Four Lectures (and Some Bonus Material) on Quantum Optimal Transport

Melchior Wirth

December 15, 2025

Chapter 1

The Classical Optimal Transport Problem

The theory of optimal transport goes back to the work of Gaspard Monge in the 18th century, who introduced the optimal transport problem, and the work of Leonid Kantorovich in the early 20th century, who initiated the theory of linear programming and formulated a relaxed version of the optimal transport that fits into the framework of linear optimization. The applications of optimal transport theory are manifold, from the transport of troops to the front, one of Monge's original motivations to study the problem, to the optimal allocation of resources in an economy, the topic that earned Kantorvich the Nobel Memorial Prize in Economic Sciences (he was the only Soviet scientist wo win this prize), and many other more recent fields like computer vision. Here we focus on setting up the basic mathematical framework of optimal transport theory and Wasserstein distances with a view towards our later study of quantum optimal transport.

Let $\Omega \subset \mathbb{R}^d$ be a measurable subset and let μ, ν be Borel probability measures on Ω. A *transport plan* from μ to ν is a probability measure π on $\Omega\times\Omega$ with marginals μ and ν, that is, $\pi(A\times X) = \mu(A)$ and $\pi(X\times B) = \nu(B)$ for all Borel sets $A, B \subset \Omega$. Let us denote the set of all transport plans from μ to ν by $\Gamma(\mu, \nu)$. This set is always non-empty since the product measure $\mu \otimes \nu$ is a transport plan.

If $c\colon \Omega\times\Omega \to [0, \infty]$ is a lower continuous function, the optimal transport problem is the optimization problem

$$\begin{cases} \int_{\Omega\times\Omega} c(x, y)\, d\pi(x, y) \to \min, \\ \pi \in \Gamma(\mu, \nu). \end{cases}$$

This is a convex optimization problem (affine objective function, convex constraint). Existence of minimizers is not hard to show since $\Gamma(\mu, \nu)$ is compact (tight and closed in the weak topology of Borel probability measures) and the objective function is lower semicontinuous.

Theorem 1.1. *The functional* $\pi \mapsto \int_{\Omega\times\Omega} c(x,y)\,d\pi(x,y)$ *attains its minimum on* $\Gamma(\mu,\nu)$.

Transport plans have several equivalent descriptions, which will be useful when moving to quantum optimal transport:

- Couplings: If $\pi \in \Gamma(\mu,\nu)$, then there exist a random variables X, Y with values in Ω such that $X \sim \mu$, $Y \sim \nu$ and $(X,Y) \sim \pi$.

  Conversely, if X, Y are random variables with values in Ω such that $X \sim \mu$, $Y \sim \nu$, then the joint law of (X,Y) is a transport plan from μ to ν.

- Markov kernels: If $\pi \in \Gamma(\mu,\nu)$, then there exists a Markov kernel $k\colon \Omega\times\mathcal{B}(\Omega) \to [0,1]$ such that $\pi(A\times B) = \int_A k(x,B)\,d\mu(x)$ for $A, B \in \mathcal{B}(\Omega)$.

  Conversely, if $k\colon \Omega\times\mathcal{B}(\Omega) \to [0,1]$ is a Markov kernel such that $\mu(A) = \int_A k(x,\Omega)\,d\mu(x)$ and $\nu(B) = \int_\Omega k(x,B)\,d\mu(x)$ for all Borel sets $A, B \in \Omega$, then the measure π defined by $\pi(A,B) = \int_A k(x,B)\,d\mu(x)$ is a transport plan from μ to ν.

- Markov maps: If $k\colon \Omega\times\mathcal{B}(\Omega) \to [0,1]$ is a Markov kernel such that $\mu(A) = \int_A k(x,\Omega)\,d\mu(x)$ and $\nu(B) = \int_\Omega k(x,B)\,d\mu(x)$ for all Borel sets $A, B \in \Omega$, then the map $\Phi\colon L^\infty(\Omega,\nu) \to L^\infty(\Omega,\mu)$, $\Phi(f)(x) = \int_\Omega f(y)k(x,dy)$ is linear, unital ($\Phi(1) = 1$), positive ($f \geq 0$ implies $\Phi(f) \geq 0$) and satisfies $\int_\Omega \Phi(f)\,d\mu = \int_\Omega f\,d\nu$ for all $f \in L^\infty(\Omega,\nu)$. A linear unital and positive map is called a *Markov map* or *channel*.

  Conversely, every Markov map $\Phi\colon L^\infty(\Omega,\nu) \to L^\infty(\Omega,\mu)$ such that $\int \Phi(\cdot)\,d\mu = \int \cdot\,\nu$ is of the form $\Phi(f)(x) = \int_\Omega f(y)k(x,dy)$ with a Markov kernel k that satisfies the marginal constraints from before.

- Hilbert bimodules/correspondences: If $\pi \in \Gamma(\mu,\nu)$, then $L^2(\Omega\times\Omega,\pi)$ becomes a $L^\infty(\mu)$-$L^\infty(\nu)$ bimodule with the left and right action given by $(f\eta g)(x,y) = f(x)\eta(x,y)g(y)$. The left and right action are $*$-homomorphisms, i.e., $\langle f\xi g,\eta\rangle_2 = \langle \xi, \overline{f}\eta\overline{g}\rangle_2$ for all $\xi,\eta \in L^2(\pi)$ and $f \in L^\infty(\mu)$, $g \in L^\infty(\nu)$, and they are weak*-continuous, i.e., if (f_n) is a bounded sequence in $L^\infty(\mu)$ such that $f_n \to f$ weak* (resp. (g_n) is a bounded sequence in $L^\infty(\nu)$ such that $g_n \to g$ weak*), then $\langle \xi, f_n\eta\rangle \to \langle \xi, f\eta\rangle$ (resp. $\langle \xi, \eta g_n\rangle \to \langle \xi, \eta g\rangle$) for all $\xi,\eta \in L^2(\pi)$. A Hilbert space with the structure of an $L^\infty(\mu)$-$L^\infty(\nu)$ bimodule so that the left and right action are weak*-continuous $*$-homomorphisms is also called a *Hilbert* $L^\infty(\mu)$-$L^\infty(\nu)$ *bimodule* or a *correspondence* from $L^\infty(\nu)$ to $L^\infty(\mu)$. Furthermore, if we let $\xi_0 = \mathbb{1}_\Omega$, then $\pi(A \times B) = \langle \xi_0, \mathbb{1}_A \xi_0 \mathbb{1}_B \rangle$ for all Borel sets $A, B \in \Omega$.

Conversely, if $\mathcal{H}$ is a Hilbert $L^\infty(\mathbb{R}^d, \mu)$-$L^\infty(\mathbb{R}^d, \nu)$ bimodule and $\xi_0 \in \mathcal{H}$ is a unit vector such that $\langle \xi_0, \mathbb{1}_A \xi_0 \rangle = \mu(A)$, $\langle \xi_0, \xi_0 \mathbb{1}_B \rangle = \nu(B)$ for all Borel sets $A, B \subset \mathbb{R}^d$, then $\pi(A \times B) = \langle \xi_0, \mathbb{1}_A \xi_0 \mathbb{1}_B \rangle$ defines a transport plan from μ to ν.

Exercise 1.2. Express the optimal transport problem in terms of couplings, Markov kernels, Markov maps (for bounded costs) and Hilbert bimodules.

In this course, we are primarily interested in the case when c is derived from the Euclidean distance, more precisely, when $c(x, y) = |x - y|^2$.

Definition 1.3 (Wasserstein distance)**.** A Borel probability measure μ on Ω is said to have *finite second moments* if $\int_\Omega |x|^2 \, d\mu(x) < \infty$. We denote the set of all Borel probability measures on Ω with finite second moments by $\mathcal{P}_2(\Omega)$.

The 2-*Wasserstein distance* W_2 is defined by

$$W_2 \colon \mathcal{P}_2(\Omega) \times \mathcal{P}_2(\Omega) \to [0, \infty),\ (\mu, \nu) \mapsto \inf_{\pi \in \Gamma(\mu,\nu)} \left(\int_{\Omega \times \Omega} |x - y|^2 \, d\pi(x, y) \right)^{1/2}.$$

Remark 1.4. With the different models of transport plans discussed above, we can equivalently write the 2-Wasserstein distance (for bounded Ω) as

$$\begin{aligned} W_2(\mu, \nu)^2 &\overset{(1)}{=} \inf \mathbb{E}[|X - Y|^2] \\ &\overset{(2)}{=} \inf \sum_{j=1}^d \int |\mathrm{pr}_j - \Phi(\mathrm{pr}_j)|^2 \, d\mu(x) \\ &\overset{(3)}{=} \inf \sum_{j=1}^d \|\mathrm{pr}_j \cdot \xi_0 - \xi_0 \cdot \mathrm{pr}_j\|_{\mathcal{H}}^2, \end{aligned}$$

where the infimum is taken over

(1) all random variables X, Y with $X \sim \mu$, $Y \sim \nu$,

(2) all Markov maps $\Phi \colon L^\infty(\Omega, \nu) \to L^\infty(\Omega, \mu)$ such that $\int_\Omega \Phi(\cdot) \, d\mu = \int_\Omega \cdot \, d\nu$,

(3) all pairs $(\mathcal{H}, \xi_0)$ with a Hilbert $L^\infty(\Omega, \mu)$-$L^\infty(\Omega, \nu)$ bimodule $\mathcal{H}$ and a vector $\xi_0 \in \mathcal{H}$ such that $\langle \xi, f \cdot \xi \rangle = \int_\Omega f \, d\mu$ and $\langle \xi, \xi \cdot g \rangle = \int_\Omega g \, d\nu$.

In (2) and (3), pr_j denotes the coordinate function $x \mapsto x_j$ on Ω.

The Wasserstein distance is indeed a metric. The non-trivial part is the triangle inequality, which is typically proven using the notion of gluing of transport plans. This becomes more transparent in terms of Markov maps or Markov kernels: If Φ_{12} is the Markov map associated with $\pi_{12} \in \Gamma(\mu_1, \mu_2)$ and Φ_{23} is the Markov map associated with $\pi_{23} \in \Gamma(\mu_2, \mu_3)$, then the gluing of π_{12} and π_{23} is the transport plans associated with $\Phi_{12} \circ \Phi_{23}$.

In terms of (pointed) Hilbert bimodules, the gluing corresponds to the relative tensor product $(\mathcal{H}_{12} \otimes_{L^\infty(\mu_2)} \mathcal{H}_{23}, \xi_{12} \otimes_{L^\infty(\mu_2)} \xi_{23})$, which we will be useful for the quantum part of this course.

Proposition 1.5. *The Wasserstein distance is a metric on $\mathcal{P}_2(\Omega)$.*

Proof. If $W_2(\mu, \nu) = 0$, then there exists $\pi \in \Gamma(\mu, \nu)$ such that $\int |x - y|^2 \, d\pi(x, y) = 0$. Hence π is supported on the diagonal $\{(x, x) \mid x \in \Omega\}$, which implies $\mu(A) = \pi(A \times \Omega) = \pi(\Omega \times A) = \nu(A)$ for all Borel sets $A \subset \Omega$.

We only show the triangle inequality for bounded Ω here, using Markov maps as models for transport plans. Let μ_1, μ_2 and μ_3 be Borel probability measures on Ω and let $\Phi_{32} \colon L^\infty(\mu_3) \to L^\infty(\mu_2)$, $\Phi_{21} \colon L^\infty(\mu_2) \to L^\infty(\mu_1)$ be Markov maps such that $\int \Phi_{32}(\cdot) \, d\mu_2 = \int \cdot \, d\mu_3$, $\int \Phi_{21}(\cdot) \, d\mu_1 = \int \cdot \, d\mu_2$. Clearly, $\Phi_{31} = \Phi_{21} \circ \Phi_{32}$ is a Markov map and $\int \Phi_{31}(\cdot) \, d\mu_1 = \int \cdot \, d\mu_3$. Moreover,

$$|\Phi_{31}(\mathrm{pr}_j) - \mathrm{pr}_j| = |\Phi_{21}(\Phi_{32}(\mathrm{pr}_j) - \mathrm{pr}_j) + (\Phi_{21}(\mathrm{pr}_j) - \mathrm{pr}_j)|$$

implies

$$\begin{aligned}
&\left(\int |\Phi_{31}(\mathrm{pr}_j) - \mathrm{pr}_j|^2 \, d\mu_1\right)^{1/2} \\
&\quad \leq \left(\int |\Phi_{21}(\Phi_{32}(\mathrm{pr}_j) - \mathrm{pr}_j)| \, d\mu_1\right)^{1/2} + \left(\int |\Phi_{21}(\mathrm{pr}_j) - \mathrm{pr}_j|^2 \, d\mu_1\right)^{1/2} \\
&\quad = \left(\int |\Phi_{32}(\mathrm{pr}_j) - \mathrm{pr}_j| \, d\mu_2\right)^{1/2} + \left(\int |\Phi_{21}(\mathrm{pr}_j) - \mathrm{pr}_j|^2 \, d\mu_1\right)^{1/2}. \qquad \square
\end{aligned}$$

Example 1.6 (Wasserstein distance between Gaussian measures). Note that the transport cost $\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \, d\pi(x, y)$ depends only on the second moments of π. This allows to compute the Wasserstein distance between Gaussian measures. Let us assume that $d = 1$ for simplicity and let μ_1, μ_2 be Gaussian measures with means m_1, m_2 and variances σ_1^2, σ_2^2. We will show that $W_2(\mu_1, \mu_2)^2 = (m_2 - m_1)^2 + (\sigma_2 - \sigma_1)^2$.

If $\tilde{\pi} \in \Gamma(\mu_1, \mu_2)$, one can always find a Gaussian measure π on $\mathbb{R}^2$ with the same mean and covariance matrix as $\tilde{\pi}$, which is necessarily a transport plan from μ_1 to μ_2. In particular, the mean of π is $\begin{pmatrix} m_1 \\ m_2 \end{pmatrix}$ and its covariance matrix is $\begin{pmatrix} \sigma_1^2 & c \\ c & \sigma_2^2 \end{pmatrix}$. Positive semi-definiteness of the covariance matrix implies $c^2 \leq \sigma_1^2 \sigma_2^2$. Thus

$$\int_{\mathbb{R}^2} |x-y|^2 \, d\pi(x,y) = \sigma_1^2 + m_1^2 + \sigma_2^2 + m_2^2 - 2(c + m_1 m_2) \geq (m_2 - m_1)^2 + (\sigma_2 - \sigma_1)^2$$

with equality if $c = \sigma_1 \sigma_2$.

In this case, we can explicitly write down an optimal coupling: Let $T \colon \mathbb{R} \to \mathbb{R}$, $x \mapsto m_2 + \frac{\sigma_2}{\sigma_1}(x - m_1)$. If $X \sim \mathcal{N}(m_1, \sigma_1^2)$, then $T(X) \sim \mathcal{N}(m_2, \sigma_2^2)$ and $T(X) - X \sim \mathcal{N}(m_2 - m_1, (\sigma_2 - \sigma_1)^2)$. Hence

$$\mathbb{E}[|X - T(X)|^2] = (m_2 - m_1)^2 + (\sigma_2 - \sigma_1)^2.$$

There is another, "dynamical" approach to optimal transport distances based on what is known as Benamou–Brenier formula.

Theorem 1.7 (Benamou–Brenier)**.** *Let $\Omega \subset \mathbb{R}^d$ be open, convex and bounded. If μ, ν are Borel probability measures on Ω, then*

$$W_2(\mu, \mu)^2 = \inf \int_0^1 \int_\Omega |v_t|^2 \, d\mu_t \, dt,$$

where the infimum is taken over all curves of measures $(\mu_t)_{t \in [0,1]}$ and all vector fields $v \colon [0,1] \times \Omega \to \mathbb{R}^d$ that satisfy the continuity equation

$$\dot{\mu}_t + \nabla \cdot (v_t \mu_t) = 0$$

in a weak sense and $\mu_0 = \mu$, $\mu_1 = \nu$.

Remark 1.8. Here is a heuristic formulation of the Benamou–Brenier formula due to Otto: Among the vector fields that satisfy the continuity equation for a given curve (μ_t), there is a unique gradient vector field $(\nabla \varphi_t)$, and this also minimizes $\int_0^1 \int_\Omega |v_t|^2 \, d\mu_t \, dt$. Let us write $\varphi_t = \mathcal{K}_{\mu_t}(\dot{\mu}_t)$. With this notation, we have

$$W_2^2(\mu, \nu) = \inf \left\{ \int_0^1 \underbrace{\int_\Omega |\mathcal{K}_{\mu_t}(\dot{\mu}_t)|^2 \, d\mu_t}_{g_{\mu_t}(\dot{\mu}_t, \dot{\mu}_t)} \, dt : \mu_0 = \mu, \mu_1 = \nu \right\}.$$

The quantity $g_{\mu_t}(\dot{\mu}_t, \dot{\mu}_t)$ is quadratic in $\dot{\mu}_t$, hence we can formally view as a metric tensor on $\mathcal{P}_2(\Omega)$ and W_2 is the distance induced by the metric g. This way, we can treat $\mathcal{P}_2(\Omega)$ like a Riemannian manifold. In particular, there is a form of differential calculus.

Chapter 2

Quantum States and Quantum Channels

Let H be a complex Hilbert space, $\mathbb{B}(H)$ the set of bounded linear operators on H and $S^1(H)$ the set of trace-class operators on H. The (often unbounded) self-adjoint operators on H play the roles of observables in quantum mechanics.

A *state* on $\mathbb{B}(H)$ is a linear functional $\varphi\colon \mathbb{B}(H) \to \mathbb{C}$ such that $\varphi(1) = \|\varphi\| = 1$. A state is called *normal* if $\varphi(\sum_{j=1}^\infty p_j) = \sum_{j=1}^\infty \varphi(p_j)$ for every sequence (p_j) of pairwise orthogonal projections in $\mathbb{B}(H)$. One can show that φ is a normal state if and only if there exists a positive operator $\rho \in S^1(H)$ with trace 1 such that $\varphi = \mathrm{tr}(\,\cdot\,\rho)$. Such an operator ρ is called a *density operator*.

The operational interpretation of quantum states and observables is as follows: If the quantum mechanical system is in state φ, then the probability of measuring a value in the interval $[a, b]$ for the observable T is $\varphi(\mathbb{1}_{[a,b]}(T))$. This is known as *Born's rule* in quantum mechanics. In this way, every state encodes a whole family of probability measures, one for each observable. In particular, this is an intrinsically probabilistic theory: The state of the system only determines a probability distribution for the measurement outcome of every observable, not a deterministic measurement outcome.

Example 2.1 (Pure and mixed states). If $\xi \in H$ is a unit vector, then $T \mapsto \langle \xi, T\xi \rangle$ is a normal state. States of this form are called *vector states* or *pure states*. This is the setting of classical quantum mechanics, where the time evolution of the vector ξ is determined by the Schrödinger equation. Note that ξ and $\lambda\xi$ for $|\lambda| = 1$ define the same vector state, which is consistent to the notion in quantum mechanics that states are indistinguishable if they only differ by a phase.

A state that is not pure is called a *mixed state*. Such states occur naturally as statistical mixtures of pure states (not to be confused with the superposition of pure states, which is again a pure state), i.e., they can be written as $\rho = \sum_{k=1}^{\infty} \lambda_k |\xi_k\rangle \langle\xi_k|$ for some non-negative sequence (λ_k) in ℓ^1 which sums to 1 and pairwise orthogonal unit vectors ξ_k, $k \in \mathbb{N}$.

An important class of states in quantum mechanics are Gaussian states. As many objects in quantum physics, they come in a Bosonic and a Fermionic version. Here, we only discuss the Bosonic version, which is a bit more popular in quantum optimal transport and related fields, especially in quantum optics and quantum communication (since photons are bosons).

Example 2.2. On $L^2(\mathbb{R}^d)$ consider the position and momentum operators $R_{2j}f(x) = x_j f(x)$ and $R_{2j-1}f(x) = -i\frac{\partial}{\partial x_j} f(x)$ for $j \in \{1, \dots, d\}$, which are (unbounded) self-adjoint operators on their natural domains. On a suitable dense domain[1], they satisfy the *canonical commutation relations (CCR)*

$$[R_{2j}, R_{2k-1}] = i\delta_{j,k} I.$$

Let $\mathbf{R} = (R_1, \dots, R_{2d})^\top$. Up to an additive constant, a Hamiltonian that is quadratic in position and momentum operators can be written as $H = \frac{1}{2}(\mathbf{R} - m)^\top \Sigma (\mathbf{R} - m)$ for a symmetric matrix $\Sigma \in M_{2d}(\mathbb{R})$ and a vector $m \in \mathbb{R}^d$. Informally, a Gaussian state is a Gibbs state of a quadratic Hamiltonian with Σ positive definite, that is, $\varphi = \mathrm{tr}(\,\cdot\,\rho_\beta)$ with

$$\rho_\beta = \frac{e^{-\beta H}}{\mathrm{tr}(e^{-\beta H})}.$$

To better handle degenerate cases (like ground states of quadratic Hamiltonians), it is more convenient to define Gaussian states via the characteristic function of Bosonic states. For that purpose, define the generalized Weyl operator or displacement operator as $W(z) = \exp\big(-i\sqrt{2}\sum_{j=1}^{d}(\operatorname{Re} z_j)R_{2j-1} - (\operatorname{Im} z_j)R_{2j}\big)$ for $z \in \mathbb{C}^d$. As a consequence of the CCR, the Weyl operators obey the so-called *Weyl relations*

$$W(z)W(w) = e^{-i\operatorname{Im}\langle z,w\rangle} W(z+w)$$

for $z, w \in \mathbb{C}^d$.

If φ is a normal state on $\mathbb{B}(L^2(\mathbb{R}^d))$, its *characteristic function* is defined as $\hat{\varphi}(z) = \varphi(W(z))$. The state φ is called *Gaussian* or *quasi-free* if there exists a positive definite matrix $\Sigma \in M_{2d}(\mathbb{R})$ and a vector $m \in \mathbb{R}^{2d}$ such that

$$\hat{\varphi}(z) = \exp\left(-\frac{1}{2}\operatorname{Re}\langle z, \Sigma z\rangle - i\operatorname{Re}\langle z, m\rangle\right).$$

[1]The standard choice is the set of all $f \in L^2(\mathbb{R}^d)$ such that $f(x) = p(x)e^{-x^2/2}$ with a polynomial p. On this set, the position and momentum operators are all *essentially self-adjoint*.

Here we identify $\mathbb{C}^d$ with $\mathbb{R}^{2d}$ to make sense of the products Σz and $\langle z, m\rangle$ (note that Σ is not necessarily complex linear).

As their name suggests, states describe the physical state of a quantum system. The possible transformations the state of a system can possibly undergo are described by quantum channels. If one assumes that every quantum channel between systems A and B extends linearly to a map Φ from $S^1(H_A)$ to $S^1(H_B)$, then two requirements are clear: The map Φ should preserve positive operators and it should preserve the trace of an operator in order to map density operators to density operators. Such maps are called *positive trace-preserving* (ptp) maps.

However, somewhat surprisingly, these conditions are not enough to ensure a physical evolution. There is a stronger positivity requirement that becomes apparent if one couples the systems to an auxiliary system or environment E. If the transformation of system A to B is described by Φ, then the transformation of the total system AE to BE is described by $\Phi\otimes \mathrm{id}_{S^1(H_E)}$, and this map may fail to be positive even if Φ is positive. If $\Phi \otimes \mathrm{id}_{S^1(H_E)}$ is positive for every Hilbert space H_E, then Φ is called *completely positive.* A completely positive trace-preserving (cptp) is called a *quantum channel.*

So far, we have described the *Schrödinger picture*, in which the state of the system evolves and the observables remain constant. Their is a dual version, called the *Heisenberg picture*, in which the state remains unchanged and the observables evolve. In the Heisenberg picture, quantum channels are described by normal unital completely positive (ucp) maps.

The duality between channels in the Schrödinger and Heisenberg picture is given by the adjoint in the following sense.

Proposition 2.3 (Heisenberg–Schrödinger duality)**.** *If* $\Phi\colon S^1(H_A) \to S^1(H_B)$ *is cptp map, then its adjoint* $\Phi^\dagger\colon \mathbb{B}(H_B) \to \mathbb{B}(H_A)$ *is a normal ucp map. Conversely, if* $\Psi\colon \mathbb{B}(H_B) \to \mathbb{B}(H_A)$ *is a normal ucp map, then its adjoint* $\Psi^\dagger$ *maps* $S^1(H_A)$ *to* $S^1(H_B)$ *and the restriction of* $\Psi^\dagger$ *to* $S^1(H_A)$ *is cptp.*

Example 2.4 (Unitary channels). If $U \in \mathbb{B}(H_B; H_A)$ is unitary, then $\Phi(\rho) = U^*\rho U$ defines a quantum channel from $S^1(H_A)$ to $S^1(H_B)$. Channels of this form are called *unitary channels.* In particular, if $H_A = H_B$ and h is a (possibly unbounded) self-adjoint operator on H_A, then $\Phi_t(\rho) = e^{-iht}\rho e^{iht}$ is a unitary channel. This channel describes the time evolution of a *closed* quantum system with Hamiltonian h according to the *von Neumann equation.*

Example 2.5 (Gaussian channels). If our system Hilbert space is $H_S = L^2(\mathbb{R}^d)$ and we have an environment Hilbert space $H_E = L^2(\mathbb{R}^m)$, let $U = \exp(\frac{i}{2}\sum_{j,k=1}^{2(d+m)} \mathbf{R}^\top M\mathbf{R}) \in \mathbb{B}(H_E\otimes H_S)$ for a symmetric matrix $M \in M_{2(d+m)}(\mathbb{R})$. The operator U is unitary, and unitaries of this form (complex exponentials of quadratic Hamiltonians in position and momentum operator) are called *Gaussian unitaries.* If φ_E is a Gaussian state on $\mathbb{B}(H_E)$ with density operator ρ_E, then

$$\Phi\colon S^1(H_S) \to S^1(H_S),\ \rho \mapsto \mathrm{tr}_E(U(\rho \otimes \rho_E)U^*)$$

is a quantum channel. Quantum channels of this form are called *Gaussian channels.* They map Gaussian states to Gaussian states. Possibly enlarging H_E, one can always achieve that ρ_E is a Gaussian state with mean 0 and covariance I.

Example 2.6 (Thermal noise channel, lossy channel). Let $H_S = H_E = L^2(\mathbb{R}^d)$. The *beam splitter* with transmissivity $\lambda \in [0,1]$ is the unitary

$$U_\lambda\colon H_S \otimes H_E \to H_S \otimes H_E,\ U_\lambda = \exp\left(\arccos\sqrt{\lambda}\sum_{j=1}^{d}(a_j^* \otimes a_j - a_j \otimes a_j^*)\right),$$

where $a_j = \frac{1}{2}(R_{2j-1} + iR_{2j})$ is the annihilation operator (its adjoint a_j^* is called creation operator). This unitary acts on Weyl operators as $U_\lambda(W(z) \otimes W(w))U_\lambda^* = W(\sqrt{\lambda}z + \sqrt{1-\lambda}w) \otimes W(-\sqrt{\lambda}z + \sqrt{1-\lambda}w)$.

If (e_n) is an orthonormal eigenbasis of the number operator $N = \sum_{j=1}^d a_j a_j^*$ and for $c \geq 1$ we are given the thermal Gibbs state

$$\rho_E = \frac{2}{c+1}\sum_{n=0}^{\infty}\left(\frac{c-1}{c+1}\right)^n |e_n\rangle\langle e_n|$$

with mean photon number $\frac{c-1}{2}$, then the channel

$$\Phi\colon S^1(H_S) \to S^1(H_S),\ \rho \mapsto \mathrm{tr}_E(U_\lambda(\rho \otimes \rho_E)U_\lambda^*)$$

is called a *thermal noise channel.* In the limit case $c = 1$, one obtains the *lossy channel* which describes photon loss with rate $1 - \lambda$.

The general form of a ucp maps between arbitrary C^*-algebras is given by Stinespring's dilation theorem. For the special case of quantum channels considered here, it takes the following very concrete form.

Theorem 2.7 (Kraus)**.** *A map $\Phi\colon S^1(H_A) \to S^1(H_B)$ is a quantum channel if and only if there exists a family $(v_i)_{i\in I}$ in $\mathbb{B}(H_A; H_B)$ such that $\sum_{i\in I} v_i^* v_i = I_A$ in the strong operator topology and*

$$\Phi(\rho) = \sum_{i\in I} v_i \rho v_i^*$$

for all $\rho \in S^1(H_A)$.

Theorem 2.8 (Paschke GNS construction)**.** *If $\Psi\colon S^1(H_A) \to S^1(H_B)$ is a quantum channel and $\rho_{A/B} \in S^1(H_{A/B})$ are non-singular density matrices such that $\Phi(\rho_A) = \rho_B$, then there exists a Hilbert $\mathbb{B}(H_B)$-$\mathbb{B}(H_A)$ bimodule $\mathcal{H}$ and a unit vector $\xi_0 \in \mathcal{H}$ such that $\operatorname{tr}(x\rho_A^{1/2}\Phi^\dagger(y)\rho_A^{1/2}) = \langle \xi_0, y\xi_0 x\rangle$ for all $x \in \mathbb{B}(H_A)$, $y \in \mathbb{B}(H_B)$.*

Conversely, if $\mathcal{H}$ is a Hilbert $\mathbb{B}(H_B)$-$\mathbb{B}(H_A)$ bimodule and $\xi_0 \in \mathcal{H}$ is a unit vector such that $\langle \xi_0, \xi_0 x\rangle = \operatorname{tr}(x\rho_A)$, $\langle \xi_0, y\xi_0\rangle = \operatorname{tr}(y\rho_B)$ for all $x \in \mathbb{B}(H_A)$, $y \in \mathbb{B}(H_B)$, then there exists a unique quantum channel $\Phi\colon S^1(H_A) \to S^1(H_B)$ such that $\langle \xi_0, y\xi_0 x\rangle = \operatorname{tr}(x\rho_A^{1/2}\Phi^\dagger(y)\rho_A^{1/2})$ for all $x \in \mathbb{B}(H_A)$, $y \in \mathbb{B}(H_B)$.

Proof. Define a sesquilinear form on the algebraic tensor product $\mathbb{B}(H_B) \odot \mathbb{B}(H_A)$ by

$$\langle y_1 \otimes x_1, y_2 \otimes x_2\rangle = \operatorname{tr}(x_1^*\rho_A^{1/2}\Phi^\dagger(y_1^* y_2)\rho_A^{1/2}x_2).$$

Complete positivity of $\Phi^\dagger$ ensures that this sesquilinear form is positive semi-definite. Let $\mathcal{H}$ denote the Hilbert space obtained after separation and completion with respect to this sesquilinear form. It is not hard to check that the actions

$$b(y \otimes x)a = by \otimes xa$$

extend continuously to $\mathcal{H}$ and that $\mathcal{H}$ with these actions is a Hilbert $\mathbb{B}(H_B)$-$\mathbb{B}(H_A)$ bimodule. Moreover, if we let ξ_0 denote the image of $I_B \otimes I_A$ in $\mathcal{H}$, then

$$\langle \xi_0, y\xi_0 x\rangle = \langle I_B \otimes I_A, y \otimes x\rangle = \operatorname{tr}(\rho_A^{1/2}\Phi^\dagger(y)\rho_A^{1/2}x) = \operatorname{tr}(x\rho_A^{1/2}\Phi^\dagger(y)\rho_A^{1/2}).$$

For the converse implication, note that

$$\|\xi_0 x\|^2 = \langle \xi_0, \xi_0 x x^*\rangle = \operatorname{tr}(xx^*\rho_A) = \|\rho_A^{1/2}x\|_2^2.$$

Since $\rho_A^{1/2}\mathbb{B}(H_A)$ is dense in $S^2(H_A)$, it follows that there exists a unique bounded linear operator $L(\xi_0)\colon S^2(H_A) \to \mathcal{H}$ such that $L(\xi_0)\rho_A^{1/2}x = \xi_0 x$ for all $x \in \mathbb{B}(H_A)$. If we define $\Psi(y) = L(\xi_0)^* y L(\xi_0)$, then

$$\operatorname{tr}(x\rho_A^{1/2}\Psi(y)\rho_A^{1/2}) = \langle L(\xi_0)\rho_A^{1/2}x^*, yL(\xi_0)\rho_A^{1/2}\rangle_2 = \langle \xi_0 x^*, y\xi_0\rangle = \langle \xi_0, y\xi_0 x\rangle.$$

Clearly, Ψ is a normal unital positive map. If H_E is a finite-dimensional Hilbert space, then $\mathcal{H} \otimes S^2(H_E)$ is a Hilbert $\mathbb{B}(H_B \otimes H_S)$-$\mathbb{B}(H_A \otimes H_S)$ bimodule, and if we let $\eta_0 = \xi_0 \otimes 1_{H_S}$, then $\Psi \otimes \mathrm{id}_{\mathbb{B}(H_S)} = L(\eta_0)^* \cdot L(\eta_0)$. Thus Ψ is completely positive. By the Schrödinger–Heisenberg duality, there exists a unique quantum channel $\Phi\colon S^1(H_A) \to S^1(H_B)$ such that $\Psi = \Phi^\dagger$. □

If $\varphi = \operatorname{tr}(\,\cdot\,\rho)$ is a normal state on $\mathbb{B}(H)$ with ρ non-singular, then its *modular group* σ^φ is defined by

$$\sigma_t^\varphi \colon \mathbb{B}(H) \to \mathbb{B}(H),\, x \mapsto \rho^{it} x \rho^{-it}.$$

For every $t \in \mathbb{R}$, the map σ_t^φ is a unitary channels (in the Heisenberg picture) on $\mathbb{B}(H)$ and the map $t \mapsto \sigma_t^\varphi(x)$ is weak* continuous for every $x \in \mathbb{B}(H)$.

Example 2.9. If $\rho_\beta = e^{-\beta h}/\operatorname{tr}(e^{-\beta h})$ is the Gibbs state of the Hamiltonian h at inverse temperature β, then $\sigma_t^{\varphi_\beta}(x) = e^{-i\beta t h} x e^{i\beta t h}$. Thus, up to a rescaling of the time parameter by $-\beta$, the modular group coincides with the time evolution of the clsoed system with Hamiltonian h under the von Neumann equation (in the Heisenberg picture).

Proposition 2.10. *Let $\varphi_A = \operatorname{tr}(\,\cdot\,\rho_A)$, $\varphi_B = \operatorname{tr}(\,\cdot\,\rho_B)$ be normal states on $\mathbb{B}(H_A)$, $\mathbb{B}(H_B)$, respectively, with ρ_A, ρ_B non-singular. If $\Phi \colon S^1(H_A) \to S^1(H_B)$ is a quantum channel such that $\Phi(\rho_A) = \rho_B$, then there exists a quantum channel $\Psi \colon S^1(H_B) \to S^1(H_A)$ such that $\Psi(\rho_B) = \rho_A$ and*

$$\varphi_A(x^* \Phi^\dagger(y)) = \varphi_B(\Psi^\dagger(x)^* y)$$

for all $x \in \mathbb{B}(H_A)$, $y \in \mathbb{B}(H_B)$ if and only if $\sigma_t^{\varphi_A} \circ \Phi = \Phi \circ \sigma_t^{\varphi_B}$ for all $t \in \mathbb{R}$.

Chapter 3

Static Quantum Wasserstein Distances

The bounded operators on H also act as bounded operators on $S^2(H)$ by left and right multiplication, i.e.,

$$L_x\colon S^2(H)\to S^2(H),\ y\mapsto xy$$
$$R_x\colon S^2(H)\to S^2(H),\ y\mapsto yx$$

for $x\in\mathbb{B}(H)$. Note that the operators L_x and R_y commute for $x,y\in\mathbb{B}(H)$ and the weak* closed linear hull of $\{L_xR_y\mid x,y\in\mathbb{B}(H)\}$ is all of $\mathbb{B}(S^2(H))$.

If φ_A,φ_B are normal states on $\mathbb{B}(H)$, then a normal state ω on $\mathbb{B}(S^2(H))$ is called a *transport plan* from φ_A to φ_B if $\omega(L_x)=\varphi_A(x)$ and $\omega(R_y)=\varphi_B(y)$ for all $x,y\in\mathbb{B}(H)$. If φ_A and φ_B are represented by non-singular density operators, then a transport plan ω is called *modular* if

$$\omega(L_{\sigma_t^{\varphi_A}(x)}R_{\sigma_t^{\varphi_B}(y)})=\omega(L_xR_y)$$

for all $x,y\in\mathbb{B}(H)$.

Remark 3.1. The space of Hilbert–Schmidt operators $S^2(H)$ is isometrically isomorphic to the tensor product $H\otimes\overline{H}$ and $\mathbb{B}(S^2(H))$ is $*$-isomorphic to the (von Neumann algebra) tensor product $\mathbb{B}(H)\mathbin{\overline{\otimes}}\mathbb{B}(\overline{H})$. Thus transport plans can also be viewed as states on $\mathbb{B}(H)\otimes\mathbb{B}(\overline{H})$, which is closer to the classical formulation of transport plans. Note however that it is important that the second tensor factor is $\mathbb{B}(\overline{H})$ instead of $\mathbb{B}(H)$ to get a one-to-one correspondence between transport plans and quantum channels. The space $\mathbb{B}(\overline{H})$ is canonically isomorphic to $\mathbb{B}(H)^{\mathrm{op}}$, the opposite algebra of $\mathbb{B}(H)$, but not canonically isomorphic to $\mathbb{B}(H)$.

Proposition 3.2. *Let φ_A, φ_B be normal states on $\mathbb{B}(H)$ with density operators ρ_A and ρ_B, respectively. There is a one-to-one correspondence between* *the set of transport plans from φ_A to φ_B and the set of quantum channels $\Phi\colon S^1(H) \to S^1(H)$ such that $\Phi(\rho_A) = \rho_B$. It assigns to a transport plan ω the unique quantum channel $\Phi\colon S^1(H) \to S^1(H)$ such that*

$$\omega(L_x R_y) = \langle \rho_A^{1/2}, x\rho_A^{1/2}\Phi^\dagger(y)\rangle$$

for all $x, y \in \mathbb{B}(H)$.

Moreover, if ρ_A, ρ_B are non-singular, then ω is modular if and only if $\Phi^\dagger \circ \sigma_t^{\varphi_B} = \sigma_t^{\varphi_A} \circ \Phi^\dagger$ for all $t \in \mathbb{R}$. In this case, we call Φ a modular *quantum channel.*

Remark 3.3. If Φ is a quantum channel such that $\Phi(\rho_A) = \rho_B$ and ω the corresponding transport plan, then the Hilbert bimodule $\mathcal{H}$ from the previous lecture can be recovered as the GNS Hilbert space of ω with the left and right action given by $\pi_\omega \circ L$ and $\pi_\omega \circ R$, and ξ_0 is the corresponding cyclic vector Ω_ω.

Moreover, the adjoint $\Phi^\dagger$ intertwines the modular groups σ^{φ_A} and σ^{φ_B} if and only if there exists a strongly continuous unitary group $(\mathcal{U}_t)_{t\in\mathbb{R}}$ on $\mathcal{H}$ such that $\mathcal{U}_t(x\xi y) = \sigma_t^{\varphi_A}(x)(\mathcal{U}_t\xi)\sigma_t^{\varphi_B}(y)$ for all $x, y \in \mathbb{B}(H)$, $\xi \in \mathcal{H}$, $t \in \mathbb{R}$ and $\mathcal{U}_t\xi_0 = \xi_0$ for all $t \in \mathbb{R}$.

Definition 3.4 (GNS and KMS transport cost)**.** Let $\rho \in S^1(H)$ be a density operator. The GNS and KMS inner product associated with ρ are defined by

$$\langle x, y\rangle_{\rho,\mathrm{GNS}} = \mathrm{tr}(x^* y\rho)$$
$$\langle x, y\rangle_{\rho,\mathrm{KMS}} = \mathrm{tr}(x^*\rho^{1/2} y\rho^{1/2})$$

for $x, y \in \mathbb{B}(H)$.

If $\underline{x} = (x_1, \dots, x_d) \in \mathbb{B}(H)^d$ is a d-tuple of self-adjoint operators and $\Phi\colon S^1(H) \to S^1(H)$ is a quantum channel, then its GNS and KMS cost with respect to $\underline{x}$ are defined as

$$C_{\rho,\underline{x},\mathrm{GNS}}(\Phi) = \sum_{j=1}^d \left(\|x_j\|^2_{\rho,\mathrm{GNS}} + \|x_j\|^2_{\Phi(\rho),\mathrm{GNS}} - 2\langle x_j, \Phi^\dagger(x_j)\rangle_{\rho,\mathrm{GNS}}\right)$$
$$C_{\rho,\underline{x},\mathrm{KMS}}(\Phi) = \sum_{j=1}^d \left(\|x_j\|^2_{\rho,\mathrm{KMS}} + \|x_j\|^2_{\Phi(\rho),\mathrm{KMS}} - 2\langle x_j, \Phi^\dagger(x_j)\rangle_{\rho,\mathrm{KMS}}\right)$$

Remark 3.5. The distinction between GNS and KMS transport cost is a purely noncommutative (or quantum) phenomenon. Both costs have their advantages: For the KMS cost, if Φ transports ρ_A to ρ_B, then there exists a channel Ψ, known as KMS adjoint or Petz recovery map in quantum information, that transports ρ_B to ρ_A and satisfies $C_{\rho_A,\underline{x},\mathrm{KMS}}(\Phi) = C_{\rho_B,\underline{x},\mathrm{KMS}}(\Psi)$. This ensures that the quantum Wasserstein distance defined in terms of the KMS transport cost is symmetric, while one has to restrict to channels that intertwine the modular groups in the definition of the Wasserstein distance in terms of the GNS transport cost to guarantee symmetry.

On the other hand, the GNS transport cost has a natural representation as a sum of squares in a Hilbert bimodule, which is key to proving the triangle inequality. Such a sum of squares representation for the KMS transport cost had not been known until recently for the KMS transport cost until recently and consequently, also the triangle inequality for the Wasserstein distance defined in terms of the KMS transport cost had been open.

Definition 3.6 (Duvenhage and de Palma–Trevisan Wasserstein distance)**.** Let φ_A, φ_B be normal states on $\mathbb{B}(H)$ with density matrices ρ_A, ρ_B and let $\underline{x} \in \mathbb{B}(H)^d$ be a d-tuple of self-adjoint operators. The *de Palma–Trevisan Wasserstein distance* between φ_A and φ_B is defined as

$$W_{\mathrm{dP,T}}(\varphi_A, \varphi_B) = \inf\{C_{\rho_A,\underline{x},\mathrm{KMS}}(\Phi)^{1/2} : \Phi \text{ quantum channel, } \Phi(\rho_A) = \rho_B\}.$$

Moreover, if ρ_A, ρ_B are non-singular, then the *Duvenhage Wasserstein distance* between φ_A and φ_B is defined as

$$W_{\mathrm{Duv}}(\varphi_A, \varphi_B) \\ = \inf\{C_{\rho_A,\underline{x},\mathrm{GNS}}(\Phi)^{1/2} : \Phi \text{ modular quantum channel, } \Phi(\rho_A) = \rho_B\}.$$

Remark 3.7. In the definition of the transport cost, the terms $\|x_j\|^2_{\rho_A}$ and $\|x_j\|^2_{\Phi(\rho_A)}$ (for both the GNS and KMS inner product) are the same for any quantum channel transporting ρ_A to ρ_B. Thus the quantum Wasserstein distances can be reformulated in terms of maximization problems for the quantity $\sum_{j=1}^d \langle x_j, \Phi^\dagger(x_j)\rangle_{\rho_A}$ for either the GNS or KMS inner product.

In the case of the KMS inner product, this term can be rewritten in terms of the associated transport plan as

$$\langle x_j, \Phi^\dagger(x_j)\rangle_{\rho_A,\mathrm{KMS}} = \mathrm{tr}(\rho_A^{1/2} x_j \rho_A^{1/2} \Phi^\dagger(x_j)) = \omega(L_{x_j} R_{x_j}).$$

Example 3.8 (Quantum optimal transport between Gaussian states). On $L^2(\mathbb{R})$ consider the self-adjoint operators R_1, R_2 given by $R_1 f = -if'$, $R_2 f(x) = x f(x)$ for f in their respective natural domains. Note that since these operators are unbounded, they do not fall into the framework presented before and the following computations are to be taken as purely formal, although they can be made rigorous. We want to compute the de

Palma–Trevisan distance between Gaussian states φ_A and φ_B for $x_j = R_j$, $j \in \{1,2\}$.

Note that $S^2(L^2(\mathbb{R})) \cong L^2(\mathbb{R}^2)$ canonically in such a way that L_{R_1}, L_{R_2} ate mapped to R_1, R_2 and R_{R_1}, R_{R_2} are mapped to R_3, R_4. Since the maximization problem for transport plans ω between φ_A and φ_B only depends on the moments $\omega(R_j R_k)$ with $j \in \{1,2\}$, $k \in \{3,4\}$, one can restrict to Gaussian transport plans (much like in the case of Gaussian measures). In particular, if φ_A and φ_B have mean zero and covariance matrices $\nu_A I$ and $\nu_B I$ with $\frac{1}{2} \leq \nu_A \leq \nu_B$, then

$$\begin{aligned} &W_{\mathrm{dPT}}\left(\varphi_A, \varphi_B\right)^2 \\ &= \sqrt{\left(\nu_B + \frac{1}{2}\right)\left(\nu_B - \frac{1}{2}\right)} + \sqrt{\left(nu_A + \frac{1}{2}\right)\left(\nu_A - \frac{1}{2}\right)} \\ &\quad - 2\sqrt{\left(\nu_B + \frac{1}{2}\right)\left(\nu_A - \frac{1}{2}\right)}. \end{aligned}$$

Theorem 3.9 (Duvenhage). *If $\underline{x} \subset \mathbb{B}(H)^d$ is a d-tuple of self-adjoint operators such that the algebra generated by $\underline{x}$ is weak* dense in $\mathbb{B}(H)$, then W_{Duv} is a metric on the set of all faithful normal states on $\mathbb{B}(H)$.*

Theorem 3.10 (Wirth). *If $\rho_A, \rho_B \in S^1(H)$ are density matrices, $\Phi\colon S^1(H) \to S^1(H)$ is a quantum channel such that $\Phi(\rho_A) = \rho_B$ and $(v_i)_{i\in I}$ is a family of Kraus operators for Φ, then*

$$\begin{aligned} &C_{\rho_A,\underline{x},\mathrm{KMS}}(\Phi) \\ &\quad = \sum_{j=1}^{d} \sum_{i\in I} \int_{\mathbb{R}} \| v_i \rho_B^{1/4+it} x_j \rho_B^{1/4-it} - \rho_A^{1/4+it} x_j \rho_A^{-1/4-it} v_j \rho_B^{1/2} \|_2^2 \frac{2\,dt}{\cosh 2\pi t}. \end{aligned}$$

Remark 3.11. The proof of this formula for the KMS transport cost relies on complex analysis methods and in particular an application of the residue theorem. This formula has surprising connections to the structure of master equations of open quantum systems coupled to an environment in equilibrium.

Corollary 3.12. *If $\rho_A, \rho_B, \rho_C \in S^1(H)$ are density matrices and $\Phi, \Psi\colon S^1(H) \to S^1(H)$ are quantum channel such that $\Phi(\rho_A) = \rho_B$ and $\Psi(\rho_B) = \rho_C$ then*

$$C_{\rho_A,\underline{x},\mathrm{KMS}}(\Psi \circ \Phi)^{1/2} \leq C_{\rho_A,\underline{x},\mathrm{KMS}}(\Phi)^{1/2} + C_{\rho_B,\underline{x},\mathrm{KMS}}(\Psi)^{1/2}.$$

Corollary 3.13. *The Wasserstein distance $W_{\mathrm{dP,T}}$ satisfies the triangle inequality.*

Chapter 4

Dynamic Quantum Wasserstein Distance

Recall that the Benamou–Brenier formula recast the Wasserstein distance as minimal value of an optimization problem over continuous-time interpolations between two given probability measures, that is,

$$W_2(\mu,\nu)^2 = \inf_{(\mu_t,v_t)} \int_0^1 \int_\Omega |v_t|^2 \, d\mu_t \, dt,$$

where μ, ν are two Borel probability measures on Ω and the infimum is taken over all curves (μ_t) of Borel probability measures with $\mu_0 = \mu$, $\mu_1 = \nu$ and vector fields $v\colon [0,1] \times \Omega \to \mathbb{R}^d$ that satisfy the continuity equation

$$\dot{\mu}_t + \nabla \cdot (\mu_t v_t) = 0$$

in a suitable weak sense.

At least formally, the Benamou–Brenier formula expresses W_2 as the distance induced by a Riemannian metric g on the space $\mathcal{P}_2(\Omega)$ of Borel probability measures with finite second moments. With respect to this Riemannian metric, the gradient of the differential entropy

$$H(\mu) = \begin{cases} \int \rho \log \rho \, dx & \text{if } d\mu(x) = \rho \, dx \\ \infty & \text{otherwise} \end{cases}$$

is given by $\nabla_g H(\mu) = -\Delta\mu$. Thus the gradient flow of H with respect to the Riemannian metric g, given by the solution of

$$\dot{\mu}_t = -\nabla_g H(\mu_t),$$

coincides with the heat flow. This is Otto's interpretation of the Jordan–Kinderlehrer–Otto theorem, which was the starting point for the work of Carlen and Maas towards a dynamic quantum Wasserstein distance.

To quantize the Benamou–Brenier formula, one first needs a notion of gradient/divergence to quantize the continuity equation. Generally speaking, the order and algebraic operations of $\mathbb{B}(H)$ provide the framework for noncommutative measure theory, but one needs additionally input data to build a noncommutative geometry.

We will present a simple framework here, which does not cover all mathematically and physically interesting examples, but it showcases many of the key structural features while avoiding some technical difficulties. In particular, we work with quantum systems with finitely many degrees of freedom, which means mathematically that the observables are elements of $\mathbb{B}(H)$ for a finite-dimensional Hilbert space H.

In this setting, the quantum analog of partial derivatives are commutators with skew-adjoint elements, as justified by the properties from the following lemma.

Lemma 4.1. *If $v \in \mathbb{B}(H)$ is skew-adjoint ($v^* = -v$), then*

$$[v, xy] = [v, x]y + x[v, y]$$

and

$$\operatorname{tr}([v, x]^* y) = -\operatorname{tr}(x^*[v, y])$$

for all $x, y \in \mathbb{B}(H)$.

The first property is a quantum version of the Leibniz rule, while the second property is a quantum version of the integration by parts formula, which asserts that the partial derivatives are skew-adjoint operators on $L^2(\mathbb{R}^d)$ on appropriate domains.

In the following, we fix skew-adjoint operators $v_1, \dots, v_d \in \mathbb{B}(H)$ such that $\operatorname{tr}(v_j) = 0$ and $\operatorname{tr}(v_i v_j) = 0$ for $i \neq j$ and define $\partial_j = [v_j, \cdot]$. The property $\operatorname{tr}(v_j)$ can always be achieved because v_j and $v_j - \operatorname{tr}(v_j)$ define the same commutator. The property $\operatorname{tr}(v_i v_j) = 0$ for $i \neq j$ is an orthogonality relation mimicking the fact that the directions of the partial derivatives are orthogonal. It can actually also always be achieved without changing the Laplacian defined in the next paragraph.

The noncommutative gradient in our setting is

$$\partial \colon \mathbb{B}(H) \to \mathbb{B}(H)^d,\ x \mapsto (\partial_j(x))_{j=1}^d,$$

where we view $\mathbb{B}(H)^d$ as space of noncommutative vector fields, and the noncommutative divergence is

$$-\partial^\dagger \colon \mathbb{B}(H)^d \to \mathbb{B}(H),\ (x_j)_{j=1}^d \mapsto \sum_{j=1}^d \partial_j(x_j).$$

Here $\partial^\dagger$ is the adjoint of ∂ if $\mathbb{B}(H)$ is endowed with the Hilbert–Schmidt inner product $\langle x, y\rangle_2 = \operatorname{tr}(x^* y)$, which is a noncommutative version of the L^2 inner product. The noncommutative Laplacian is

$$\Delta = -\partial^\dagger \partial \colon \mathbb{B}(H) \to \mathbb{B}(H),\ x \mapsto \sum_{j=1}^d \partial_j^2(x).$$

Beyond the formal analogy to the classical Laplacian, the following result gives more evidence that this may indeed be a suitable noncommutative version of the Laplace operator.

Proposition 4.2. *For every $t \geq 0$, the map $e^{t\Delta}$ is a quantum channel and symmetric with respect to the Hilbert–Schmidt inner product.*

Conversely, if $(\Phi_t)_{t\geq 0}$ is a continuous family of quantum channels on $\mathbb{B}(H)$ such that $\Phi_t^\dagger = \Phi_t$, $\Phi_0 = \mathrm{id}_{\mathbb{B}(H)}$ and $\Phi_{s+t} = \Phi_s \Phi_t$, then there exist skew-adjoint $v_1, \dots, v_d \in \mathbb{B}(H)$ such that $\operatorname{tr}(v_j) = 0$, $\operatorname{tr}(v_i v_j) = 0$ for $i \neq j$ and $\Phi_t = e^{t\Delta}$.

One aspect of differential calculus that does not generalize quite so straightforwardly is the chain rule. The reason is that in noncommutative the Leibniz rule, the order of the factors plays a crucial role. For that reason, one cannot expect a chain rule of the form $\partial_j(f(x)) = f'(x)\partial_j(x)$ either. Instead of developing the general noncommutative chain rule here, we will focus on a particular identity that is crucial for Otto calculus. In the commutative case, it is $\rho\partial_j(\log \rho) = \partial_j(\rho)$, which is an immediate consequence of the chain rule. In the noncommutative case, it takes on the following somewhat more complicated form.

Lemma 4.3. *If $\rho \in \mathbb{B}(H)$ is positive definite, then*

$$\int_0^1 \rho^s \partial_j(\log \rho)\rho^{1-s}\, ds = \partial_j(\rho).$$

Here $\log \rho$ is defined in terms of functional calculus, that is, if $\rho = u \operatorname{diag}(\lambda_1, \dots, \lambda_n) u^*$ with $\lambda_j > 0$ and $u \in \mathbb{B}(H)$ unitary, then $\log \rho = u \operatorname{diag}(\log \lambda_1, \dots, \log \lambda_n) u^*$.

Proof. If $F(s) = -\rho^s v_j \rho^{1-s}$, then $F(1) = -\rho v_j$, $F(0) = -v_j \rho$ and $F'(s) = -\rho^s(\log \rho v_j - v_j \log \rho)\rho^{1-s} = \rho^s \partial_j(\log \rho)\rho^{1-s}$. Now the statement follows from the fundamental theorem of calculus. □

As a consequence we define

$$\hat{\rho}x = \int_0^1 \rho^s x \rho^{1-s}\, ds,$$

and we say that $(\rho_t, \xi_t)_{t\in[0,1]}$ with (ρ_t) a smooth curve of non-singular density matrices and (ξ_t) a smooth curve in $\mathbb{B}(H)^d$ satisfies the *noncommutative continuity equation* if

$$\dot{\rho}_t - \partial^\dagger(\hat{\rho}_t \xi_t) = 0.$$

Definition 4.4 (Carlen–Maas Wasserstein distance)**.** If $\rho_A, \rho_B \in S^1(H)$ are non-singular density matrices, then the *Carlen–Maas distance* between them is defined as

$$W_{\mathrm{CM}}(\rho_A, \rho_B) = \inf_{(\rho_t,\xi_t)} \left(\int_0^1 \langle \xi_t, \hat{\rho}_t \xi_t \rangle_2 \, dt \right)^{1/2},$$

where the infimum is taken over all pairs (ρ_t, ξ_t) that satisfy the noncommutative continuity equation and $\rho_0 = \rho_A$, $\rho_1 = \rho_B$.

Since we are in finite dimensions, the Otto formalism viewing the Wasserstein distance as the distance induced by a Riemannian metric on the space of probability measures can be made rigorous. To see this, first note that the non-singular density matrices form an open subset of the hyperplane $\{x \in \mathbb{B}(H) \mid x = x^*, \mathrm{tr}(x) = 1\}$ and thus a smooth submanifold $\mathfrak{S}_+(H)$ of the self-adjoint operators on H. Moroever, the tangent space at an arbitrary $\rho \in \mathfrak{S}_+(H)$ is canonically identified with $\{x \in \mathbb{B}(H) \mid x = x^*, \mathrm{tr}(x) = 0\}$.

Proposition 4.5. *Assume that the algebra generated by $\{v_1, \dots, v_d\}$ is $\mathbb{B}(H)$. If $x \in \mathbb{B}(H)$ is self-adjoint with $\mathrm{tr}(x) = 0$ and $\rho \in \mathfrak{S}_+(H)$, then there exists a unique element $K_\rho(x) \in \mathbb{B}(H)^d$ such that*

- *$K_\rho(x)$ is the noncommutative gradient of a self-adjoint element $y \in \mathbb{B}(H)$,*
- *$x = \partial^\dagger(\hat{\rho} K_\rho(x))$.*

Moreover,

$$g_\rho \colon T_\rho \mathfrak{S}_+(H) \times T_\rho \mathfrak{S}_+(H) \to \mathbb{R},\ (x, y) \mapsto \langle K_\rho(x), \hat{\rho} K_\rho(y) \rangle_2$$

defines a Riemannian metric on $\mathfrak{S}_+(H)$ and the induced distance function is W_{CM}.

So far, the choice of the multiplication operator $\hat{\rho}$ played no crucial role, and in fact, one could also define a metric for $\hat{\rho}x = \rho x$ or $\hat{\rho}x = \frac{1}{2}(\rho x + x\rho)$, for example. The choice is crucial, however, to recover the noncommutative heat flow as gradient flow of the quantum relative entropy.

Theorem 4.6 (Carlen, Maas)**.** *Let E be the orthogonal projection onto* $\ker \Delta$ *(with respect to the Hilbert–Schmidt inner product). If*

$$H\colon \mathfrak{S}_+(H) \to [0,\infty),\ \rho \mapsto \operatorname{tr}(\rho(\log \rho - \log E(\rho))),$$

then $\nabla_g H(\rho) = -\Delta\rho$.

The quantity $H(\rho)$ is the relative entropy of ρ with respect to the fixed-point algebra of $(e^{t\Delta})_{t\geq 0}$, which quantifies how many measurements one needs to distinguish ρ from a fixed point of the evolution. If the algebra generated by $\{v_1, \dots, v_d\}$ is $\mathbb{B}(H)$, then $\ker \Delta = \mathbb{C}1$ and $E(\rho) = I/\dim H$.

One of the striking applications of Otto calculus, going back to the work of Otto and Villani, is that convexity properties of the entropy in Wasserstein space imply functional inequalities for the heat flow. Similar results can also be deduced in the quantum setting. Here we focus particularly on the decay of relative entropy along the noncommutative heat flow. To state the result, recall that a functional S on a Riemannian manifold (M, g) is called (strongly) geodesically K-convex if $\frac{d^2}{dt^2} S(\gamma_t) \geq K g(\dot{\gamma}_t, \dot{\gamma}_t)$ for every constant-speed geodesic (γ_t) in M.

Theorem 4.7 (Carlen–Maas)**.** *If $K > 0$ and H is geodesically K-convex on $(\mathfrak{S}_+(H), g)$, then*

$$H(e^{t\Delta}\rho) \leq e^{-2Kt} H(\rho)$$

for all $\rho \in \mathfrak{S}_+(H)$.

With this approach, Carlen and Maas could prove optimal decay rates for the relative entropy for certain evolutions on Bosonic and Fermionic Fock space.