma)).$$

The metric $\mathcal{W}$ is intimately connected to the relative entropy and therefore well-suited to study its decay properties along the QMS. For other applications, variants of the metric $\mathcal{W}$ have also proven useful (e.g. [10, 11]), for which the operator $[\rho]_{\boldsymbol{\omega}}$ is replaced. A systematic framework of these metrics has been developed in [9, 30]. It can be conveniently phrased in terms of so-called operator connections.

Let H be an infinite-dimensional Hilbert space. A map $\Lambda \colon B(H)_+ \times B(H)_+ \to B(H)_+$ is called an *operator connection* [21] if

- $A \le C$ and $B \le D$ imply $\Lambda(A, B) \le \Lambda(C, D)$ for $A, B, C, D \in B(H)_+$,
- $C\Lambda(A, B)C \le \Lambda(CAC, CBC)$ for $A, B, C \in B(H)_+$,
- $A_n \searrow A$, $B_n \searrow B$ imply $\Lambda(A_n, B_n) \searrow \Lambda(A, B)$ for $A, A_n, B, B_n \in B(H)_+$.

For example, for every $\alpha \in \mathbb{R}$ the map

$$\Lambda_\alpha \colon (A, B) \mapsto \int_0^1 e^{\alpha(s-1/2)} A^s B^{1-s} \, ds$$

is an operator connection.

It can be shown that every operator connection Λ satisfies

$$U^* \Lambda(A, B) U = \Lambda(U^* A U, U^* B U)$$

for $A, B \in B(H)_+$ and unitary $U \in B(H)$ [21, Sect. 2]. Embedding $\mathbb{C}^n$ into H, one can view $A, B \in M_n(\mathbb{C})$ as bounded linear operators on H, and the unitary invariance of Λ ensures that $\Lambda(A, B)$ does not depend on the embedding of $\mathbb{C}^n$ into H.

For $X \in \mathcal{A}$ define

$$L(X) \colon \mathfrak{H}_{\mathcal{A}} \to \mathfrak{H}_{\mathcal{A}}, \quad A \mapsto XA$$
$$R(X) \colon \mathfrak{H}_{\mathcal{A}} \to \mathfrak{H}_{\mathcal{A}}, \quad A \mapsto AX.$$

Note that if $X \in \mathcal{A}_+$, then

$$\langle A, L(X)A \rangle_{\mathfrak{H}_{\mathcal{A}}} = \tau(A^* X A) = \tau((X^{1/2} A)^* (X^{1/2} A)) \ge 0,$$

so that $L(X)$ is a positive operator, and the same holds for $R(X)$.

Thus we can define

$$[X]_\Lambda = \Lambda(L(X), R(X)).$$

If $\lambda, \mu \ge 0$ and 1_n denotes the identity matrix, then $\Lambda(\lambda 1_n, \mu 1_n)$ is a scalar multiple of the identity as a consequence the unitary invariance of Λ discussed above. By a slight abuse of notation, this scalar will be denoted by $\Lambda(\lambda, \mu)$.

Since $L(X)$ and $R(X)$ commute, we have

$$\Lambda(L(X), R(X))A = \sum_{k,l=1}^{n} \Lambda(\lambda_k, \lambda_l) E_k A E_l \tag{3}$$

for $X \in \mathcal{A}_+$ and $A \in \mathfrak{H}_{\mathcal{A}}$, where (λ_k) are the eigenvalues of X and E_k the corresponding spectral projections.

More generally let $\mathbf{\Lambda} = (\Lambda_j)_{j \in \mathcal{J}}$ be a family of operator connections and define

$$[\rho]_{\Lambda_j} = \Lambda_j(L(\rho), R(\rho)),$$
$$[\rho]_{\mathbf{\Lambda}} = \bigoplus_{j \in \mathcal{J}} [\rho]_{\Lambda_j}.$$

Clearly, $[\rho]_\alpha = [\rho]_{\Lambda_\alpha}$ with the operator connection Λ_α from above.

Then one can define a distance $\mathcal{W}_{\mathbf{\Lambda}, \varepsilon}$ by

$$\mathcal{W}_{\mathbf{\Lambda}, \varepsilon}(\rho_0, \rho_0)^2 = \inf_{(\rho, \mathbf{V}) \in \mathsf{CE}_\varepsilon(\rho_0, \rho_1)} \int_0^1 \langle [\rho(t)]_{\mathbf{\Lambda}}^{-1} \mathbf{V}(t), \mathbf{V}(t) \rangle_{\mathfrak{H}_{\mathcal{A}, \mathcal{J}}} \, dt.$$

If $\Lambda_j = \Lambda_{\omega_j}$ as above, then we retain the original metric $\mathcal{W}_\varepsilon$, while for $\Lambda_j(A, B) = \frac{1}{2}(A + B)$ (and $\varepsilon = 0$) one obtains the distance studied in [10, 11].

Later we will make the additional assumption that $\Lambda_{j*}(A, B) = \Lambda_j(B, A)$, where $j^* \in \mathcal{J}$ is the unique index in the Alicki representation of $\mathscr{L}$ such that $V_{j*} = V_j^*$. It follows from the representation theorem of operator means [21] that the class of metrics $\mathcal{W}_{\mathbf{\Lambda}, 0}$ with $\mathbf{\Lambda}$ subject to this symmetry condition is exactly the class of metrics satisfying Assumptions 7.2 and 9.5 in [9].

For technical reasons in the proof of Theorem 2, it will be necessary to allow for curves of density matrices that are not necessarily invertible. For this purpose, we make the following convention: If $\mathcal{K} \colon \mathfrak{H}_{\mathcal{A}, \mathcal{J}} \to \mathfrak{H}_{\mathcal{A}, \mathcal{J}}$ is a positive operator and $\mathbf{V} \in \mathfrak{H}_{\mathcal{A}, \mathcal{J}}$, we define

$$\langle \mathbf{V}, \mathcal{K}^{-1} \mathbf{V} \rangle_{\mathfrak{H}_{\mathcal{A}, \mathcal{J}}} = \begin{cases} \langle \mathcal{K}\mathbf{W}, \mathbf{W} \rangle_{\mathfrak{H}_{\mathcal{A}, \mathcal{J}}} & \text{if } \mathbf{V} \in (\ker \mathcal{K})^\perp, \mathcal{K}\mathbf{W} = \mathbf{V}, \\ \infty & \text{otherwise.} \end{cases}$$

Since $(\ker \mathcal{K})^\perp = \operatorname{ran} \mathcal{K}$ and $\mathcal{K}$ is injective on $(\ker \mathcal{K})^\perp$, the element $\mathbf{W}$ in this definition exists and is unique. Moreover, this convention is clearly consistent with the usual definition if $\mathcal{K}$ is invertible.

Alternatively, as a direct consequence of the spectral theorem, this expression can equivalently be defined as

$$\langle \mathbf{V}, \mathcal{K}^{-1} \mathbf{V} \rangle_{\mathfrak{H}_{\mathcal{A}, \mathcal{J}}} = \sum_{k=1}^{m} \frac{1}{\lambda_k} |\langle \mathbf{V}, \mathbf{W}_k \rangle_{\mathfrak{H}_{\mathcal{A}, \mathcal{J}}}|^2,$$

where $\lambda_1, \ldots, \lambda_m$ are the eigenvalues of $\mathcal{K}$ and $\mathcal{W}_1, \ldots, W_m$ an orthonormal basis of corresponding eigenvectors.

Lemma 1 *If $\mathcal{K}_n \colon \mathfrak{H}_{\mathcal{A}, \mathcal{J}} \to \mathfrak{H}_{\mathcal{A}, \mathcal{J}}$, $n \in \mathbb{N}$, are positive invertible operators that converge monotonically decreasing to $\mathcal{K}$, then*

$$\langle \mathbf{V}, \mathcal{K}_n^{-1} \mathbf{V} \rangle_{\mathfrak{H}_{\mathcal{A}, \mathcal{J}}} \nearrow \langle \langle \mathbf{V}, \mathcal{K}^{-1} \mathbf{V} \rangle_{\mathfrak{H}_{\mathcal{A}, \mathcal{J}}}$$

for all $\mathbf{V} \in \mathfrak{H}_{\mathcal{A}, \mathcal{J}}$.

Proof From the spectral expression it is easy to see that

$$\langle \mathbf{V}, \mathcal{K}_n^{-1}\mathbf{V}\rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} = \sup_{\delta > 0}\langle \mathbf{V}, (\mathcal{K}_n + \delta)^{-1}\mathbf{V}\rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}}$$

and the same for $\mathcal{K}_n$ replaced by $\mathcal{K}$. Moreover, since $\mathcal{K}_n \searrow \mathcal{K}$, we have $(\mathcal{K}_n + \delta)^{-1} \nearrow (\mathcal{K} + \delta)^{-1}$. Thus

$$
\begin{aligned}
\langle \mathbf{V}, \mathcal{K}^{-1}\mathbf{V}\rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} &= \sup_{\delta > 0}\langle \mathbf{V}, (\mathcal{K} + \delta)^{-1}\mathbf{V}\rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} \\
&= \sup_{\delta > 0}\sup_{n \in \mathbb{N}}\langle \mathbf{V}, (\mathcal{K}_n + \delta)^{-1}\mathbf{V}\rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} \\
&= \sup_{n \in \mathbb{N}}\sup_{\delta > 0}\langle \mathbf{V}, (\mathcal{K}_n + \delta)^{-1}\mathbf{V}\rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} \\
&= \sup_{n \in \mathbb{N}}\langle \mathbf{V}, \mathcal{K}_n^{-1}\mathbf{V}\rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}}.
\end{aligned}
$$

Since $(\langle \mathbf{V}, \mathcal{K}_n^{-1}\mathbf{V}\rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}})$ is monotonically increasing, this settles the claim.

Write $\mathsf{CE}'_\varepsilon(\rho_0, \rho_1)$ for the set of all pairs $(\rho, \mathbf{V})$ such that $\rho \in H^1([0, 1]; \mathfrak{S}(\mathcal{A}))$ with $\rho(0) = \rho_0$, $\rho(1) = \rho_1$, $\mathbf{V} \in L^2([0, 1]; \mathfrak{H}_{\mathcal{A},\mathcal{J}})$ and

$$\dot{\rho}(t) + \operatorname{div}\mathbf{V}(t) = \varepsilon\mathscr{L}^\dagger\rho(t)$$

for a.e. $t \in [0, 1]$. The only difference to the definition of $\mathsf{CE}_\varepsilon(\rho_0, \rho_1)$ is that $\rho(t)$ is not assumed to be invertible.

Proposition 1 *For $\rho_0, \rho_1 \in \mathfrak{S}_+(\mathcal{A})$ we have*

$$\mathcal{W}_{\mathbf{\Lambda},\varepsilon}^2(\rho_0, \rho_1) = \inf_{(\rho,\mathbf{V})\in\mathsf{CE}'_\varepsilon(\rho_0,\rho_1)} \int_0^1 \langle \mathbf{V}(t), [\rho(t)]_{\mathbf{\Lambda}}^{-1}\mathbf{V}(t)\rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}}\, dt.$$

Proof It suffices to show that every curve $(\rho, \mathbf{V}) \in \mathsf{CE}'_\varepsilon(\rho_0, \rho_1)$ can be approximated by curves in $\mathsf{CE}_\varepsilon(\rho_0, \rho_1)$ such that the action integrals converge.

For that purpose let

$$\rho^\delta : [0, 1] \to \mathfrak{S}_+(\mathcal{A}),$$
$$t \mapsto \begin{cases} (1 - t)\rho_0 + t1_{\mathcal{A}} & \text{if } t \in [0, \delta], \\ (1 - \delta)\rho((1 - 2\delta)^{-1}(t - \delta)) + \delta 1_{\mathcal{A}} & \text{if } t \in (\delta, 1 - \delta), \\ t\rho_1 + (1 - t)1_{\mathcal{A}} & \text{if } t \in [1 - \delta, 1]. \end{cases}$$

Since (P_t) is assumed to be ergodic, by [8, Theorem 5.4] there exists for every $t \in [0, 1]$ a unique $X(t) \in \mathcal{A}_h$ with $\tau(X(t))) = 0$ such that

$$1_{\mathcal{A}} - \rho_0 + \operatorname{div}(\nabla X(t)) = \varepsilon(1 - t)\mathscr{L}^\dagger\rho_0,$$

and $X(t)$ depends continuously on t. For $t \in [0, \delta]$ let $\mathbf{V}^\delta(t) = \nabla X(t)$.

Moreover, if λ is the smallest eigenvalue of ρ_0, which is strictly positive by assumption, then $\rho^\delta(t) \geq ((1 - t)\lambda + t)1_{\mathcal{A}} \geq \lambda 1_{\mathcal{A}}$.

Thus

$$\int_0^\delta \langle \mathbf{V}^\delta(t), [\rho^\delta(t)]_\Lambda^{-1} \mathbf{V}^\delta(t) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} \, dt \le \int_0^\delta \langle \mathbf{V}^\delta(t), [\lambda 1_\mathcal{A}]_\Lambda^{-1} \mathbf{V}^\delta(t) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} \, dt$$

$$\le \|[\lambda 1_\mathcal{A}]_\Lambda^{-1}\| \int_0^\delta \|\nabla X(t)\|_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}}^2 \, dt$$

$$\to 0$$

as $\delta \to 0$. Similarly one can show

$$\lim_{\delta \to 0} \int_{1-\delta}^1 \langle \mathbf{V}^\delta(t), [\rho^\delta(t)]_\Lambda^{-1} \mathbf{V}^\delta(t) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} \, dt = 0.$$

By the same argument as above, for a.e. $t \in (\delta, 1 - \delta)$ there exists a unique gradient $\mathbf{W}^\delta(t)$ such that

$$\operatorname{div} \mathbf{W}^\delta(t) = -\frac{2\delta\varepsilon}{1 - 2\delta} \mathscr{L}^\dagger \rho((1 - 2\delta)^{-1}(t - \delta))$$

and

$$\|\mathbf{W}^\delta(t)\|_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} \le \frac{2\delta\varepsilon}{1 - 2\delta} \|\mathscr{L}^\dagger \rho((1 - 2\delta)^{-1}(t - \delta))\|_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}}.$$

Since $\rho \in H^1([0,1]; \mathfrak{S}(\mathcal{A})) \subset C([0,1]; \mathfrak{S}(\mathcal{A}))$, the norm on the right side is bounded independent of δ, so that

$$\|\mathbf{W}^\delta(t)\|_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} \le \tilde{C}\delta$$

with a constant $\tilde{C} > 0$ independent of δ. As $\rho^\delta(t) \ge \delta 1_\mathcal{A}$ for $t \in (\delta, 1 - \delta)$, this implies

$$\int_\delta^{1-\delta} \langle \mathbf{W}^\delta(t), [\rho^\delta(t)]_\Lambda^{-1} \mathbf{W}^\delta(t) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} \, dt \le \frac{1}{\delta} \int_\delta^{1-\delta} \langle \mathbf{W}^\delta(t), [1_\mathcal{A}]_\Lambda^{-1} \mathbf{W}^\delta(t) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} \, dt$$

$$\le \tilde{C}\|[1_\mathcal{A}]_\Lambda^{-1}\|\delta$$

$$\to 0$$

as $\delta \to 0$.

With

$$\mathbf{V}^\delta(t) = \frac{1}{1 - 2\delta} \mathbf{V}((1 - 2\delta)^{-1}(t - \delta)) + \mathbf{W}^\delta(t)$$

we have

$$\dot{\rho}^\delta(t) + \operatorname{div} \mathbf{V}^\delta(t) = \varepsilon\mathscr{L}\rho^\delta(t).$$

Furthermore,

$$\int_\delta^{1-\delta} \left\langle \mathbf{V}\left(\frac{t-\delta}{1-2\delta}\right), [\rho^\delta(t)]_\Lambda^{-1} \mathbf{V}\left(\frac{t-\delta}{1-2\delta}\right) \right\rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} dt$$

$$= \frac{1 - 2\delta}{1 - \delta} \int_0^1 \left\langle \mathbf{V}(s), \left[\rho(s) + \frac{\delta}{1-\delta}\right]_\Lambda^{-1} \mathbf{V}(s) \right\rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} ds,$$

where we used the substitution $s = (1 - 2\delta)^{-1}(t - \delta)$.

By Lemma 1 and the monotone convergence theorem we obtain

$$\lim_{\delta \to 0} \int_\delta^{1-\delta} \left\langle \mathbf{V}\left(\frac{t-\delta}{1-2\delta}\right), [\rho^\delta(t)]_{\boldsymbol{\Lambda}}^{-1} \mathbf{V}\left(\frac{t-\delta}{1-2\delta}\right) \right\rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} dt$$
$$= \int_0^1 \langle \mathbf{V}(s), [\rho(s)]_{\boldsymbol{\Lambda}}^{-1} \mathbf{V}(s) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} \, ds.$$

Together with the convergence result for $\mathbf{W}^\delta$ from above, this implies

$$\int_\delta^{1-\delta} \langle \mathbf{V}^\delta(t), [\rho^\delta(t)]_{\boldsymbol{\Lambda}}^{-1} \mathbf{V}^\delta(t) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} \, dt \to \int_0^1 \langle \mathbf{V}(t), [\rho(t)]_{\boldsymbol{\Lambda}}^{-1} \mathbf{V}(t) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} \, dt.$$

Altogether we have shown

$$\lim_{\delta \to 0} \int_0^1 \langle \mathbf{V}^\delta(t), [\rho^\delta(t)]_{\boldsymbol{\Lambda}}^{-1} \mathbf{V}^\delta(t) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} \, dt = \int_0^1 \langle \mathbf{V}(t), [\rho(t)]_{\boldsymbol{\Lambda}}^{-1} \mathbf{V}(t) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} \, dt.$$

□

3 Real subspaces

Since the proof of the main result relies on convex analysis methods for real Banach spaces, we need to identify suitable real subspaces for our purposes. For $\mathcal{A}$ this is simply $\mathcal{A}_h$, but for $\mathfrak{H}_{\mathcal{A},\mathcal{J}}$ this is less obvious and will be done in the following.

For $j \in \mathcal{J}$ denote by j^* the unique index in $\mathcal{J}$ such that $V_j^* = V_{j^*}$. Let $\tilde{\mathfrak{H}}_{\mathcal{A}}^{(j)}$ be the linear span of $\{X\partial_j A \mid A, X \in \mathcal{A}\}$, and define a linear map $J: \tilde{\mathfrak{H}}_{\mathcal{A}}^{(j)} \to \tilde{\mathfrak{H}}_{\mathcal{A}}^{(j^*)}$ by

$$J(X\partial_j A) = \partial_{j^*}(A^*)X^*.$$

By the product rule (1), $(\partial_j A)X$ also belongs to $\tilde{\mathfrak{H}}_{\mathcal{A},\mathcal{J}}^{(j)}$ and

$$J((\partial_j A)X) = X^* \partial_{j^*}(A^*).$$

Thus J interchanges left and right multiplication, that is, $J(A\mathbf{V}B) = B^* J(\mathbf{V}) A^*$ for $A, B \in \mathcal{A}$ and $\mathbf{V} \in \tilde{\mathfrak{H}}_{\mathcal{A}}^{(j)}$.

Lemma 2 *The map J is anti-unitary.*

Proof For $A, B, X, Y \in \mathcal{A}$ we have

$$\langle J(X\partial_j A), J(Y\partial_j B) \rangle_{\mathfrak{H}_{\mathcal{A}}} = \tau(X(AV_j - V_j A)(V_j^* B^* - B^* V_j^*)Y^*)$$
$$= \tau((B^* V_j^* - V_j^* B^*)Y^* X(V_j A - AV_j))$$
$$= \langle Y\partial_j B, X\partial_j A \rangle.$$

□

Let

$$\mathcal{H}_{\mathcal{A},\mathcal{J}}^{(h)} = \left\{ \mathbf{V} \in \bigoplus_{j \in \mathcal{J}} \tilde{\mathfrak{H}}_{\mathcal{A}}^{(j)} \mid J(\mathbf{V}_j) = \mathbf{V}_{j^*} \right\}.$$

By the previous lemma, $\mathcal{H}_{\mathcal{J},\mathcal{A}}^{(h)}$ is a real Hilbert space.

Lemma 3 *Let $(\Lambda_j)_{j \in \mathcal{J}}$ be a family of operator connections such that*

$$\Lambda_j(B, A) = \Lambda_{j^*}(A, B)$$

for all $j \in \mathcal{J}$. If $A \in \mathcal{A}_h$ and $\rho \in \mathfrak{S}(\mathcal{A})$, then $\nabla A, [\rho]_\Lambda \nabla A \in \mathcal{H}_{\mathcal{A}, \mathcal{J}}^{(h)}$.

__Proof__ For ∇A the statement follows directly from the definitions. For $[\rho]_\Lambda \nabla A$ first note that

$$J\Lambda(L(\rho), R(\rho)) = \Lambda(R(\rho), L(\rho))J$$

as a consequence of the spectral representation (3) and the fact that J interchanges left and right multiplication.

Thus

$$
\begin{aligned}
J([\rho]_{\Lambda_j} \partial_j A) &= J\Lambda_j(L(\rho), R(\rho))\partial_j A \\
&= \Lambda_j(R(\rho), L(\rho))J\partial_j A \\
&= \Lambda_{j^*}(L(\rho), R(\rho))\partial_{j^*} A.
\end{aligned}
$$

$\square$

4 Duality

In this section we prove the duality theorem announced in the introduction. Our strategy follows the same lines as the proof in the commutative case in [15]. It crucially relies on the Rockafellar–Fenchel duality theorem quoted below. Throughout this section we fix an ergodic QMS with generator $\mathscr{L}$ satisfying the σ-DBC for some $\sigma \in \mathfrak{S}_+(\mathcal{A})$ and a family $(\Lambda_j)_{j \in \mathcal{J}}$ of operator connections such that $\Lambda_{j^*}(A, B) = \Lambda_j(B, A)$ for all $j \in \mathcal{J}$.

We need the following definition for the constraint of the dual problem. Here and in the following we write

$$\langle \mathbf{V}, \mathbf{W} \rangle_\rho = \langle \mathbf{V}, [\rho]_\Lambda \mathbf{W} \rangle_{\mathfrak{H}_{\mathcal{A}, \mathcal{J}}}$$

for $\mathbf{V}, \mathbf{W} \in \mathfrak{H}_{\mathcal{A}, \mathcal{J}}$ and $\rho \in \mathcal{A}_+$.

Definition 1 A function $A \in H^1((0, T); \mathcal{A}_h)$ is said to be a *Hamilton–Jacobi–Bellmann subsolution* if for a.e. $t \in (0, T)$ we have

$$\tau((\dot{A}(t) + \varepsilon \mathscr{L} A(t))\rho) + \frac{1}{2}\|\nabla A(t)\|_\rho^2 \le 0 \quad \text{for all } \rho \in \mathfrak{S}(\mathcal{A}).$$

The set of all Hamilton–Jacobi–Bellmann subsolutions is denoted by $\mathsf{HJB}_{\Lambda, \varepsilon}$.

Our proof will establish equality between the primal and dual problem, but before we begin, let us show that one inequality is actually quite easy to obtain.

Proposition 2 *For all $\rho_0, \rho_1 \in \mathfrak{S}_+(\mathcal{A})$ we have*

$$\sup\{\tau(A(1)\rho_1 - A(1)\rho_0) \mid A \in \mathsf{HJB}_{\Lambda, \varepsilon}\}$$

$$\le \frac{1}{2}\inf\left\{\left.\int_0^1 \langle \mathbf{V}(t), [\rho(t)]_\Lambda^{-1}\mathbf{V}(t) \rangle_{\mathfrak{H}_{\mathcal{A}, \mathcal{J}}} \, dt \,\right|\, (\rho, \mathbf{V}) \in \mathsf{CE}_\varepsilon(\rho_0, \rho_1)\right\}.$$

Proof For $A \in \mathsf{HJB}_{\Lambda,\varepsilon}$ and $(\rho, \mathbf{V}) \in \mathsf{CE}_\varepsilon(\rho_0, \rho_1)$ we have

$$
\begin{aligned}
\tau(A(1)\rho_1 - A(0)\rho_0) &= \int_0^1 \tau(\dot{A}(t)\rho(t) + A(t)\dot{\rho}(t)) \, dt \\
&\leq -\int_0^1 \left(\varepsilon\tau((\mathscr{L}A(t))\rho(t)) + \frac{1}{2}\|\nabla A(t)\|_{\rho(t)}^2 \right) dt \\
&\quad + \int_0^1 (\langle \nabla A(t), \mathbf{V}(t)\rangle_{\mathfrak{H}_{A,\mathcal{J}}} + \varepsilon\tau(A(t)\mathscr{L}^\dagger\rho(t))) \, dt \\
&= \int_0^1 \langle [\rho(t)]_\Lambda^{1/2}\nabla A(t), [\rho(t)]_\Lambda^{-1/2}\mathbf{V}(t)\rangle_{\mathfrak{H}_{A,\mathcal{J}}} \, dt \\
&\quad - \frac{1}{2}\int_0^1 \langle [\rho(t)]_\Lambda^{1/2}\nabla A(t), [\rho(t)]_\Lambda^{1/2}\nabla A(t)\rangle_{\mathfrak{H}_{A,\mathcal{J}}} \, dt \\
&\leq \frac{1}{2}\int_0^1 \langle [\rho(t)]_\Lambda^{1/2}\nabla A(t), [\rho(t)]_\Lambda^{1/2}\nabla A(t)\rangle_{\mathfrak{H}_{A,\mathcal{J}}}^2 \, dt \\
&\quad + \frac{1}{2}\int_0^1 \langle [\rho(t)]_\Lambda^{-1/2}\nabla A(t), [\rho(t)]_\Lambda^{-1/2}\nabla A(t)\rangle_{\mathfrak{H}_{A,\mathcal{J}}}^2 \, dt \\
&\quad - \frac{1}{2}\int_0^1 \langle [\rho(t)]_\Lambda^{1/2}\nabla A(t), [\rho(t)]_\Lambda^{1/2}\nabla A(t)\rangle_{\mathfrak{H}_{A,\mathcal{J}}} \, dt \\
&= \frac{1}{2}\int_0^1 \langle \mathbf{V}(t), [\rho(t)]_\Lambda^{-1}\mathbf{V}(t)\rangle_{\mathfrak{H}_{A,\mathcal{J}}} \, dt,
\end{aligned}
$$

where we used $A \in \mathsf{HJB}_{\Lambda,\varepsilon}$ and $(\rho, \mathbf{V}) \in \mathsf{CE}_\varepsilon(\rho_0, \rho_1)$ for the first inequality and Young's inequality $2|\langle \xi, \eta\rangle| \leq \langle \xi, \xi\rangle + \langle \eta, \eta\rangle$ for the second inequality. $\qquad\square$

To prove actual equality, our crucial tool is the Rockafellar–Fenchel duality theorem (see e.g. [28, Theorem 1.9], which we quote here for the convenience of the reader. Recall that if E is a (real) normed space, the Legendre–Fenchel transform F^* of a proper convex function $F\colon E \to \mathbb{R} \cup \{\infty\}$ is defined by

$$
F^*\colon E^* \to \mathbb{R} \cup \{\infty\}, \quad F^*(x^*) = \sup_{x \in E}(\langle x^*, x\rangle - F(x)).
$$

Theorem 1 *Let E be a real normed space and $F, G\colon E \to \mathbb{R}\cup\{\infty\}$ proper convex functions with Legendre–Fenchel transforms F^*, G^*. If there exists $z_0 \in E$ such that G is continuous at z_0 and $F(z_0), G(z_0) < \infty$, then*

$$
\sup_{z \in E}(-F(z) - G(z)) = \min_{z^* \in E^*}(F^*(z^*) + G^*(-z^*)).
$$

Before we state the main result, we still need the following useful inequality.

Lemma 4 *For any operator connection Λ the map*

$$
f_\Lambda\colon \mathcal{A}_{++} \to B(\mathfrak{H}_A), \quad A \mapsto [A]_\Lambda
$$

is smooth and its Fréchet derivative satisfies

$$
df_\Lambda(B)A \geq f_\Lambda(A)
$$

for $A, B \in \mathcal{A}_{++}$ with equality if $A = B$.

Proof Smoothness of f_Λ is a consequence of the representation theorem of operator connections [21, Theorem 3.4]. For the claim about the Fréchet derivative first note that f_Λ is concave [21, Theorem 3.5]. Therefore $d^2 f_\Lambda(X)[Y, Y] \leq 0$ for all $X \in \mathcal{A}_{++}$ and $Y \in \mathcal{A}_h$ by [18, Proposition 2.2].

The fundamental theorem of calculus implies

$$(df_\Lambda(A) - df_\Lambda(B))(A - B) = \int_0^1 d^2 f_\Lambda(tA + (1-t)B)[A - B, A - B]\, dt$$
$$\leq 0.$$

Since f_Λ is 1-homogeneous by [21, Eq. (2.1)], its derivative is 0-homogeneous. Thus, if we replace B by εB and let $\varepsilon \searrow 0$, we obtain

$$df_\Lambda(A)A \leq df_\Lambda(B)A.$$

Moreover, the 1-homogeneity of f_Λ implies $df_\Lambda(A)A = f_\Lambda(A)$, which settles the claim.

Theorem 2 (Duality formula) *For* $\rho_0, \rho_1 \in \mathfrak{S}_+(\mathcal{A})$ *we have*

$$\frac{1}{2}\mathcal{W}_{\boldsymbol{\Lambda},\boldsymbol{\varepsilon}}(\rho_0, \rho_1)^2 = \sup\{\tau(A(1)\rho_1) - \tau(A(0)\rho_0) : A \in HJB_{\boldsymbol{\Lambda},\boldsymbol{\varepsilon}}\}$$
$$= \sup\{\tau(A(1)\rho_1) - \tau(A(0)\rho_0) : A \in HJB_{\boldsymbol{\Lambda},\boldsymbol{\varepsilon}} \cap C^\infty([0, 1]; \mathcal{A})\}.$$

Proof The second inequality follows easily by mollifying. We will show the duality formula for Hamilton–Jacobi subsolutions in H^1. For this purpose we use the Rockafellar–Fenchel duality formula from Theorem 1.

Let E be the real Banach space

$$H^1([0, 1]; \mathcal{H}_{\mathcal{A}}^{(h)}) \times L^2([0, 1]; \mathcal{H}_{\mathcal{A},\mathcal{J}}^{(h)}).$$

By the theory of linear ordinary differential equations, the map

$$H^1([0, 1]; \mathcal{H}_{\mathcal{A}}^{(h)}) \to \mathcal{H}_{\mathcal{A}}^{(h)} \times L^2([0, 1]; \mathcal{H}_{\mathcal{A}}^{(h)}), \quad A \mapsto (A(0), \dot{A} + \varepsilon \mathscr{L} A)$$

is a linear isomorphism.

Thus the dual space E^* can be isomorphically identified with

$$\mathcal{H}_{\mathcal{A}}^{(h)} \times L^2([0, 1]; \mathcal{H}_{\mathcal{A}}^{(h)}) \times L^2([0, 1]; \mathcal{H}_{\mathcal{A},\mathcal{J}}^{(h)})$$

via the dual pairing

$$\langle (A, \mathbf{V}), (B, C, \mathbf{W}) \rangle = \tau(A(0)B) + \int_0^1 \tau((\dot{A}(t) + \varepsilon \mathscr{L} A(t))C(t))\, dt$$
$$+ \int_0^1 \langle \mathbf{V}(t), \mathbf{W}(t) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}}\, dt.$$

Define functionals $F, G \colon E \longrightarrow \mathbb{R} \cup \{\infty\}$ by

$$F(A, \mathbf{V}) = \begin{cases} -\tau(A(1)\rho_1) + \tau(A(0)\rho_0) & \text{if } \mathbf{V} = \nabla A, \\ \infty & \text{otherwise,} \end{cases}$$

$$G(A, \mathbf{V}) = \begin{cases} 0 & \text{if } (A, \mathbf{V}) \in \mathcal{D}, \\ \infty & \text{otherwise.} \end{cases}$$

Here $\mathcal{D}$ denotes the set of all pairs $(A, \mathbf{V})$ such that

$$\tau((\dot{A}(t) + \varepsilon \mathscr{L} A(t))\rho) + \frac{1}{2}\|\mathbf{V}(t)\|_{\rho}^2 \leq 0$$

for all $t \in [0, 1]$, $\rho \in \mathfrak{S}(\mathcal{A})$.

It is easy to see that F and G are convex. Moreover, for $A_0(t) = -t 1_{\mathcal{A}}$ and $\mathbf{V}_0 = 0$ we have $\mathbf{V}_0 = \nabla A_0$, hence $F(A_0, \mathbf{V}_0) = 0$, and

$$\tau((\dot{A}_0(t) + \varepsilon \mathscr{L} A_0(t))\rho) + \frac{1}{2}\|\mathbf{V}_0(t)\|_{\rho}^2 = -1$$

for all $t \in [0, 1]$, $\rho \in \mathfrak{S}(A)$, hence $G(A_0, \mathbf{V}_0) = 0$. Furthermore, G is clearly continuous at $(A_0, \mathbf{V}_0)$.

Moreover,

$$\sup_{(A, \mathbf{V}) \in E} (-F(A, \mathbf{V}) - G(A, \mathbf{V})) = \sup_{A \in \mathsf{HJB}_{\mathbf{\Lambda}, \varepsilon}(\rho_0, \rho_1)} (\tau(A(1)\rho_1) - \tau(A(0)\rho_0)).$$

Let us calculate the Legendre transforms of F and G, keeping in mind the identification of E^*. For F we obtain

$$F^*(B, C, \mathbf{W}) = \sup_{(A, \mathbf{V}) \in E} \left\{ \langle (A, \mathbf{V}), (B, C, \mathbf{W}) \rangle - F(A, \mathbf{V}) \right\}$$

$$= \sup_A \left\{ \tau(A(0)B) + \int_0^1 \tau((\dot{A}(t) + \varepsilon \mathscr{L} A(t)) C(t)) \, dt \right.$$

$$\left. + \int_0^1 \langle \nabla A(t), \mathbf{W}(t) \rangle_{\mathfrak{H}_{A, \mathcal{J}}} \, dt + \tau(A(1)\rho_1) - \tau(A(0)\rho_0) \right\}.$$

Since the last expression is homogeneous in A, we have $F^*(B, C, \mathbf{W}) = \infty$ unless

$$-\tau(A(1)\rho_1) + \tau(A(0)(\rho_0 - B)) = \int_0^1 \tau((\dot{A}(t) + \varepsilon \mathscr{L} A(t)) C(t)) \, dt$$

$$+ \int_0^1 \langle \nabla A(t), \mathbf{W}(t) \rangle_{\mathfrak{H}_{A, \mathcal{J}}} \, dt$$

for all $A \in H^1([0, 1]; \mathcal{H}_{\mathcal{A}}^{(h)})$.

This implies $C(0) = -(\rho_0 - B)$ and $C(1) = -\rho_1$ and

$$\dot{C}(t) + \operatorname{div} \mathbf{W}(t) = \varepsilon \mathscr{L}^\dagger C(t).$$

Thus

$$F^*(B, C, \mathbf{W}) = \begin{cases} 0 & \text{if } (-C, -\mathbf{W}) \in \mathsf{CE}_\varepsilon''(\rho_0 - B, \rho_1), \\ \infty & \text{otherwise.} \end{cases}$$

Here $\mathsf{CE}_\varepsilon''(\rho_0 - B, \rho_1)$ denotes the set of all pairs $(X, \mathbf{U}) \in H^1((0, 1); \mathcal{H}_{\mathcal{A}}^{(h)}) \times L^2((0, 1); \mathcal{H}_{\mathcal{A}, \mathcal{J}}^{(h)})$ satisfying $X(0) = \rho_0 - B$, $X(1) = \rho_1$ and

$$\dot{X}(t) + \operatorname{div} \mathbf{U}(t) = \varepsilon \mathscr{L}^\dagger X(t).$$

The difference to the definitions of CE (or CE') and CE'' is that we do not make any positivity or normalization constraints. Note however that if $(X, \mathbf{U}) \in \mathsf{CE}''(\rho_0 - B, \rho_1)$, then

$$\frac{d}{dt}\tau(X(t)) = \tau(\varepsilon\mathscr{L}^{\dagger}X(t) - \mathrm{div}\,U(t)) = 0$$

so that $\tau(X(t)) = \tau(\rho_1) = 1$ (and $\tau(B) = 0$).

Now let us turn to the Legendre transform of G. We have

$$G^*(B, C, \mathbf{W}) = \sup_{(A,\mathbf{V})\in E} \left\{ \langle (A, \mathbf{V}), (B, C, \mathbf{W}) \rangle - G(A, \mathbf{V}) \right\}$$

$$= \sup_{(A,\mathbf{V})\in\mathcal{D}} \left\{ \tau(A(0)B) + \int_0^1 \tau((\dot{A}(t) + \varepsilon\mathscr{L}A(t))C(t))\,dt \right.$$

$$\left. + \int_0^1 \langle \mathbf{V}(t), \mathbf{W}(t) \rangle_{\mathfrak{H}_{A,\mathcal{J}}})\,dt \right\}.$$

Since $(A, \mathbf{V}) \in \mathcal{D}$ implies $(A + \lambda\mathrm{id}_{\mathcal{A}}, \mathbf{V}) \in \mathcal{D}$ for all $\lambda \in \mathbb{R}$, we have $G^*(B, C, \mathbf{V}) = \infty$ unless $B = 0$. Furthermore, it follows from the definition of $\mathcal{D}$ that $G^*(0, C, \mathbf{W}) = \infty$ unless $C(t) \geq 0$ for a.e. $t \in [0, 1]$.

For $B = 0$ we have

$$G^*(0, C, \mathbf{W}) = \sup_{(A,\mathbf{V})\in\mathcal{D}} \left\{ \int_0^1 \tau((\dot{A}(t) + \varepsilon\mathscr{L}A(t))C(t))\,dt \right.$$

$$\left. + \int_0^1 \langle \mathbf{V}(t), \mathbf{W}(t) \rangle_{\mathfrak{H}_{A,\mathcal{J}}}\,dt \right\}$$

$$\leq \sup_{(A,\mathbf{V})\in\mathcal{D}} \left\{ -\int_0^1 \frac{1}{2}\|\mathbf{V}(t)\|^2_{C(t)}\,dt \right.$$

$$\left. + \int_0^1 \langle [C(t)]_{\mathbf{\Lambda}}^{1/2}\mathbf{V}(t), [C(t)]_{\mathbf{\Lambda}}^{-1/2}\mathbf{W}(t) \rangle_{\mathfrak{H}_{A,\mathcal{J}}}\,dt \right\}$$

$$\leq \frac{1}{2}\int_0^1 \langle [C(t)]_{\mathbf{\Lambda}}^{-1}\mathbf{W}(t), \mathbf{W}(t) \rangle_{\mathfrak{H}_{A,\mathcal{J}}}\,dt.$$

We will show next that the inequalities are in fact equalities. Let $C^{\delta} = C + \delta$ and $\mathbf{V}^{\delta}(t) = [C(t)^{\delta}]^{-1}\mathbf{W}(t)$. Moreover, let $f_j = f_{\Lambda_j}$ with the notation from Lemma 4. Since

$$\mathcal{H}_{\mathcal{A}}^{(h)} \to \mathbb{R}, \quad B \mapsto \sum_{j\in\mathcal{J}} \langle (df_j(C^{\delta}(t))B)\mathbf{V}_j^{\delta}(t), \mathbf{V}_j^{\delta}(t) \rangle_{\mathfrak{H}_{\mathcal{A}}}$$

is a bounded linear map that depends continuously on t, there exists a unique continuous map $X^{\delta}: [0, 1] \to \mathcal{H}_{\mathcal{A}}^{(h)}$ such that

$$\tau(BX^{\delta}(t)) = \sum_{j\in\mathcal{J}} \langle (df_j(C^{\delta}(t))B)\mathbf{V}_j^{\delta}(t), \mathbf{V}_j^{\delta}(t) \rangle_{\mathfrak{H}_{\mathcal{A}}}$$

for every $B \in \mathcal{H}_{\mathcal{A}}^{(h)}$ and $t \in [0, 1]$.

Let

$$A^{\delta}: [0, 1] \to \mathcal{A}_h, \quad A^{\delta}(t) = -\frac{1}{2}\int_0^t X^{\delta}(s)\,ds.$$

We claim that $(A^\delta, \mathbf{V}^\delta) \in \mathcal{D}$. Indeed,

$$\tau(\dot{A}^\delta(t)\rho) = -\frac{1}{2} \sum_{j \in \mathcal{J}} \langle (df_j(C^\delta(t))\rho)\mathbf{V}_j^\delta(t), \mathbf{V}_j^\delta(t) \rangle_{\mathfrak{H}_\mathcal{A}}$$

$$\leq -\frac{1}{2} \sum_{j \in \mathcal{J}} \langle [\rho]_{\Lambda_j} \mathbf{V}_j^\delta(t), \mathbf{V}_j^\delta(t) \rangle_{\mathfrak{H}_\mathcal{A}}$$

$$= -\frac{1}{2} \|\mathbf{V}^\delta(t)\|_\rho^2,$$

where the inequality follows from Lemma 4. Note that we have equality for $\rho = C^\delta(t)$.

In particular, for $\rho = C(t)$ we obtain

$$\tau(\dot{A}^\delta(t)C(t)) + \langle \mathbf{V}^\delta(t), \mathbf{W}(t) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} \leq \frac{1}{2} \langle [C(t)]_\Lambda^{-1} \mathbf{W}(t), \mathbf{W}(t) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}}.$$

On the other hand,

$$\tau(\dot{A}^\delta(t)C(t)) = -\frac{1}{2} \sum_{j \in \mathcal{J}} \langle (df_j(C^\delta(t))(C^\delta(t) - \delta))\mathbf{V}_j^\delta(t), \mathbf{V}_j^\delta(t) \rangle_{\mathfrak{H}_\mathcal{A}}$$

$$\geq -\frac{1}{2} \langle [C^\delta(t)]_\Lambda \mathbf{V}^\delta(t), \mathbf{V}^\delta(t) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} + \frac{1}{2} \langle [\delta]_\Lambda \mathbf{V}^\delta(t), \mathbf{V}^\delta(t) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}}$$

$$\geq -\frac{1}{2} \langle \mathbf{V}^\delta(t), \mathbf{W}(t) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}},$$

where we again used Lemma 4 for the first inequality.

Put together, we have

$$\frac{1}{2} \langle [C^\delta(t)]_\Lambda^{-1} \mathbf{W}(t), \mathbf{W}(t) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} \leq \tau(\dot{A}^\delta(t)C(t)) + \langle \mathbf{V}^\delta(t), \mathbf{W}(t) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}}$$

$$\leq \frac{1}{2} \langle [C(t)]_\Lambda^{-1} \mathbf{W}(t), \mathbf{W}(t) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}},$$

and

$$\lim_{\delta \searrow 0} \int_0^1 (\tau(\dot{A}^\delta(t)C(t)) + \langle \mathbf{V}^\delta(t), \mathbf{W}(t) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}}) \, dt$$

$$= \frac{1}{2} \int_0^1 \langle [C(t)]_\Lambda^{-1} \mathbf{W}(t), \mathbf{W}(t) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} \, dt$$

follows from the monotone convergence theorem.

Hence

$$G^*(0, C, \mathbf{W}) = \frac{1}{2} \int_0^1 \langle [C(t)]_\Lambda^{-1} \mathbf{W}(t), \mathbf{W}(t) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} \, dt$$

if $C(t) \geq 0$ for a.e. $t \in [0, 1]$. Together with the formula for F^*, we obtain

$$\inf_{(B,C,\mathbf{W}) \in E^*} (F^*(-B, -C, -\mathbf{W}) + G^*(B, C, \mathbf{W}))$$

$$= \inf_{(\rho, \mathbf{V}) \in \mathrm{CE}'_\varepsilon(\rho_0, \rho_1)} \frac{1}{2} \int_0^1 \langle [\rho(t)]_\Lambda^{-1} \mathbf{V}(t), \mathbf{V}(t) \rangle_{\mathfrak{H}_{\mathcal{A},\mathcal{J}}} \, dt$$

$$= \frac{1}{2} \mathcal{W}_{\Lambda,\varepsilon}^2(\rho_0, \rho_1),$$

where the last equality follows from Proposition 1.

An application of the Rockafellar–Fenchel theorem yields the desired conclusion.

References

1. Alicki, R.: On the detailed balance condition for non-Hamiltonian systems. Rep. Math. Phys. **10**(2), 249–258 (1976). https://doi.org/10.1016/0034-4877(76)90046-X

2. Ambrosio, L., Erbar, M., Savaré, G.: Optimal transport, Cheeger energies and contractivity of dynamic transport distances in extended spaces. Nonlinear Anal. **137**, 77–134 (2016). https://doi.org/10.1016/j.na.2015.12.006

3. Becker, S., Li, W.: Quantum statistical learning via quantum Wasserstein natural gradient. J. Stat. Phys. (2021). https://doi.org/10.1007/s10955-020-02682-1

4. Benamou, J.D., Brenier, Y.: A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. Numer. Math. **84**(3), 375–393 (2000). https://doi.org/10.1007/s002110050002

5. Bobkov, S.G., Gentil, I., Ledoux, M.: Hypercontractivity of Hamilton–Jacobi equations. J. Math. Pures Appl. (9) **80**(7), 669–696 (2001). https://doi.org/10.1016/S0021-7824(01)01208-9

6. Brenier, Y., Vorotnikov, D.: On optimal transport of matrix-valued measures. SIAM J. Math. Anal. **52**(3), 2849–2873 (2020). https://doi.org/10.1137/19M1274857

7. Carlen, E.A., 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**(3), 887–926 (2014). https://doi.org/10.1007/s00220-014-2124-8

8. Carlen, E.A., Maas, J.: Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance. J. Funct. Anal. **273**(5), 1810–1869 (2017). https://doi.org/10.1016/j.jfa.2017.05.003

9. Carlen, E.A., Maas, J.: Non-commutative calculus, optimal transport and functional inequalities in dissipative quantum systems. J. Stat. Phys. **178**(2), 319–378 (2020). https://doi.org/10.1007/s10955-019-02434-w

10. Chen, Y., Gangbo, W., Georgiou, T.T., Tannenbaum, A.: On the matrix Monge–Kantorovich problem. Eur. J. Appl. Math. **31**(4), 574–600 (2020). https://doi.org/10.1017/s0956792519000172

11. Chen, Y., Georgiou, T.T., Tannenbaum, A.: Matrix optimal mass transport: a quantum mechanical approach. IEEE Trans. Autom. Control **63**(8), 2612–2619 (2018). https://doi.org/10.1109/tac.2017.2767707

12. Datta, N., Rouzé, C.: Relating relative entropy, optimal transport and Fisher information: a quantum HWI inequality. Ann. Henri Poincaré (2020). https://doi.org/10.1007/s00023-020-00891-8

13. De Palma, G., Trevisan, D.: Quantum optimal transport with quantum channels. Ann. Henri Poincaré (2021). https://doi.org/10.1007/s00023-021-01042-3

14. Duvenhage, R.: Quadratic Wasserstein metrics for von Neumann algebras via transport plans (2020). arXiv:2012.03564

15. Erbar, M., Maas, J., Wirth, M.: On the geometry of geodesics in discrete optimal transport. Calc. Var. Partial Differ. Equ. **58**(1), 19 (2019). https://doi.org/10.1007/s00526-018-1456-1. (**Article 19**)

16. Gangbo, W., Li, W., Mou, C.: Geodesics of minimal length in the set of probability measures on graphs. ESAIM Control Optim. Calc. Var. **25**, 36 (2019). https://doi.org/10.1051/cocv/2018052. (**Paper No. 78**)

17. Golse, F., Mouhot, C., Paul, T.: On the mean field and classical limits of quantum mechanics. Commun. Math. Phys. **343**(1), 165–205 (2016). https://doi.org/10.1007/s00220-015-2485-7

18. Hansen, F.: Operator convex functions of several variables. Publ. Res. Inst. Math. Sci. **33**(3), 443–463 (1997). https://doi.org/10.2977/prims/1195145324

19. Hornshaw, D.F.: L^2-Wasserstein distances of tracial W^*-algebras and their disintegration problem (2018). arXiv:1806.01073

20. Kantorovitch, L.: On the translocation of masses. C. R. (Dokl.) Acad. Sci. USSR (N.S.) **37**, 199–201 (1942)

21. Kubo, F., Ando, T.: Means of positive linear operators. Math. Ann. **246**(3), 205–224 (1980). https://doi.org/10.1007/BF01371042

22. 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). https://doi.org/10.1007/s10955-017-1756-4

23. Ning, L., Georgiou, T.T., Tannenbaum, A.: On matrix-valued Monge–Kantorovich optimal mass transport. IEEE Trans. Autom. Control **60**(2), 373–382 (2015). https://doi.org/10.1109/TAC.2014.2350171

24. Otto, F., Villani, C.: Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. J. Funct. Anal. **173**(2), 361–400 (2000). https://doi.org/10.1006/jfan.1999.3557

25. Palma, G.D., Marvian, M., Trevisan, D., Lloyd, S.: The quantum Wasserstein distance of order 1 (2020). arXiv:2009.04469

26. Peyré, G., Chizat, L., Vialard, F.X., Solomon, J.: Quantum entropic regularization of matrix-valued optimal transport. Eur. J. Appl. Math. **30**(6), 1079–1102 (2019). https://doi.org/10.1017/s0956792517000274

27. Rouzé, C., Datta, N.: Concentration of quantum states from quantum functional and transportation cost inequalities. J. Math. Phys. **60**(1), 012202, 22 (2019). https://doi.org/10.1063/1.5023210

28. Villani, C.: Topics in Optimal Transportation. Graduate Studies in Mathematics, vol. 58. American Mathematical Society, Providence (2003). https://doi.org/10.1007/b12016

29. Villani, C.: Optimal transport. Old an new. In: Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences), vol. 338. Springer, Berlin (2009). https://doi.org/10.1007/978-3-540-71050-9

30. Wirth, M.: A Noncommutative Transport Metric and Symmetric Quantum Markov Semigroups as Gradient Flows of the Entropy (2018). arXiv:1808.05419

31. Wirth, M., Zhang, H.: Complete gradient estimates of quantum Markov semigroups (2020). arXiv:2007.13506