Duality for optimal couplings in free probability

David Jekel (joint work with Wilfrid Gangbo, Kyeongsik Nam, and Dimitri Shlyakhtenko)

University of California, San Diego

COSy, May 31, 2021

Acknowledgements

Land Acknowledgement: UCSD stands on the land of the Kumeyaay people.

Funding Acknowledgement: D.J. was supported by the NSF postdoc grant DMS-2002826. W.G. was supported by DMS-1700202 and U.S. Air Force grant FA9550-18-1-0502. D.S. was supported by NSF grant DMS-1762360.

Inspiration acknowledgement: We thank Alice Guionnet, Yoann Dabrowski, Wuchen Li, Ben Hayes, and Srivatsav Kunnawalkam Elayavalli for useful conversations.

Thanks to the conference organizers!

Non-commutative probability spaces

A W^*-*algebra* is a unital C^*-algebra A that is also a dual space (Sakai showed that this is equivalent to a von Neumann algebra).
A tracial W^*-algebra is a pair (A, τ) where A is a W^*-algebra and $\tau : A \to \mathbb{C}$ is a linear functional that is

- Unital: $\tau(1) = 1$,
- Positive: $\tau(a^* a) \geq 0$,
- Tracial: $\tau(ab) = \tau(ba)$,
- Faithful: $\tau(a^* a) = 0 \implies a = 0$.

It is a well-known theorem that every commutative tracial W^*-algebra is isomorphic to $L^\infty(\Omega, P)$ for some probability space (Ω, P), with the trace being given by $\tau(f) = \int f \, dP$.

Hence, a tracial W^*-algebra may be viewed as a non-commutative probability space.

Non-commutative probability spaces

classical	non-commutative
$L^\infty(\Omega, P)$	A
expectation $\mathbb{E}$	trace τ
bounded random variable Z	$Z \in A$
bounded real random variable	$Z \in A_{\text{sa}}$ (self-adjoint)
bounded $\mathbb{R}^m$-valued random variable	$X = (X_1, \dots, X_m) \in A_{\text{sa}}^m$
unbounded random variables	operators affiliated to A
$L^2(\Omega, P)$	$L^2(A, \tau)$

Here $L^2(A, \tau)$ is the Hilbert space from the GNS construction, and this can naturally be identified with the space of square-integrable affiliated operators.

Non-commutative laws

In probability theory, if $(X_1, \ldots, X_m)$ are bounded real, random variables, then their law is a probability measure on $[-R, R]^m$ for some R, that is, a trace on $C[-R, R]^{\otimes m}$.

A *non-commutative law* is a trace on the C^*-universal free product $C([-R, R])^{*m}$ (for some R).

$\Sigma_{m,R}$ denotes this trace space with weak-$*$ topology.

More combinatorially, an element of $\Sigma_{m,R}$ is a unital, positive, tracial map $\mu : \mathbb{C}\langle X_1, \ldots, X_m \rangle \to \mathbb{C}$ satisfying

$$|\mu(X_{i_1} \ldots X_{i_n})| \leq R^n.$$

This encodes the *non-commutative moments* of some *tuple of non-commutative random variables*.

Non-commutative laws

For every $X \in A_{sa}^m$, the *non-commutative law of* X, denoted λ_X, is given by $\lambda_X(p) = \tau(p(X))$ for every non-commutative polynomial p.

In fact, tracial non-commutative laws $\leftrightarrow$ tracial W^*-algebras with a specified generating m-tuple up to isomorphism.

- $\rightarrow$ GNS construction.
- $\leftarrow$ evaluate moments of your generators.

Classical optimal couplings

Given $\mu, \nu \in \mathcal{P}([-R,R]^m)$:

- A coupling of μ and ν is a probability space (Ω, P) and random variables X and Y such that $X \sim \mu$ and $Y \sim \nu$.
- Couplings can alternatively be defined by looking at the joint law π of (X,Y) on $[-R,R]^{2m}$, which is some probability distribution with marginals μ and ν.
- The *cost* of the coupling is $\|X - Y\|_{L^2(\Omega,P)}$.
- A coupling is said to be *optimal* if it achieves the minimal cost.
- The minimal value of the cost is called the Wasserstein distance between μ and ν and is denoted by $d_W(\mu,\nu)$.
- d_W defines a metric on $\mathcal{P}([-R,R]^m)$, and it metrizes the weak-$*$ topology on $\mathcal{P}([-R,R]^m)$.

Classical optimal couplings — intuition

Couplings provide a mathematical description of a *transportation problem*. Imagine that μ represents the distribution of mass in some pile of dirt. We want to move the dirt and rearrange it into the shape of ν. The measure π describes a plan to transport the dirt. The "mass" at (x, y) under π represents the "mass" at the point x for μ that will be moved to the point y for ν.

Hence, an optimal coupling represents a way to transport the dirt with the least possible work. It is a classical example of an optimization problem.

Non-commutative optimal couplings (Biane-Voiculescu)

Given $\mu, \nu \in \Sigma_{m,R}$:

- A coupling of μ and ν is a tracial W^*-algebra (A, τ) and random variables $X, Y \in A_{\text{sa}}^m$ with $\lambda_X = \mu$ and $\lambda_Y = \nu$.
- We can also consider the joint law $\pi = \lambda_{(X,Y)} \in \Sigma_{2m,R}$. By the GNS construction any $\pi \in \Sigma_{2m,R}$ with marginals μ and ν produces a coupling in the above sense.
- The *cost* of the coupling is $\|X - Y\|_{L^2(A,\tau)^m}$.
- A coupling is said to be *optimal* if it achieves the minimal cost. Optimal couplings exist by compactness of $\Sigma_{2m,R}$.
- The minimal value of the cost is called the (non-commutative) *Wasserstein distance of μ and ν* and is denoted by $d_W(\mu, \nu)$.
- d_W defines a metric on $\Sigma_{m,R}$. But does it generate the weak-$*$ topology?

Non-separability of the NC Wasserstein space

Proposition

For $m > 1$, while $\Sigma_{m,R}$ is compact in the weak-$*$ topology, it is *not separable* with respect to the Wasserstein distance.

This relies on the following result:

Theorem (Gromov, Olshanskii, Ozawa)

There exists a group G with property (T) and an uncountable family $(G_\alpha)_{\alpha \in I}$ of non-isomorphic quotients of G.

Let G be as above and let $g_1, \dots, g_m$ be the generators. Let $q_\alpha : G \to G_\alpha$ be the quotient map. Let $X_\alpha \in L(G_\alpha)_{\text{sa}}^{2m}$ be given by the real and imaginary parts of $q_\alpha(g_1), \dots, q_\alpha(g_m)$. Because of property (T), if λ_{X_α} and λ_{X_β} were coupled sufficiently close together, then we would have $\sup_{g \in G} |\tau(q_\alpha(g)) - \tau(q_\beta(g))| < 1$. However, since $\tau(q_\alpha(g)) = \delta_{q_\alpha(g)=e}$, this would imply that $G_\alpha = G_\beta$. Hence, the laws $(\lambda_{X_\alpha})_{\alpha \in I}$ are ϵ-separated in Wasserstein distance for some ϵ.

Classical Monge-Kantorovich duality

Recall that if $f : \mathbb{R}^m \to \mathbb{R}^m$ is convex, then y is said to be *subgradient* to f at x if

$$f(x') - f(x) \geq \langle x' - x, y \rangle \text{ for all } x' \in \mathbb{R}^m.$$

We denote the set of subgradients by $\eth f(x)$.

Theorem (Classical)

Let (Ω, P, X, Y) be a coupling of μ and ν. The following are equivalent:

1. The coupling is optimal.
2. There exists a convex function $f : \mathbb{R}^m \to (-\infty, \infty]$ such that Y is almost surely in $\eth f(X)$.
3. There exist convex functions $f, g : \mathbb{R}^m \to (-\infty, \infty]$ such that $f(x) + g(y) \geq \langle x, y \rangle$ everywhere and $\mathbb{E} f(X) + \mathbb{E} g(Y) = \langle X, Y \rangle_{L^2(\Omega, P)}$.

Classical Monge-Kantorovich duality

Let $C(\mu,\nu)$ be the supremum of $\langle X, Y\rangle_{L^2(\Omega,P)}$ over couplings (Ω, P, X, Y). Since $\|X\|_{L^2(\Omega,P)}$ and $\|Y\|_{L^2(\Omega,P)}$ are determined by μ and ν, minimizing $\|X - Y\|_{L^2(\Omega,P)}$ is equivalent to maximizing $\langle X, Y\rangle_{L^2(\Omega,P)}$. We denote the maximal value of $\langle X, Y\rangle_{L^2(\Omega,P)}$ by $C(\mu,\nu)$.

The supremum $C(\mu,\nu)$ can be expressed as an infimum over pairs of convex functions f and g as in the previous theorem.

Theorem (Classical)

Let $\mu,\nu \in \mathcal{P}([-R,R]^m)$. Then $C(\mu,\nu)$ is the infimum of $\int f\,d\mu + \int g\,d\nu$ over pairs (f,g) of convex functions $f,g : \mathbb{R}^m \to (-\infty,\infty]$ satisfying $f(x) + g(y) \geq \langle x, y\rangle$.

In other words, minimizing $\int f\,d\mu + \int g\,d\nu$ over such pairs (f,g) is the dual problem (in the sense of linear programming) to optimal coupling.

Non-commutative Monge-Kantorovich duality

In adapting this result to the non-commutative setting, the challenge was to find an appropriate analog of convex functions. Although tracial W^*-algebras are an analog of $L^\infty(\Omega, P)$, the non-commutative probability spaces have no points!

Hence, we consider functions that can be evaluated on random variables rather than points. As motivation, if $f : \mathbb{R}^m \to (-\infty, \infty]$ is convex and (Ω, P) is a probability space with σ-algebra $\mathcal{F}$, then we can define

$$\tilde{f} : L^2(\Omega, P; \mathbb{R}^m) \to (-\infty, \infty], \qquad \tilde{f}(X) = \mathbb{E} f(X).$$

Then $\tilde{f}$ is convex. Moreover, by Jensen's inequality if $\mathcal{G}$ is a σ-subalgebra of $\mathcal{F}$, then by Jensen's inequality,

$$\tilde{f}(\mathbb{E}[X|\mathcal{G}]) \leq \tilde{f}(X).$$

Moreover, $f, g : \mathbb{R}^m \to (-\infty, \infty]$ satisfy $f(x) + g(y) \geq \langle x, y \rangle$ if and only if $\tilde{f}(X) + \tilde{g}(Y) \geq \langle X, Y \rangle_{L^2(\Omega, P; \mathbb{R}^m)}$ for all $X, Y \in L^2(\Omega, P; \mathbb{R}^m)$.

Non-commutative Monge-Kantorovich duality

An additional complication of the non-commutative setting is that we have to consider multiple non-commutative probability spaces at the same time.

In the classical setting, every probability space can be modeled in $[0,1]$ with Lebesgue measure. By contrast, Ozawa showed based on Gromov and Olshanskii's work that there is no separable tracial W^*-algebra that contains a copy of all others.

Hence, to study optimal couplings, we need to study functions f that are defined for all tracial W^*-algebras, not only for one particular tracial W^*-algebra.

E-convex functions

To save space, we will use $\mathcal{A}$ to denote a pair (A, τ) of a von Neumann algebra with faithful normal tracial state.

Definition

A *tracial* W^**-function* is a collection $f^{\mathcal{A}} : L^2(\mathcal{A})^m_{\mathrm{sa}} \to [-\infty, \infty]$ of functions for each tracial W^*-algebra such that if $\iota : \mathcal{A} \to \mathcal{B}$ is a trace-preserving unital $*$-homomorphism, then $f^{\mathcal{A}} = f^{\mathcal{B}} \circ \iota$.

Definition

A tracial W^*-function $f = (f^{\mathcal{A}})_{\mathcal{A}}$ with values in $(\infty, \infty]$ is said to be *E-convex* if

1. Each $f^{\mathcal{A}}$ is convex and lower semi-continuous on $L^2(\mathcal{A})^m_{\mathrm{sa}}$.
2. If $\iota : \mathcal{A} \to \mathcal{B}$ is a trace-preserving inclusion and $E = \iota^* : \mathcal{B} \to \mathcal{A}$ is the corresponding trace-preserving conditional expectation, then $f^{\mathcal{A}}(E[X]) \leq f^{\mathcal{B}}(X)$.

Non-commutative Monge-Kantorovich duality

If f is a tracial W^*-function, then $f^{\mathcal{A}}(X)$ only depends on λ_X because f respects inclusions. Hence, $\mu(f)$ is well-defined for $\mu \in \Sigma_{m,R}$.

We call a pair (f, g) of tracial W^*-functions *admissible* if

$$f^{\mathcal{A}}(X) + g^{\mathcal{A}}(Y) \geq \langle X, Y \rangle_{L^2(\mathcal{A})^m_{\mathrm{sa}}}$$

for all $\mathcal{A}$ and all $X, Y \in L^2(\mathcal{A})^m_{\mathrm{sa}}$.

Proposition (GJNS)

For $\mu, \nu \in \Sigma_{m,R}$, the coupling constant $C(\mu, \nu)$ is equal to the infimum of $\mu(f) + \nu(g)$ over all admissible pairs of E-convex functions. Furthermore, a coupling $(\mathcal{A}, X, Y)$ of μ and ν is optimal if and only if there exists such an admissible pair (f, g) with $f^{\mathcal{A}}(X) + g^{\mathcal{B}}(Y) = \langle X, Y \rangle_{L^2(\mathcal{A})^m_{\mathrm{sa}}}$, and in this case both quantities are equal to $C(\mu, \nu)$.

Applications — isomorphism of von Neumann algebras

If (Ω, P, X, Y) is a classical optimal coupling of μ, $\nu \in \mathcal{P}([-R, R]^m)$, then we saw that $Y \in \eth f(X)$ almost surely for some convex function f. If f is differentiable everywhere, then $\eth f(X)$ consists of a single point $\nabla f(X)$, and hence Y can be expressed as a function of X.

In the non-commutative setting, if $(\mathcal{A}, X, Y)$ is an optimal coupling, then $Y \in \eth f^{\mathcal{A}}(X)$ for some E-convex function f. E-convexity implies that $\eth f^{\mathcal{A}}(X)$ contains a point in $L^2(\mathrm{W}^*(X))_{\mathrm{sa}}^m$, and hence if $f^{\mathcal{A}}$ is sufficiently regular, then Y must belong to $\mathrm{W}^*(X)_{\mathrm{sa}}^m$. This led to the following result:

Theorem (GJNS)

Let $(\mathcal{A}, X, Y)$ be an optimal coupling of $\mu, \nu \in \Sigma_{m,R}$. Then for every $t \in (0, 1)$, we have $\mathrm{W}^*((1-t)X + tY) = \mathrm{W}^*(X, Y)$.

Applications — demonstrating optimality

Guionnet and Shlyakhtenko (2014) considered the case where μ is a log-concave free Gibbs law and ν is the law of a semi-circular family S. They showed that there was some type of convex function f such that $\nabla f(S) \sim \mu$ when $S = (S_1, \dots, S_m) \sim \nu$ is a free semicircular family. However, they did not show that $(\mathrm{W}^*(S), S, \nabla f(S))$ is an optimal coupling of ν and μ.

The preprint of J.-Li-Shlyakhtenko (Jan. 2021) used a special case of non-commutative Monge-Kantorovich duality to prove that the coupling was indeed optimal.

One of the main goals of the current work (Gangbo-J.-Nam-Shlyakhtenko) was set up the non-commutative Monge-Kantorovich duality in greater generality.

Connections — quantum information theory

Definition (Anantharaman-Delaroche)

If $\mathcal{A}$ and $\mathcal{B}$ are tracial W^*-algebras, then a completely positive map $\Phi : \mathcal{A} \to \mathcal{B}$ is said to be *factorizable* if it is realized as a trace-preserving inclusion $\mathcal{A} \to \mathcal{C}$ followed by a trace-preserving conditional expectation $\mathcal{C} \to \mathcal{B}$ (and in this case, we say Φ factorizes through $\mathcal{C}$).

Observation (GJNS)

Let $X \in \mathcal{A}_{\mathrm{sa}}^m$ and $Y \in \mathcal{B}_{\mathrm{sa}}^m$ be non-commutative tuples. Then

$$C(\lambda_X, \lambda_Y) = \sup_{\Phi \in \mathrm{FM}(\mathcal{A},\mathcal{B})} \langle \Phi(X), Y \rangle_{L^2(\mathcal{B})_{\mathrm{sa}}^m},$$

where $\mathbb{FM}(\mathcal{A}, \mathcal{B})$ denotes the space of factorizable maps (also known as quantum channels).

Connections — quantum information theory

Theorem (Musat-Rørdam 2020)

For large enough n, there exist factorizable maps $M_n(\mathbb{C}) \to M_n(\mathbb{C})$ that factorize through the hyperfinite II_1 factor but not through any finite-dimensional algebra.

Corollary (GJNS)

For sufficiently large n, for all d, there exist $X, Y \in M_n(\mathbb{C})_{\mathrm{sa}}^{n^2}$ such that an optimal coupling of λ_X and λ_Y requires an algebra of dimension at least d.

Problem

Study the optimal couplings of $X, Y \in M_n(\mathbb{C})_{\mathrm{sa}}^m$ for explicit examples. Can you show directly that d may need to be infinite?

Connections — quantum information theory

Theorem (Haagerup-Musat 2015)

A completely positive map $M_n(\mathbb{C}) \to M_n(\mathbb{C})$ factorizes through a Connes-embeddable tracial W^*-algebra if and only if it can be approximated by maps that factorize through finite-dimensional algebras. Moreover, the Connes-embedding problem has a positive answer if and only if every factorizable map can be approximated by those that factorize through finite-dimensional algebras.

Ji-Natarajan-Vidick-Wright-Yuen 2020 announced a negative solution to the Connes embedding problem, which would imply the following corollary.

Corollary (GJNS)

For sufficiently large n, there exist $X, Y \in M_n(\mathbb{C})_{sa}^{n^2}$ such that every optimal coupling of λ_X and λ_Y uses an non-Connes-embeddable tracial W^*-algebra.

Connections — lifting properties

Proposition (GJNS)

Let $\mu \in \Sigma_{m,R}$, and let $X \in \mathcal{A}_{\text{sa}}^m$ such that $\lambda_X = \mu$ and $\mathcal{A} = \mathrm{W}^*(X)$. Then the following are equivalent:

1. The weak-$*$ topology and the Wasserstein topology on $\Sigma_{m,R}$ agree at at the point μ.
2. Every trace-preserving embedding of $\mathcal{A}$ into a tracial ultraproduct $\prod_{n \to \mathcal{U}} \mathcal{A}_n$ lifts to a sequence of factorizable completely positive maps $\Phi_n : \mathcal{A} \to \mathcal{A}_n$.

Corollary (GJNS)

In the above proposition, if $\mathcal{A}$ is Connes-embeddable, then the weak-$*$ and Wasserstein topologies agree at μ if and only if $\mathcal{A}$ is amenable.

This relies on Connes' 1976 paper of course, and it is related to recent results of Atkinson and Kunnawalkam Elayavalli characterizing amenability through tracial stability properties.

References (in order of appearance) I

1. Wilfrid Gangbo, David Jekel, Kyeongsik Nam, and Dimitri Shlyakhtenko. Duality for optimal couplings in free probability. arXiv:2105.12351
2. Philippe Biane and Dan-Virgil Voiculescu. A free probability analogue of the Wasserstein metric on the trace-state space. Geometric and Functional Analysis, 11:1125–1138, 2001.
3. Narutaka Ozawa. There is no separable universal II_1 factor. Proc. Amer. Math. Soc., 132:487–90, 2004.
4. Alice Guionnet and Dimitri Shlyakhtenko. Free monotone transport. Inventiones Mathematicae, 197(3):613–661, 09 2014.
5. David Jekel, Wuchen Li, and Dimitri Shlyakhtenko. Tracial non-commutative smooth functions and the free Wasserstein manifold. arXiv:2101.06572, 2021.

References (in order of appearance) II

6. Claire Anantharaman-Delaroche. On ergodic theorems for free group actions on noncommutative spaces. Probab. Theory Rel. Fields, 135:520–546, 2006.
7. Magdalena Musat and Mikael Rørdam. Non-closure of quantum correlation matrices and factorizable channels that require infinite dimensional ancilla (with an appendix by Narutaka Ozawa). Comm. Math. Phys., 375:1761–1776, 2020.
8. Uffe Haagerup and Magdalena Musat. An asymptotic property of factorizable completely positive maps and the Connes embedding problem. Comm. Math. Phys., 338(2):721–752, 2015.
9. Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, Henry Yuen. MIP*=RE. arXiv:2001.04383
10. Alain Connes. Classification of injective factors. Cases II_1, II_∞, III_λ, $\lambda \neq 1$. Ann. of Math. (2), 104(1):73–115, 1976.

References (in order of appearance) III

11. Scott Atkinson and Srivatsav Kunnawalkam Ellayavalli. On ultraproduct embeddings and amenability for tracial von Neumann algebras. International Mathematics Research Notices, 2021(4):2882–2918, 2021.