Quantum Relative Entropy and Quantum Optimal Transport

Eric A. Carlen

Rutgers University

Oberwolfach, May 9, 2018

with A. Vershynina, *Recovery map stability for the Data Processing Inequality*, arXiv:1710.08080

with J. Maas, *An analog of the 2-Wasserstein metric in non-commutative probability under which the fermionic Fokker-Planck equation is gradient flow for the entropy*, Comm. Math. Phys. **331**, 887–926, 2014.

with J. Maas, *Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance*, Jour. Func. Analysis, **273**, no. 5, (2017) 1810-1869.

Related work Chen, Georgiou, Tannenbaum and also Mielke, Mittnenzweig. (Also on the arXiv).

Quantum Entropy

A *density matrix* on a Hilbert space $\mathcal{H}$ is a positive trace class operator ρ with $\mathrm{Tr}[\rho] = 1$.

On a bipartite space $\mathcal{H}_1 \otimes \mathcal{H}_2$, $\{u_j\}$ is an orthonormal basis of $\mathcal{H}_2$, define $V_j : \mathcal{H}_1 \to \mathcal{H}_1 \otimes \mathcal{H}_2$ by

$$V_j v = v \otimes u_j \ .$$

Then for an operator A on $\mathcal{H}_1 \otimes \mathcal{H}_2$, $\mathrm{Tr}_2[A] = \sum_j V_j^* A V_j$.

Notice that

$$\sum_j V_j^* V_j = \mathbf{1}_{\mathcal{H}_1} \quad \text{and} \quad \sum_j V_j V_j^* = \mathbf{1}_{\mathcal{H}_1 \otimes \mathcal{H}_2}$$

The von Neumann entropy of a density matrix ρ is

$$S(\rho) = -\mathrm{Tr}[\rho \log \rho] \ .$$

Given two density matrices ρ and σ on the same Hilbert space $\mathcal{H}$, the Umegaki relative entropy $S(\rho||\sigma)$ is defined by

$$S(\rho||\sigma) = \mathrm{Tr}[\rho(\log \rho - \log \sigma)] \ .$$

In the early 1970's G. Lindblad proved the DPI for the Umegaki relative entropy, Let $\mathcal{M}$ be a von Neumann subalgebra of $\mathcal{B}(\mathcal{H})$, and let Let $\mathcal{N}$ be a subalgebra of $\mathcal{M}$. In the simplest case, $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$, both $\mathcal{H}_1$ and $\mathcal{H}_2$ finite dimensional $\mathcal{N} = \mathcal{B}(\mathcal{H}_1) \otimes \mathbf{1}_{\mathcal{H}_2}$; i.e., all operators in $\mathcal{M}$ of the form $A \otimes \mathbf{1}_{\mathcal{H}_2}$, $A \in \mathcal{B}(\mathcal{H}_1)$.

For a normal state ρ on $\mathcal{M}$, let $\rho_{\mathcal{N}}$ denote the element of $\mathcal{N}$ obtained by restricting the state induced by ρ to $\mathcal{N}$. In the example, $\rho_{\mathcal{N}} = \mathrm{Tr}_2 \rho$.
Lindblad proved

$$S(\rho||\sigma) \geq S(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}}) \ ,$$

and more generally, $S(\rho||\sigma) \geq S(\Phi(\rho)||\Phi(\sigma))$ for any CPT map Φ.

Equality in the DPI

The conditon for equality in the DPI, discovered by Petz, involves the *Accardi–Cechini coarse graining operator*. In the commuative setting, there is always a conditional expectation operator from $\mathcal{E}$ from $\mathcal{M}$ to $\mathcal{N}$ such that $\rho_{\mathcal{N}} \circ \mathcal{E} = \rho$. In the non-commutative case, these rarely exist.

$$\langle A, X \rangle_{GNS,\rho} = \rho(A^*X) = \rho(\mathscr{E}(A^*X)) = \rho(A^*\mathscr{E}(X)) = \langle A, \mathscr{E}(X) \rangle_{GNS,\rho} \ .$$

Thus $\mathscr{E}$ woud have to be the orthogonal projection form $\mathcal{M}$ onto $\mathcal{N}$ with resepct to the GNS inner product.

Theorem 0.1. *Let $\mathcal{M}$ be a finite dimensional von Neumann algebra, and let $\mathcal{N}$ be a von Neumann subalgebra of $\mathcal{M}$. Let ρ be a faithful state on $\mathcal{M}$, and let Δ_ρ be the* modular operator *on $\mathcal{M}$ defined by $\Delta_\rho(X) = \rho X \rho^{-1}$. Let $\mathscr{P}_\rho$ be the orthogonal projection from $\mathcal{M}$ onto $\mathcal{N}$ in the GNS inner product induced by ρ. Then: $\mathscr{P}_\rho$ is Hermitian; i.e., it preserves self-adjointness, if and only if $\mathcal{N}$ is invariant under Δ_ρ.*

There is another inner product on $\mathcal{M}$ that is naturally induced by a faithful state ρ, namely the *Kubo-Martin-Schwinger (KMS) inner product*.

$$\langle X, Y \rangle_{KMS,\rho} = \mathrm{Tr}[\rho^{1/2} X^* \rho^{1/2} Y] = \mathrm{Tr}[(\rho^{1/4} X \rho^{1/4})^* (\rho^{1/4} Y \rho^{1/4})] \,.$$

For any $X \in \mathcal{M}$ and $Y \in \mathcal{N}$, $Y \mapsto \langle X, Y \rangle_{KMS,\rho}$ is a bounded linear functional on $(\mathcal{N}, \langle, \cdot, \cdot \rangle_{KMS,\rho_{\mathcal{N}}})$.

$$\langle X, Y \rangle_{KMS,\rho} = \langle \mathscr{A}_\rho(X), Y \rangle_{KMS,\rho_{\mathcal{N}}} \ ,$$

This is the *Accardi-Cecchini coarse graining operator* $\mathscr{A}_\rho$.

By a simple calculuation,

$$\mathscr{A}_\rho(X) = \rho_{\mathcal{N}}^{-1/2}\mathscr{E}_\tau(\rho^{1/2}X\rho^{1/2})\rho_{\mathcal{N}}^{-1/2} \ ,$$

where $\mathscr{E}_\tau$ is the tracial conditional expectation.

The Petz recovery map $\mathscr{R}_\rho$ is the predual to $\mathscr{A}_\rho$.

Evidently, in this simple setting,

$$\mathscr{R}_\rho(\gamma) = \rho^{1/2}(\rho_{\mathcal{N}}^{-1/2}\gamma\rho_{\mathcal{N}}^{-1/2})\rho^{1/2} \ .$$

It is clear that $\mathscr{R}_\rho(\rho_{\mathcal{N}}) = \rho$, and that $\mathscr{R}_\rho$ is a CPTP map. Hence **if** it is also the case that $\mathscr{R}_\rho(\sigma_{\mathcal{N}}) = \sigma$, then

$$S(\rho||\sigma) \leq S(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}}) \ .$$

Let ρ and σ be two density matrices on a finite dimensional Hilbert space $\mathcal{H}$. Make $\mathcal{B}(\mathcal{H})$ into a Hilbert space by equipping it with the Hilbert-Schmidt inner product $\langle A, B\rangle_{\mathrm{HS}} = \mathrm{Tr}[A^*B]$. Then the *relative modular operator* $\Delta_{\sigma.\rho}$ is the operator on $\mathcal{B}(\mathcal{H})$ given by

$$\Delta_{\sigma.\rho}(X) = \sigma X \rho^{-1} \ .$$

Theorem 0.2 (C., Vershynina)**.**

$$S(\rho||\sigma) - S(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}}) \geq \left(\frac{1}{4\pi}\right)^4 \|\Delta_{\sigma,\rho}\|^{-2} \|\sigma_{\mathcal{N}}^{1/2}\rho_{\mathcal{N}}^{-1/2}\rho^{1/2} - \sigma^{1/2}\|_2^4 \ .$$

Then since

$$\|(\sigma_{\mathcal{N}})^{1/2}(\rho_{\mathcal{N}})^{-1/2}\rho^{1/2} - \sigma^{1/2}\|_2 \geq \frac{1}{2}\|\mathcal{R}_\rho(\sigma_{\mathcal{N}}) - \sigma\|_1 \ ,$$

$$S(\rho||\sigma) - S(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}}) \geq \left(\frac{1}{8\pi}\right)^4 \|\Delta_{\sigma,\rho}\|^{-2}\|\mathcal{R}_\rho(\sigma_{\mathcal{N}}) - \sigma\|_1^4 \ .$$

Let $f : (0, +\infty) \to \mathbb{R}$ be an operator convex function, so that for all positive $n \times n$ matrices A and B,

$$f(\tfrac{1}{2}A + \tfrac{1}{2}B) \leq \tfrac{1}{2}f(A)) + \tfrac{1}{2}f(B) \ .$$

Petz has defined the f*-relative quasi-entropy* as

$$S_f(\rho||\sigma) = \mathrm{Tr}(f(\Delta_{\sigma,\rho}\rho)) = \langle \rho^{1/2}, f(\Delta_{\sigma,\rho})\rho^{1/2}\rangle_{HS} \ . \qquad (0.1)$$

Since $-\log(\Delta_{\sigma,\rho})\rho^{1/2} = \rho^{1/2}\log\rho - \log\sigma\rho^{1/2}$, the choice $f(x) = -\log x$ yields the Umegaki relative entropy.

Since for each $t > 0$, the function $x \mapsto (t+x)^{-1}$ is operator convex, this construction yields a one parameter family a quasi relative entropies, $S_{(t)}$, defined by

$$S_{(t)}(\rho||\sigma) = \mathrm{Tr}\left[(t + \Delta_{\sigma,\rho})^{-1}\rho\right] \ .$$

From the integral representation of the logarithm

$$-\log(A) = \int_0^\infty \left(\frac{1}{t+A} - \frac{1}{1+t}\right) \mathrm{d}t \ ,$$

it follows that

$$S(\rho||\sigma) = \int_0^\infty \left(S_{(t)}(\rho||\sigma) - \frac{1}{1+t}\right) \mathrm{d}t \ ,$$

Define a partial isometry U by $UX = X\rho_{\mathcal{N}}^{-1/2}\rho^{1/2}$. Then $U\rho_{\mathcal{N}}^{1/2} = \rho^{1/2}$.

Lemma 0.3. *Let U be a partial isometry embedding a Hilbert space $\mathcal{K}$ into a Hilbert space $\mathcal{H}$. Let B be an invertible positive operator on $\mathcal{K}$, let A be an invertible positive operator on $\mathcal{H}$, and suppose that*

$$U^*AU = B \ .$$

Then for all $v \in \mathcal{K}$,

$$\langle v, U^*A^{-1}Uv\rangle = \langle v, B^{-1}v\rangle + \langle w, Aw\rangle \ ,$$

where

$$w := UB^{-1}v - A^{-1}Uv \ .$$

This leads to

$$\begin{aligned} S_{(t)}(\rho||\sigma) &- S_{(t)}(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}}) \\ &= \langle \rho^{1/2}, (t+\Delta_{\sigma,\rho})^{-1}\rho^{1/2}\rangle - \langle \rho_{\mathcal{N}}^{1/2}, (t+\Delta_{\sigma_{\mathcal{N}},\rho_{\mathcal{N}}})^{-1}\rho_{\mathcal{N}}^{1/2}\rangle \\ &= \langle w_t, (t+\Delta_{\sigma,\rho})w_t\rangle \geq t\|w_t\|^2, \end{aligned}$$

where

$$w_t := U(t+\Delta_{\sigma_{\mathcal{N}},\rho_{\mathcal{N}}})^{-1}(\rho_{\mathcal{N}})^{1/2} - (t+\Delta_{\sigma,\rho})^{-1}\rho^{1/2}\ .$$

Recall

$$S(\rho||\sigma) = \int_0^\infty \left(S_{(t)}(\rho||\sigma) - \frac{1}{1+t} \right) \mathrm{d}t\ ,$$

$$X^{1/2} = \pi \int_0^\infty t^{1/2} \left(\frac{1}{t} - \frac{1}{t+X} \right) \mathrm{d}t,$$

and using this with $U(\rho_{\mathcal{N}})^{1/2} = \rho^{1/2}$ once more,

$$U(\Delta_{\sigma_{\mathcal{N}},\rho_{\mathcal{N}}})^{1/2}(\rho_{\mathcal{N}})^{1/2} - (\Delta_{\sigma,\rho})^{1/2}\rho^{1/2} = \pi \int_0^\infty t^{1/2} w_t \mathrm{d}t \ .$$

$$\begin{aligned} U(\Delta_{\sigma_{\mathcal{N}},\rho_{\mathcal{N}}})^{1/2}(\rho_{\mathcal{N}})^{1/2} \quad &- \quad (\Delta_{\sigma,\rho})^{1/2}\rho^{1/2} \\ &= \quad U(\sigma_{\mathcal{N}})^{1/2} - \sigma^{1/2} \\ &= \quad (\sigma_{\mathcal{N}})^{1/2}(\rho_{\mathcal{N}})^{-1/2}\rho^{1/2} - \sigma^{1/2} \ . \end{aligned}$$

This brings us back around to our quatitiative for of the DPI:

$$S(\rho||\sigma) - S(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}}) \geq$$

$$\left(\frac{1}{4\pi}\right)^4 \|\Delta_{\sigma,\rho}\|^{-2} \|\sigma_{\mathcal{N}}^{1/2} \rho_{\mathcal{N}}^{-1/2} \rho^{1/2} - \sigma^{1/2}\|_2^4 \ .$$

A standard argument extends this to the general CPTP case.

The classical starting point

Let $\sigma(x)$ be a smooth strictly positive probability density on R^n. If $\rho(x)$ as any other such density, the relative entropy of ρ with respect to σ is

$$D(\rho||\sigma) = \int_{\mathrm{R}^n} \rho(x)[\log \rho(x) - \log \sigma(x)]\mathrm{d}x \ .$$

The evolution equation

$$\begin{aligned} \frac{\partial}{\partial t}\rho(x,t) &= \nabla \cdot (\rho(x,t)[\nabla \log \rho(x,t) - \nabla \log \sigma(x)]) \\ &= \nabla \cdot \left(\rho(x,t)\left[\nabla \frac{\delta}{\delta \rho} D(\rho||\sigma)\right]\right) \end{aligned}$$

is the Kolmogorov forward equation for a diffusion process.

The quantum counterpart

Let $\mathcal{A}$ be a finite-dimensional C^*-algebra with unit $\mathbf{1}$. If you like, take $\mathcal{A} = M_n(\mathbb{C})$, the $n \times n$ matrices over $\mathbb{C}$. Let $\mathfrak{S}_+(\mathcal{A})$ denote the set of faithful states of $\mathcal{A}$: In the matricial case this is the set of all psotive $n \times n$ matrices ρ with unit trace, and the *state* corresponding to ρ is the linear functional $A \mapsto \mathrm{Tr}[\rho A]$.

A *Quantum Markov Semigroup* (QMS) is a continuous one-parameter semigroup of linear transformations

$$(\mathscr{P}_t)_{t \geq 0}$$

on $\mathcal{A}$ such that for each $t \geq 0$, $\mathscr{P}_t$ is *completely positive* and $\mathscr{P}_t \mathbf{1} = \mathbf{1}$.

The QMS $\mathscr{P}_t$ is *ergodic* in case $\mathbf{1}$ spans the eigenspace of $\mathscr{P}_t$ for the eigenvalue 1. In that case, there is a unique invariant state σ.

We consider a class of ergodic QMS that satisfy a quantum *detailed balance condition* with respect to their unique invariant state σ.

We show that all such semigroups (in the Schrödinger picture) are *gradient flow for the relative entropy with respect to a natural analog of the 2-Wasserstein metric*, and we use this to prove new functional inequalities, one of which proves a recent conjecture of Huber, König and Vershynina.

Detailed balance

Let $P_{i,j}$ be the Markov transition matrix for a Markov chain on a finite state space $S = \{x_1, \ldots, x_n\}$. Suppose that σ is a probability density on S with

$$\sigma_j = \sum_{i=1}^{n} \sigma_i P_{i,j} \ .$$

The transition matrix satisfies the *detailed balance condition with respect to* σ in case

$$\sigma_i P_{i,j} = \sigma_j P_{j,i} \qquad \text{for all } i, j \ .$$

The matrix P is self-adjoint on $\mathbb{C}^n$ equipped with the inner product $\langle f, g\rangle_\sigma = \sum_{k=1}^{n} \sigma_k \overline{f_k} g_k$ if and only if the detailed balance condition is satisfied.

There are a number of different ways one might generalize this inner product to the quantum setting, and these give different notions of self-adjointness.

Definition 0.4. Let $\sigma \in \mathfrak{S}_+$ be a non-degenerate density matrix. For each $s \in \mathbb{R}$, and each $A, B \in \mathcal{A}$, define

$$\langle A, B\rangle_s = \mathrm{Tr}[(\sigma^{(1-s)/2} A \sigma^{s/2})^* (\sigma^{(1-s)/2} B \sigma^{s/2})] = \mathrm{Tr}[\sigma^s A^* \sigma^{1-s} B]\,.$$

Let $\sigma \in \mathfrak{S}_+$ and note that

$$\mathrm{Tr}[A^*\Delta_\sigma B] = \mathrm{Tr}[(\Delta_\sigma A)^* B] \quad \text{and} \quad \mathrm{Tr}[A^*\Delta_\sigma A] = \mathrm{Tr}[|\sigma^{1/2} A \sigma^{-1/2}|^2]$$

so that Δ_σ is a positive operator on $\mathfrak{H}_{\mathcal{A}}$.

Since Δ_σ is strictly positive, all eigenvalues of Δ_σ are strictly positive, hence we may write them in the form $e^{-\omega_\gamma}$. Since $(\Delta_\sigma A)^* = \Delta_\sigma^{-1} A^*$, it follows that for all $E \in \mathfrak{H}_{\mathcal{A}}$,

$$\Delta_\sigma E = e^{-\omega} E \quad \iff \quad \Delta_\sigma E^* = e^{\omega} E^* \ .$$

The following is due to Alicki:

Lemma 0.5. *Let $\sigma \in \mathfrak{S}_+$ be a non-degenerate density matrix, and let $s \in [0,1]$, $s \neq 1/2$. Let $\mathcal{K}$ be any operator on $\mathcal{A}$ that is self-adjoint with respect to $\langle \cdot, \cdot \rangle_s$ and also preserves self-adjointness. Then $\mathcal{K}$ commutes with α_t, for all t, real and complex.*

Definition 0.6 (Detailed balance)**.** A QMS $\mathscr{P}_t$ on $\mathcal{A}$ satisfies the *detailed balance condition* with respect to $\sigma \in \mathfrak{S}_+(\mathcal{A})$ in case for each $t > 0$, $\mathscr{P}_t$ is self-adjoint for the σ-GNS inner product. In this case σ is invariant under $\mathscr{P}_t^\dagger$, and we say that the QMS $\mathscr{P}_t$ satisfies the σ-DBC.

A QMS $\mathscr{P}_t = e^{t\mathcal{L}}$ on $\mathcal{A}$ that satisfies the σ-DBC for $\sigma \in \mathfrak{S}_+(\mathcal{A})$ has a generator $\mathcal{L}$ that commutes with the modular operator Δ_σ. Hence Δ_σ and $\mathcal{L}$ can be simultaneously diagonalized.

In the case $\mathcal{A} = \mathcal{M}_n(\mathbb{C})$, the diagonalization of Δ_σ reduces immediately to the diagonalization of σ: Let $\sigma = e^{-h}$ be a density matrix on $\mathbb{C}^n$. Let $\{\eta_1, \dots, \eta_n\}$ be an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of $h = -\log\sigma$:

$$h\eta_j = \lambda_j \eta_j \ .$$

For $\alpha = (\alpha_1, \alpha_2) \in \{(i,j) \ : 1 \leq i,j \leq n\}$, define numbers ω_α (called the *Bohr frequencies*) by

$$\omega_\alpha = \lambda_{\alpha_1} - \lambda_{\alpha_2} \ ,$$

and rank-one operators F_α given by $F_\alpha = |\eta_{\alpha_1} \mathbf{k} \eta_{\alpha_2}|$ where for $\eta, \xi \in \mathbb{C}^n$, $|\eta \mathbf{k} \xi|$ is the rank-one operator sending ζ to $\langle \xi, \zeta \rangle_{\mathbb{C}^n} \eta$. Evidently

$$\Delta_\sigma F_\alpha = e^{-\omega_\alpha} F_\alpha \qquad \text{and} \qquad F_\alpha^* = F_{\alpha'} \quad \text{where} \quad \alpha' = (\alpha_2, \alpha_1) \ .$$

Theorem 0.7 (Alicki). *Let $\mathscr{P}_t = e^{t\mathcal{L}}$ be a QMS on $\mathcal{M}_n(\mathbb{C})$ that satisfies the σ-DBC for $\sigma \in \mathfrak{S}_+$. Then the generator $\mathcal{L}$ of $\mathscr{P}_t$ has the form*

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

where:

(i) $\tau[V_j^* V_k] = c_j \delta_{j,k}$ for all $j, k \in \mathcal{J}$. (ii) $\tau[V_j] = 0$ for all $j \in \mathcal{J}$. (iii) $\{V_j\}_{j\in\mathcal{J}} = \{V_j^*\}_{j\in\mathcal{J}}$. (iv) $\{V_j\}_{j\in\mathcal{J}}$ consists of eigenvectors of the modular operator Δ_σ with

$$\Delta_\sigma V_j = e^{-\omega_j} V_j \ .$$

Conversely, given any set any set $\{V_j\}_{j\in\mathcal{J}}$ satisfying *(i)*, *(ii)*, *(iii)*, the operator $\mathcal{L}$ given by this formula is the generator of a QMS $\mathscr{P}_t$ that satisfies the σ-DBC.

The fact that the operators V_j, $j \in \mathcal{J}$ are eigenfunctions of Δ_σ, and hence Δ_σ^s for all s, has the following consequence:

$$\sigma^s V_j = \sigma^s V_j \sigma^{-s}\sigma^s = (\Delta^s V_j)\sigma^s = e^{-s\omega_j} V_j \sigma^s \ .$$

Differentiating in s at $s = 0$,

$$[V_j, h] = -\omega_j V_j \ .$$

Non-commutative derivatives

Fix such a generator $\mathcal{L}$, and the sets $\{V_j\}_{j\in\mathcal{J}}$ and $\{\omega_j\}_{j\in\mathcal{J}}$ as above.

Define operators ∂_j on $\mathcal{A}$ by

$$\partial_j A = [V_j, A] \quad \text{so that} \quad \partial_j^\dagger A = [V_j^*, A] \ .$$

Define an operator $\mathcal{L}_0$ on $\mathfrak{H}_\mathcal{A}$ by

$$\mathcal{L}_0 A = -\sum_{j\in\mathcal{J}} \partial_j^\dagger \partial_j A = -\sum_{j\in\mathcal{J}} [V_j^*, [V_j, A]] \ .$$

We may write $\mathcal{L}_0 A = -\sum_{j\in\mathcal{J}} (V_j^*[V_j, A] + [A, V_j]V_j^*)$, and hence $\mathcal{L}_0$ is the generator of QMS.

Define the Hilbert space $\mathfrak{H}_{\mathcal{A},\mathcal{J}}$ by

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

where each $\mathfrak{H}_{\mathcal{A}}^{(j)}$ is a copy of $\mathfrak{H}_{\mathcal{A}}$. For $\mathbf{A} \in \mathfrak{H}_{\mathcal{A},\mathcal{J}}$ and $j \in \mathcal{J}$, let A_j denote the component of $\mathbf{A}$ in $\mathfrak{H}_{\mathcal{A}}^{(j)}$. Thus, picking some linear ordering of $\mathcal{J}$, we can write

$$\mathbf{A} = (A_1, \dots, A_{|\mathcal{J}|}) \ .$$

Define an operator $\nabla : \mathfrak{H}_{\mathcal{A}} \to \mathfrak{H}_{\mathcal{A},\mathcal{J}}$ by

$$\nabla A = (\partial_1 A, \dots, \partial_{|\mathcal{J}|} A) \ .$$

We define the operator $\mathrm{div} : \mathfrak{H}_{\mathcal{A},\mathcal{J}} \to \mathfrak{H}_{\mathcal{A}}$ by

$$\mathrm{div}\, \mathbf{A} = -\sum_{j\in\mathcal{J}} \partial_j^\dagger A_j = \sum_{j\in\mathcal{J}} [A_j, V_j^*] \ .$$

Note that div is minus the adjoint of the map $\nabla : \mathfrak{H}_{\mathcal{A}} \to \mathfrak{H}_{\mathcal{A},\mathcal{J}}$, so that $\mathcal{L}_0$ is negative semi-definite. With these definitions, $\mathcal{L}_0 = \mathrm{div} \circ \nabla$. We call ∇ the *non-commutative gradient* associated to $\mathcal{L}$, and div the *non-commutative divergence* associated to $\mathcal{L}$.

Note that each ∂_j is a derivation: For all A, B,

$$\partial_j(AB) = (\partial_j A)B + A\partial_j(B) \ .$$

Chain rule

The evolution equation

$$\frac{\partial}{\partial t}\rho(x,t) = \nabla \cdot (\rho(x,t)[\nabla \log \rho(x,t) - \nabla \log \sigma(x)])$$

is a linear equation because of the chain rule identity

$$\rho \nabla \log \rho = \nabla \rho \ .$$

To obtain a non-commutartive analog, write

$$\rho = \lim_{n \to \infty} \left(\mathbf{1} + \frac{1}{n} \log \rho \right)^n \ .$$

for any V,

$$[V,\rho] = \lim_{n\to\infty} \sum_{m=0}^{n-1} \frac{1}{n} \left(\mathbf{1} + \frac{1}{n}\log\rho\right)^m [V,\log\rho] \left(\mathbf{1} + \frac{1}{n}\log\rho\right)^{n-m-1}$$

$$= \int_0^1 \rho^s [V,\log\rho]\rho^{1-s} \mathrm{d}s \ .$$

The operation $A \mapsto \int_0^1 \rho^s A \rho^{1-s}\mathrm{d}s = R_\rho \int_0^1 \Delta_\rho^s A \mathrm{d}s$, where R_ρ is right multiplication by ρ, is a non-commutative anolog of multiplication by ρ, and it takes self-adjoint operators to self-adjoint operators.

Consider the function f_ω defined by

$$f_\omega(t) := \int_0^1 e^{\omega(s-1/2)} t^s \, \mathrm{d}s = e^{\omega/2} \frac{t - e^{-\omega}}{\log t + \omega} \ .$$

Definition 0.8. For $\rho \in \mathfrak{S}_+$, and $\omega \in \mathbb{R}$, define the operator $[\rho]_\omega : \mathcal{M}_n(\mathbb{C}) \to \mathcal{M}_n(\mathbb{C})$ by

$$[\rho]_\omega = R_\rho \circ f_\omega(\Delta_\rho)$$

For each ω, $[\rho]_\omega$ is invertible, and its inverse, $[\rho]_\omega^{-1} = (1/f_0)(\Delta_\rho) \circ R_{\rho^{-1}}$ may then be viewed as the corresponding non-commutative form of *division by* ρ.

Simple lemmas say:

$$R_\rho f_\omega(\Delta_\rho)\Big(V\log(e^{-\omega/2}\rho) - \log(e^{\omega/2}\rho)V\Big) = e^{-\omega/2}V\rho - e^{\omega/2}\rho V\ .$$

For $\omega = 0$, and $V = V_j$, this is

$$R_\rho f_0(\Delta_\rho)(\partial_j \log\rho) = \partial_j\rho\ .$$

Moreover,

$$\partial_j(\log\rho - \log\sigma) = V_j\log(e^{-\omega_j/2}\rho) - \log(e^{\omega_j/2}\rho)V_j\ .$$

Combining, we can write $\mathcal{L}^\dagger\rho = \sum_{j\in\mathcal{J}}[V^*, e^{-\omega/2}V\rho - e^{\omega/2}\rho V]$ in terms of $D(\rho||\sigma)$.

Theorem 0.9. *Let $\mathscr{P}_t = e^{t\mathcal{L}}$ be QMS on $\mathcal{A}$ that satisfies the σ-DBC for $\sigma \in \mathfrak{S}_+(\mathcal{A})$, and let $\mathcal{L}$ be given in standard from. Then, for all $\rho \in \mathfrak{S}_+$,*

$$-\mathcal{L}^\dagger \rho = \sum_{j \in \mathcal{J}} \partial_j^\dagger \Big([\rho]_{\omega_j} \partial_j (\log \rho - \log \sigma) \Big) \ .$$

We have now arrived at a quantum analog of the classical Kolmogorov forward equation The evolution equation

$$\frac{\partial}{\partial t}\rho(x,t) = \nabla \cdot (\rho(x,t)[\nabla \log \rho(x,t) - \nabla \log \sigma(x)])$$

This is the Kolmogorov forward equation for a diffusion process.

Riemannian metrics on $\mathfrak{S}$.

We now turn to the construction of a Riemannian metric $g_{\mathcal{L}}$ on $\mathfrak{S}$ for which the quantum equation is gradient flow for the relative entropy.

Let $\rho(t), t \in (t_0, t_1)$, be any differentiable path in $\mathfrak{S}_+$ regarded as a convex subset of $\mathcal{A}$. For each $t \in (t_0, t_1)$, let $\dot{\rho}(t) \in \mathcal{A}$ denote the derivative of $\rho(t)$ in t. If $\rho(t)$ is any differentiable path in $\mathfrak{S}_+$ defined on $(-\epsilon, \epsilon)$ for some $\epsilon > 0$ such that $\rho(0) = \rho_0$, then $\mathrm{Tr}[\dot{\rho}(0)] = 0$, so that there is an affine subspace of $\mathfrak{H}_{\mathcal{A},\mathcal{J}}$ consisting of elements $\mathbf{A}$ for which

$$\dot{\rho}(0) = \mathrm{div}\,\mathbf{A}\ .$$

We wish to rewrite this as an analog of the classical continuity equation for the time evolutions of a probability density $\rho(x,t)$ on R^n:

$$\frac{\partial}{\partial t}\rho(x,t) + \mathrm{div}[\mathbf{v}(x,t)\rho(x,t)] = 0 \ .$$

In the classical case, for ρ strictly positive, any expression of the form

$$\frac{\partial}{\partial t}\rho(x,t) = \mathrm{div}[\mathbf{a}(x,t)]$$

gives rise to a continuity equation with $\mathbf{v}(x,t) = -\rho^{-1}(x,t)\mathbf{a}(x,t)$. In the quantum case, there are many different ways to multiply and divide by $\rho \in \mathfrak{S}_+$.

Definition 0.10. Let $\boldsymbol{\omega} \in \mathrm{R}^{|\mathcal{J}|}$. For $\rho \in \mathfrak{S}_+$ we define the linear operator $[\rho]_{\boldsymbol{\omega}}$ on $\mathfrak{H}_{\mathcal{A},\mathcal{J}}$ by

$$[\rho]_{\boldsymbol{\omega}}\big(A_1, \ldots, A_{|\mathcal{J}|}\big) = \big([\rho]_{\omega_1} A_1, \ldots, [\rho]_{\omega_{|\mathcal{J}|}} A_{|\mathcal{J}|}\big) .$$

Note that $[\rho]_{\boldsymbol{\omega}}$ is invertible with

$$[\rho]_{\boldsymbol{\omega}}^{-1}\big(A_1, \ldots, A_{|\mathcal{J}|}\big) = \big([\rho]_{\omega_1}^{-1} A_1, \ldots, [\rho]_{\omega_{|\mathcal{J}|}}^{-1} A_{|\mathcal{J}|}\big) .$$

where we have used the fact that R_ρ and Δ_ρ commute.

Theorem 0.11. *Let $\rho(t)$ be a differentiable path in $\mathfrak{S}_+$ defined on $(-\epsilon, \epsilon)$ for some $\epsilon > 0$ such that $\rho(0) = \rho_0$. Then there is a unique vector field $\mathbf{V} \in \oplus^{|\mathcal{J}|}\mathcal{A}$ of the form $\mathbf{V} = \nabla U$ with $U \in \mathcal{A}$, for which the non-commutative continuity equation*

$$\dot{\rho}(0) = -\operatorname{div}([\rho_0]_{\boldsymbol{\omega}}\mathbf{V}) = -\operatorname{div}([\rho_0]_{\boldsymbol{\omega}}\nabla U) \tag{0.2}$$

holds. Moreover, U can be taken to be traceless, and is then self-adjoint. Furthermore, if $\mathbf{W}$ is any other vector field such that $\dot{\rho}(0) = -\nabla^\dagger([\rho_0]_{\boldsymbol{\omega}}\mathbf{W})$, then

$$\|\mathbf{V}\|_{\mathcal{L},\rho_0} < \|\mathbf{W}\|_{\mathcal{L},\rho_0} \ .$$

Definition 0.12. For each $\rho \in \mathfrak{S}_+$, we identify the tangent space T_ρ at ρ, with the set of gradients vector fields $\{\nabla U \; : \; U \in \mathcal{A} \, , \; U = U^*\}$. We define the Riemannian metric $g_{\mathcal{L}}$ on $\mathfrak{S}_+$ by

$$\|\dot{\rho}(0)\|^2_{g_{\mathcal{L},\rho(0)}} = \|\mathbf{V}\|^2_{\mathcal{L},\rho(0)}$$

where $\dot{\rho}(0)$ and $\mathbf{V}$ are related by (0.2). If $\mathcal{F}$ is any differentiable function on $\mathfrak{S}_+$, the corresponding gradient vector field, denoted $\mathrm{grad}_{g_{\mathcal{L}}} \mathcal{F}(\rho)$ is given by

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{F}(\rho(t)) \bigg|_{t=0} = g_{\mathcal{L}}\big(\dot{\rho}(0), \mathrm{grad}_{g_{\mathcal{L}}} \mathcal{F}(\rho)\big)$$

for all differentiable paths $\rho(t)$ defined on $(-\epsilon, \epsilon)$ for some $\epsilon > 0$ with $\rho(0) = \rho$.

Let $\dfrac{\delta \mathcal{F}}{\delta \rho}(\rho)$ denote the derivative of $\mathcal{F}$: For all self-adjoint $A \in \mathcal{A}$,

$$\lim_{t \to 0} \frac{1}{t}(\mathcal{F}(\rho + tA) - \mathcal{F}(\rho) = \mathrm{Tr}\left[\frac{\delta \mathcal{F}}{\delta \rho}(\rho) A\right] \ .$$

In particular, when

$$\dot{\rho}(0) + \mathrm{div}([\rho_0]_{\boldsymbol{\omega}} \nabla U) = 0$$

is satisfied for some U,

$$\mathrm{Tr}\left[\frac{\delta \mathcal{F}}{\delta \rho}(\rho)\,\mathrm{div}([\rho]_{\boldsymbol{\omega}} \nabla U)\right] = -g_{\boldsymbol{\omega}}\left(\nabla \frac{\delta \mathcal{F}}{\delta \rho}(\rho), \nabla U\right) \ .$$

Theorem 0.13. *Let* $\mathscr{P}_t = e^{t\mathcal{L}}$ *be QMS on* $\mathcal{A}$ *that satisfies the* σ*-DBC for* $\sigma \in \mathfrak{S}_+(\mathcal{A})$*. Then*

$$\frac{\partial}{\partial t}\rho = \mathcal{L}^\dagger \rho \tag{0.3}$$

is gradient flow for the relative entropy $D(\cdot||\sigma)$ *in the metric* $g_{\rho,\boldsymbol{\omega}}$ *canonically associated to* $\mathcal{L}$*.*

Geodesic convexity

We now develop the advantages of having written the evolution equation $\frac{\partial}{\partial t}\rho = \mathcal{L}^\dagger \rho$ as gradient flow for the relative entropy. We draw on work of Otto and Westdickenberg and also of Daneri and Savaré.

Let $(\mathcal{M}, g)$ be any smooth Riemannian manifold. The Riemannian distance $d_g(x,y)$ between x and y is given by

$$d_g^2(x,y) = \inf\left\{ \int_0^1 \|\dot{\gamma}(s)\|^2_{g(\gamma(s))}\,\mathrm{d}s \ :\ \gamma(0) = x,\ \gamma(1) = y \right\} \ ,$$

where

$$\|\dot{\gamma}(s)\|^2_{g(\gamma(s))} = g_{\gamma(s)}(\dot{\gamma}(s), \dot{\gamma}(s)) \ .$$

If F is a smooth function on $\mathcal{M}$, let $\mathrm{grad}_g F$ denote its Riemannian gradient. Consider the semigroup S_t of transformations on $\mathcal{M}$ given by solving $\dot{\gamma}(t) = -\mathrm{grad}_g F(\gamma(t))$; assume that nice global solutions exist. The semigroup S_t, $t \geq 0$, is *gradient flow for* F.
For $\lambda \in \mathbb{R}$, the function F is λ-convex in case whenever $\gamma : [0,1] \to \mathcal{M}$ is a distance minimizing geodesic, then for all $s \in (0,1)$,

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2} F(\gamma(s)) \geq \lambda g(\dot{\gamma}(s), \dot{\gamma}(s)) \ .$$

It is a standard result that whenever F is λ-convex, the gradient flow for F is λ-contracting in the sense that for all $x, y \in \mathcal{M}$ and $t > 0$,

$$\frac{\mathrm{d}}{\mathrm{d}t} d_g^2(S_t(x), S_t(y)) \leq -2\lambda d_g^2(S_t(x), S_t(y)) \ .$$

Otto and Westdickenberg developed an approach to geodesic convexity that takes this contraction as its starting point.

They use the gradient flow transformation S_t to define a one-parameter family of paths $\gamma^t : [0,1] \to \mathcal{M}$, $t \geq 0$ defined by

$$\gamma^t(s) = S_t \gamma(s) \ .$$

From their work and that of Daneri and Savaré, we have that if for all smooth curves $\gamma : [0,1] \to \mathcal{M}$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{0+} \left(\left\| \frac{\mathrm{d}}{\mathrm{d}s}\gamma^t(s) \right\|^2_{g(\gamma^t(s))} \right) \leq -2\lambda \left\| \frac{\mathrm{d}}{\mathrm{d}s}\gamma^0(s) \right\|^2_{g(\gamma^0(s))} ,$$

for all $s \in (0,1)$, then F is geodesically λ-convex.

We now return the QMS setting.

Let $\rho : [0,1] \to \mathfrak{S}_+$ be a smooth path in $\mathfrak{S}_+$, and define the one-parameter family of paths, $\rho^t(s)$, $(s,t) \in [0,1] \times [0,\infty)$ by

$$\rho^t(s) = \mathscr{P}_t^\dagger \rho(s) \ .$$

By what has been explained above, it we can prove that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\left\| \frac{\mathrm{d}}{\mathrm{d}s}\rho^t(s) \right\|^2_{g(\rho^t(s))} \right)\Bigg|_{0+} \leq -2\lambda \left\| \frac{\mathrm{d}}{\mathrm{d}s}\rho^0(s) \right\|^2_{g(\rho^0(s))}$$

for all smooth $\rho : [0,1] \to \mathcal{M}$ and all $s \in (0,1)$, we will have proved the geodesic convexity of the relative entropy functional, and consequently, we shall have proved

$$D(\mathscr{P}_t^\dagger \rho || \sigma) \leq e^{-2\lambda t} D(\rho || \sigma) \ .$$

Lemma 0.14. *Suppose that for some numbers* a_j, $j \in \mathcal{J}$,

$$[\partial_j, \mathcal{L}] = -a_j \partial_j$$

for each $j \in \mathcal{J}$. *Then defining* $\mathscr{P}_t$ *on* $\bigoplus^{|\mathcal{J}|}\mathcal{A}$ *by*

$$\mathscr{P}_t(A_1, \dots, A_{|\mathcal{J}|}) = (e^{-ta_1}\mathscr{P}_t A_1, \dots, e^{-ta_{|\mathcal{J}|}}\mathscr{P}_t A_{|\mathcal{J}|}) ,$$

we have the intertwining relation $\partial_j \mathscr{P}_t = \mathscr{P}_t \partial_j$ *on* $\mathcal{A}$.

Let Z and Z^* be Bose annihilation and creation operators: $[Z, Z^*] = \mathbf{1}$. Define

$$\sigma_\beta = \left(\mathrm{Tr}\left[e^{-\beta h}\right]\right)^{-1} e^{-\beta h} \ .$$

Theorem 0.15. *Let $\mathscr{P}_t$ be the Bose Ornstein-Uhlenbeck semigroup with generator $\mathcal{L}_\beta$ given by*

$$\mathcal{L}_\beta = e^{\beta/2}\left(Z^*AZ - \tfrac{1}{2}\{ZZ^*, A\}\right) + e^{-\beta/2}\left(ZAZ^* - \tfrac{1}{2}\{Z^*Z, A\}\right) \ ,$$

and let σ_β be its invariant state. Then for all $\rho \in \mathfrak{S}_+$,

$$D(\mathscr{P}_t\rho||\sigma_\beta) \leq e^{-2\sinh(\beta/2)t} D(\rho||\sigma_\beta) \ .$$

Let $\{Z_1, \dots, Z_m\}$ and $\{Z_1^*, \dots, Z_m^*\}$ be the sets of annihilation and creation operators for m Fermi degree of freedom:

$$Z_j Z_k + Z_k Z_j = 0 \quad \text{and} \quad Z_j Z_k^* + Z_k^* Z_j = 2\delta_{j,k}\mathbf{1} \quad \text{for all } 1 \leq j, k \leq m \ .$$

Consider the complementary orthogonal projections N_j and $N_j^\perp$ defined by

$$N_j = \frac{1}{2} Z_j^* Z_j \quad \text{and} \quad N_j^\perp = \frac{1}{2} Z_j Z_j^* \quad \text{for all } \ 1 \leq j \leq m \ .$$

For any set of m real numbers $\{e_1, \ldots, e_m\}$, and any parameter $\beta \in (0, \infty)$, to be interpreted as the *inverse temperature*, define the *free Hamiltonian* h and the *Gibbs state* σ_β by

$$h = \sum_{j=1}^{m} e_j N_j \quad \text{and} \quad \sigma_\beta = \frac{1}{\tau[e^{-\beta h}]} e^{-\beta h} \ .$$

where τ is the canonical trace.

Write $Z_j = (Q_j + iP_j)/\sqrt{2}$, and put $W = i^m \prod_{j=1}^{m} Q_j P_j$ so that W is self-adjoint and unitary, and for all $A \in \mathcal{A}$, $\Gamma(A) = WAW$.

Define the operators

$$V_j = W Z_j, \qquad 1 \le j \le m \ ,$$

so that $\{V_1, \dots, V_m, V_1^*, \dots, V_m^*\}$ is set of operators on $\mathcal{A}$ satisfying the conditions *(i)*, *(ii)*, *(iii)* and *(iv)* of the structure theorem. Therefore, the operator $\mathcal{L}_\beta$ defined by

$$\mathcal{L}_\beta A = \frac{1}{2} \sum_{j=1}^{m} \left[e^{\beta e_j/2} V_j^* [A, V_j] + e^{-\beta e_j/2} [V_j, A] V_j^* \right] \ .$$

generates a QMS that is the Fermi Ornstein-Uhlenbeck semigroup at inverse temperature β.

Theorem 0.16. *For $\beta > 0$, let $\mathscr{P}_t$ be the Fermi Ornstein-Uhlenbeck semigroup with generator*

$$\mathcal{L}_\beta A = \frac{1}{2}\sum_{j=1}^{m}\left[e^{\beta e_j/2}V_j^*[A,V_j] + e^{-\beta e_j/2}[V_j,A]V_j^*\right] .$$

and let σ_β be its invariant state. Then for all $\rho \in \mathfrak{S}_+$,

$$D(\mathscr{P}_t\rho||\sigma_\beta) \leq e^{-2\lambda_\beta t}D(\rho||\sigma_\beta)$$

where $\lambda_\beta = \min\{\cosh(\beta e_j/2)\ :\ j = 1,\ldots,m\}$.