Imperial College London

How to discretise some Optimal Transport problems with linear constraints

Pietro Siorpaes, *Joint work with Marco Massa*

25/03/22, Stochastic Mass Transports, BIRS

The OT problem

Setup

- $\mathbb{B}$ separable Banach space (if $\mathbb{B} = \mathbb{R}^n$ take Euclidean $||\cdot||$).
- μ, ν proba. on $\mathbb{B}$ with finite r-moment, $r \in [1, \infty)$.

A proba. θ on $\mathbb{B} \times \mathbb{B}$ is a *transport* from μ to ν ($\theta \in \Theta(\mu, \nu)$) if it has marginals μ, ν, i.e. if θ is the law of $(X, Y) : X \sim \mu$, $Y \sim \nu$.

Given a cost function v, the Optimal Transport problem is:

$$p(\mu, \nu) := \inf_{\theta \in \Theta(\mu,\nu)} \int v d\theta = \inf_{X \sim \mu, Y \sim \nu} \mathbb{E} v(X, Y). \tag{OT}$$

If $v(x, y) = ||x - y||^r$ then

$$W_r(\mu, \nu) := p(\mu, \nu)^{\frac{1}{r}}$$

is the Wasserstein (a.k.a. Monge-Kantorovich) distance.

Often one takes v cont. and s.t. $0 \le v \le c(1 + ||x||^r + ||y||^r)$.

Variants of OT

The OT problem admits many interesting variants, e.g.:

- μ, ν defined on different spaces
- Multiple Marginals $\mu_1, \ldots, \mu_N$
- Unbalanced OT: $\mu(B) \neq \nu(B)$

Some variants impose additional (linear) constraints, e.g.:

1. OT with capacity contraints: $\frac{d\theta}{d\mathcal{L}^{2n}} \leq c$
2. Invariant OT: $\theta = \theta \circ g^{-1} \, \forall g \in G$, G group acting on $\mathbb{B} \times \mathbb{B}$
3. Martingale OT: $\mathbb{E}^{\theta}[Y|X] = X$.
4. Causal OT: $\mathbb{P}\left((Y_1, \ldots, Y_t) \in B \mid X_1, \ldots, X_N\right) = \mathbb{P}\left((Y_1, \ldots, Y_t) \in B \mid X_1, \ldots X_t\right)$ for all meas. B

See respectively e.g.: Kormal and McCann ('14), Zaev ('15), Beiglböck, Henry-Labordère, Penkner ('13), Backhoff, Beiglböck, Lin and Zalashko ('17).

Discretisation of measures

How can one approximate μ with *finitely supported* $\hat{\mu}$?

The Optimal r-Quantisation problem of order k:

$$\inf\{W_r(\mu,\hat{\mu}) : \hat{\mu} \text{ proba } : \# \text{ supp } (\hat{\mu}) \leq k\} \qquad \text{(OQ)}$$

Discretisation which satisfy additional constraints often exist:

Tchakaloff ('57), Beiglböck, Nutz ('14)

If $\mathbb{B} = \mathbb{R}^n$, given $f \in L^1(\mu;\mathbb{R}^m)$ there exists proba. $\hat{\mu}$ s.t.

$$\#supp(\hat{\mu}) \leq b_n^m, \qquad \text{supp } \hat{\mu} \subseteq \text{ supp } \mu, \qquad \int f d\mu = \int f d\hat{\mu}. \qquad \text{(C)}$$

Let $\mathcal{M}(x_0)$ be the family of laws of martingales $(M_0,\ldots,M_K)$ s.t. $M_0 = x_0$. If $\mu \in \mathcal{M}(x_0)$ then $\exists \hat{\mu} \in \mathcal{M}(x_0)$ s.t. (C) holds.

Discretisation of the OT problem

Applications of discretisation to OT?

If μ, ν have *finite support*, then (OT) is a finite-dimensional LP, so it can be solved numerically (with great efficiency if an entropic regularisation is considered).

To compute $p(\mu, \nu)$, construct fin. sup. proba. $(\hat{\mu}^k, \hat{\nu}^k) \to (\mu, \nu)$ such that $p(\mu, \nu) = \lim_k p(\hat{\mu}^k, \hat{\nu}^k)$. Then easily compute $p(\hat{\mu}^k, \hat{\nu}^k)$, so get $p(\mu, \nu)$.

Can one adapt the above method to constrained OT ?

Discretising constrained OT

Let $\Theta_c(\mu, \nu)$ be the set of *constrained* transports from μ to ν.
Call (μ, ν) *viable* if $\Theta_c(\mu, \nu) \neq \emptyset$.

Questions :

(Q1) If (μ, ν) viable, can find viable fin. sup. $(\hat{\mu}^k, \hat{\nu}^k) \to (\mu, \nu)$?

(Q2) How can $(\hat{\mu}^k, \hat{\nu}^k)$ be computed?

(Q3) Given $(\hat{\mu}^k, \hat{\nu}^k) \to (\mu, \nu)$ as in (Q1), if

$$p_c(\mu, \nu) := \inf_{\Theta_c(\mu,\nu)} \mathbb{E} v(X, Y)$$

does $p_c(\hat{\mu}^k, \hat{\nu}^k) \to p_c(\mu, \nu)$?

(Q4) Can choose $(\hat{\mu}^k, \hat{\nu}^k)$ which satisfies optimality property?

(Q5) Can choose $(\hat{\mu}^k, \hat{\nu}^k)$ which satisfies additional constraints ?

Martingale OT and Strassen's Thm

We focus on MOT; if (μ, ν) fin. supp. it is an LP, which can be solved efficiently with entropic regularisation, see De March ('18).

Let $\mathcal{M}(\mu, \nu) := \Theta_c(\mu, \nu)$ be the set of martingale transports from μ to ν, i.e. $\theta \in \mathcal{M}(\mu, \nu)$ if:

θ law of $(X, Y) : X \sim \mu, Y \sim \nu, \mathbb{E}^\theta[Y|X] = X$, or equiv. if

$\theta \in \Theta(\mu, \nu) : \int g(x)(y - x) d\theta(x, y) = 0 \quad \forall g$ cont. bdd.

Strassen's Thm ('65)

$\mathcal{M}(\mu, \nu) \neq \emptyset \iff \mu \leq_c \nu \iff \ldots$

$\mu \leq_c \nu$ means $\int f d\mu \leq \int f d\nu$ for all $f : \mathbb{B} \to \mathbb{R}$ convex cont.

Discretisations preserving the convex order

So, (Q1) and (Q2) become: given $\mu \leq_c \nu$, $\exists$ fin. sup. proba. $(\hat{\mu}^k, \hat{\nu}^k) \to (\mu, \nu)$ s.t. $\hat{\mu}^k \leq_c \hat{\nu}^k$? How can one compute them?

Find discretisation $D_k : \{\text{Proba.}\} \to \{\text{Proba. on } k \text{ points}\}$ preserving $\leq_c$, take $\hat{\mu}_k = D_k(\mu), \hat{\nu}_k = D_k(\nu)$. Known D_k's:

1. $D_k(\alpha) := \frac{1}{k} \sum_{i=1}^k \delta_{x_i(\alpha)}$, where $x_i(\alpha) := k \int_{\frac{i-1}{k}}^{\frac{i}{k}} F_\alpha^{-1}(t) dt$

   Baker ('12): *Considers only* $\mathbb{B} = \mathbb{R}$

2. Pagès and Wilbertz ('12). *Defined for* $\mathbb{B} = \mathbb{R}^n$*, but preserves* $\leq_c$ *only for* $n = 1$*. Defined only for proba. with cpt. supp.. Does not generalise to several marginals.*

Other ways?

3. Apply different operators to μ and ν.
4. Relax convex order/martingale constraint

Discretisation via Sampling and projections

Alfonsi, Corbetta, Jourdain ('19):
Given given $\mu \leq_c \nu$ on $\mathbb{R}^n$, and arbitrary fin. sup. proba. $(\hat{\mu}^k, \hat{\nu}^k) \to (\mu, \nu)$, replace $\hat{\mu}^k$ with its W_r-projection $\hat{\alpha}^k$ on $\{\alpha : \alpha \leq_c \hat{\nu}^k\}$, then $(\hat{\alpha}^k, \hat{\nu}^k) \to (\mu, \nu)$.

Analog.: can replace $\hat{\nu}^k$ with its W_r-projection $\hat{\beta}^k$ on $\{\beta : \hat{\mu}^k \leq_c \beta\}$, then $(\hat{\mu}^k, \hat{\beta}^k) \to (\mu, \nu)$.

$\hat{\beta}^k$ cannot be computed. If $\hat{\mu}^k$ is the empirical meas. $\frac{1}{k}\sum_{i=1}^k \delta_{X_i}$ with $X_i \sim \mu$ IDD, and analog. $\hat{\nu}^k$, then $\hat{\alpha}^k$ can be computed numerically, and $(\hat{\alpha}^k, \hat{\nu}^k) \to (\mu, \nu)$ a.s..

Guo and Obłój ('19):
Although $\mathbb{E}[Y|X] = X$, only ask that $||\mathbb{E}[Y^k|X^k] - X^k||_{L^1} \to 0$

Our approach: discretise martingales !

Instead of (Q1),(Q2), consider the analog. statement for rv:

Given $X, Y \in L^1(\mathbb{P};\mathbb{B})$ such that $\mathbb{E}[Y|X] = X$, how to build finitely valued $X^k, Y^k \in L^1(\mathbb{P};\mathbb{B})$ s.t.

$$\mathbb{E}[Y^k|X^k] = X^k, \quad (X^k, Y^k) \to (X, Y) \text{ in } L^1?$$

Idea: given $C(k)$ partition of $\mathbb{B}$ with k elements and s.t. $\mathcal{B}^k := \sigma(C(k)) \uparrow \mathcal{B}(\mathbb{B})$, let

$$X^k := \mathbb{E}[X|\sigma^k(X)], \quad Y^k := \mathbb{E}[Y|\sigma^k(X, Y)];$$

$\sigma^k(X) = X^{-1}(\mathcal{B}^k)$ is the smallest σ-alg. s.t. X is $\mathcal{B}^k$-meas (resp. $\sigma^k(X, Y) = (X, Y)^{-1}(\mathcal{B}^k \times \mathcal{B}^k)$...$(X, Y)$ is $\mathcal{B}^k \times \mathcal{B}^k$-meas).

Proof: Clearly $\#Im(X^k) \leq k$ and $\#Im(Y^k) \leq k^2$. The tower property gives $\mathbb{E}[Y^k|X^k] = X^k$. Since $\sigma^k(X) \uparrow \sigma(X)$ and $\sigma^k(X, Y)) \uparrow \sigma(X, Y)$, by martingale convergence thm $(X^k, Y^k) \to (X, Y)$ in L^1.

Evaluating our approach

Pros:

- simple proof
- works for infinite dimensional $\mathbb{B}$
- explicit expression of X^k, Y^k
- can easily be computed numerically by evaluating integrals
- outputs non-random $(\hat{\mu}^k, \hat{\nu}^k)$

Cons:

- Needs a $\theta \in \mathcal{M}(\mu, \nu)$ as an input. Only μ, ν are given, but one such θ can be calculated: if $\mathbb{B} = \mathbb{R}$ in many ways, if $\mathbb{B} = \mathbb{R}^n$ by extending Bass' construction (Henry-Labordère)

Optimality: link with Voronoi's quantisation

Theorem

If $\mathbb{B} = \mathbb{R}^n$, $\exists \bar{\mathcal{B}}^k = \sigma(\bar{C}(k))$ which minimises $||X - X^k||_{L^2}$, and it is given by the optimal Voronoi 2-quantisation of μ or order k.

Sketch of Proof: $\mathcal{S}_k := \{f : \mathbb{B} \to \mathbb{B} : \# Im(f) \leq k\}$ k-simple fns.

$$\mathcal{S}_k = \{S_C^b(x) := \sum_{i=1}^k b^i 1_{C^i}(x), C := (C^i)_{i=1}^k \text{ } k\text{-partition of } \mathbb{B}\}.$$

Call $b^i \in \mathbb{B}$ 'point', and $C^i \subseteq \mathbb{B}$ 'cell'. Fix $b = (b^i)^i$. Clearly the Voronoi partition

$$\bar{C}_i(b) := \{x : ||x - b^i|| = \min_j ||x - b^j||\}$$

minimizes $||S_C^b(x) - x||$ at each x over all k-partitions; in partic. $\bar{C}(b)$ minimizes $||S_C^b(X) - X||_{L^r}$.

Proof of optimal Voronoi quant.= optimal mart. quant.

Let $\bar{b}$ minimize

$$f(b) := \min_C ||S_C^b(X) - X||_{L^r};$$

then $S_{\bar{C}}^{\bar{b}}$ solves $\inf_{f \in \mathcal{S}_k} ||f(X) - X||_{L^r}$, which solves (OQ) if $X \sim \mu$ has density. $S_{\bar{C}}^{\bar{b}}$ is the optimal Voronoi quantisation.

Let us instead first fix C and minimise over b; if $r = 2$ then the 'martingale quantisation' $\mathbb{E}[X|\sigma(C)]$ equals

$$\min_b ||S_C^b(X) - X||_{L^r}, \text{ solved by } \tilde{b}^i = \text{bar}\mu(\cdot|C^i).$$

The optimal martingale quantisation is given by $\tilde{C}$ which minimizes $||\mathbb{E}[X|\sigma(C)] - X||_{L^r}$. Since $\inf_b \inf_C = \inf_C \inf_b$ we get $\tilde{C} = \bar{C}, \tilde{b} = \bar{b}$, i.e. optimal Voronoi quant.=optimal mart. quant. $\square$

Generalisation of martingale discretisation

More generally: take any finitely valued $Y^k \to Y$ in L^1, define $X^k := \mathbb{E}[Y^k | \sigma^k(X)]$, then

$$(X^k, Y^k) \to (X, Y) \text{ in } L^1, \quad \mathbb{E}[Y^k | X^k] = X^k.$$

Useful if want Y^k to have fewer than k^2 values; however, link to optimal quantisation is lost.

Could be useful to satisfy additional constraints, since if $\mathbb{B} = \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}^m$ then we know $\exists Y^k$ s.t. $\mathbb{E}f(Y) = \mathbb{E}f(Y^k)$; however, we don't normally know how to compute such Y^k.

Analog, given $g : \mathbb{R}^{2n} \to \mathbb{R}^m$ we know $\exists (X^k, Y^k)$ in L^1 s.t. $\mathbb{E}g(X, Y) = \mathbb{E}g(X^k, Y^k)$ and $\mathbb{E}[Y^k | X^k] = X^k$...but we don't know how to compute (X^k, Y^k).

Stability of Martingale OT

Backhoff-Veraguas and Pammer ('19):
If $(\mu^k, \nu^k) \to (\mu, \nu)$ and $v^k \to v \geq 0$ uniformly then

$$\inf_{\mathcal{M}(\mu^k,\nu^k)} \mathbb{E}v^k(X, Y) \to \inf_{\mathcal{M}(\mu,\nu)} \mathbb{E}v(X, Y) \tag{1}$$

holds if $\mathbb{B} = \mathbb{R}$, and 'We think that our approach can also be adapted to cover higher dimensions.'

Remark

Let π^k be a martingale law with $\pi^k \to \pi^*$ and with marginals (μ^k, ν^k), then

$$\mathbb{E}^{\pi^*}(v(X, Y)) \geq \liminf_k \inf_{\mathcal{M}(\mu^k,\nu^k)} \mathbb{E}v^k(X, Y) \geq \inf_{\mathcal{M}(\mu,\nu)} \mathbb{E}v(X, Y)$$

so (1) holds along minimising subsequence if $\pi^* \in \operatorname{argmin}_{\pi \in \mathcal{M}(\mu,\nu)} \mathbb{E}^{\pi}(v(X, Y))$

Summary

1. Given $\mu \leq_c \nu$, we found simple construction of fin. sup. proba. $(\hat{\mu}^k, \hat{\nu}^k) \to (\mu, \nu)$ s.t. $\hat{\mu}^k \leq_c \hat{\nu}^k$. This construction admits several variants.
2. $(\hat{\mu}^k, \hat{\nu}^k)$ can be chosen to satisfy some optimality property, e.g. $\hat{\mu}^k$ is the Voronoi quantisation of μ and so it minimises $W_2(\cdot, \mu)$ over $\{\hat{\mu} : \#\text{supp}\hat{\mu} \leq k\}$.
3. We are working on satisfying additional constraints and optimality properties. Once done, we'll submit.