Causal Transport in Discrete Time and Applications

Anastasiia Zalashko

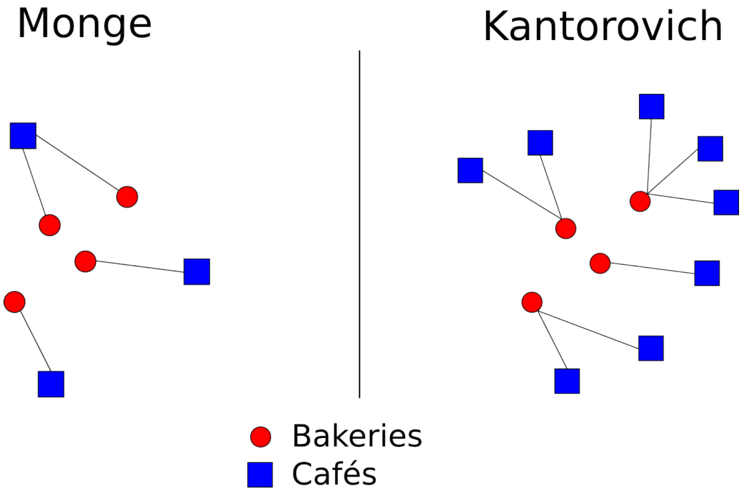
University of Vienna
Faculty of Mathematics

joint work with

Dr. Julio Backhoff (Vienna University of Technology)
Prof. Mathias Beiglböck (Vienna University of Technology)
Dr. Yiqing Lin (École Polytechnique)

June 29, 2016

Bakeries and croissants



Mathematical formulation

Production at the bakeries = source probability measure $\mu \in \mathcal{P}(\mathcal{X})$

Demand at the cafés = target probability measure $\nu \in \mathcal{P}(\mathcal{Y})$

Price for delivery of croissants from bakery x to café y - Borel cost function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+ \cup \{+\infty\}$

Transport maps - $T : \mathcal{X} \to \mathcal{Y}$ such that $T\#\mu = \nu$,

Transport plans - probability measures
$\gamma \in \Pi(\mu, \nu) = \{\gamma \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}), p_1\#\gamma = \mu, p_2\#\gamma = \nu\}$.

Optimal transport with linear constraints

Zaev D., On the Monge–Kantorovich Problem with Additional Linear Constraints (2015)

Let $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$, $\mathbb{F} \subset C_b(\mathcal{X} \times \mathcal{Y})$

$$\inf \left\{ \int_{\mathcal{X}\times\mathcal{Y}} c d\gamma; \quad \gamma \in \Pi(\mu, \nu), \quad \int f d\gamma = 0, f \in \mathbb{F} \right\}.$$

Example: Martingale optimal transport

$$\Pi_M(\mu, \nu) = \left\{ \gamma \in \Pi(\mu, \nu) : \int y d\gamma^x(y) = x \right\}.$$

$$\inf_{\gamma \in \Pi_M(\mu,\nu)} \int c d\gamma =$$

$$\inf \left\{ \int c d\gamma; \gamma \in \Pi(\mu, \nu), \int g(x)(y - x) d\gamma = 0, g(x) \in C_b(\mathcal{X}) \right\}.$$

Causality on $\mathbb{R}^N$

Transport between discrete time stochastic processes

$$(X_1, \ldots, X_N) \quad \text{to} \quad (Y_1, \ldots, Y_N)$$

in such a way that

$$(Y_1, \ldots, Y_t) \perp\!\!\!\perp_{(X_1, \ldots, X_t)} (X_{t+1}, \ldots, X_N).$$

(classical) Transport maps - $T : \mathcal{X} \to \mathcal{Y}$ such that $T \# \mu = \nu$

$T : \mathbb{R}^N \mapsto \mathbb{R}^N$ is causal if and only if it is adapted, i.e. there exists $T^t : \mathbb{R}^t \mapsto \mathbb{R}$ - Borel measurable s.t. for μ-a.e. $(x_1, \ldots, x_N)$

$$T(x_1, \ldots, x_N) = (T^1(x_1), T^2(x_1, x_2), \ldots, T^N(x_1, \ldots, x_N))$$

Motivation

Lassalle R., Causal transference plans and their Monge-Kantorovich problems

Weak solutions (X_t, B_t) of SDE

$$dX_t = dB_t + b_t(X_t)dt; \quad X_0 = 0.$$

are optimizers of the transport problem

$$\inf \left\{ \int |x-y|^2_{CM} d\gamma, \gamma \in \Pi_c(\mu,\nu) \right\},$$

with Cameron-Martin cost and μ - Wiener measure

(B, X) are causal couplings

$$(W_2\text{-distance})^2 \leq (\text{ causal } W_2\text{-distance})^2 = 2\,\mathcal{H}(\mu|\nu)$$

Talagrand inequality on Wiener space

Causal transportation problem

Given μ, ν, cost c defined on $\mathbb{R}^N \times \mathbb{R}^N$,
Causal transportation problem

$$\inf_{\gamma \in \Pi_c(\mu,\nu)} \int c d\gamma \tag{Pc}$$

Transport plans that are causal in both directions are called *bicausal*.

$$\Pi_{bc}(\mu, \nu) = \{\gamma \in \Pi_c(\mu, \nu) \text{ s.t. } e\#\gamma \in \Pi_c(\nu, \mu)\},$$

Bicausal transportation problem

$$\inf_{\gamma \in \Pi_{bc}(\mu,\nu)} \int c d\gamma \tag{Pbc}$$

Rüschendorf L., The Wasserstein distance and approximation theorems (1985)
Pflug G., Version-independence and nested distributions in multistage stochastic optimization (2009)

Goals/ Results

- Primal attainability
- Duality
- Recursive factorization
- Identification of causal Brenier map
- Multistage stochastic optimization

Characterization of causal plans

Proposition (Part 1)

The following are equivalent:

1. $\gamma \in \Pi_c(\mu, \nu)$,
2. $\gamma \in \Pi(\mu, \nu)$ *and for all* $t \in \{1, \dots, N\}$, $h \in C_b(\mathbb{R}^t)$ *and* $g \in C_b(\mathbb{R}^N)$

$$\int h(y_1, \dots, y_t)\Big\{ g(x_1, \dots, x_N) - \int g(x_1, \dots, x_t, \bar{x}_{t+1}, \dots, \bar{x}_N)\mu^{x_1, \dots, x_t}(d\bar{x}_{t+1}, \dots, d\bar{x}_N)\Big\} d\gamma = 0$$

Characterization of causal plans

Proposition

The following are equivalent:

1. $\gamma \in \Pi_c(\mu, \nu)$,
2. $\gamma \in \Pi(\mu, \nu)$ *and for every bounded continuous* $\mathcal{F}^{\mathcal{Y}}$*-adapted process* H *and each bounded* $\mathcal{F}^{\mathcal{X}}$*-martingale* M

$$\int \sum_{t<N} H_t(y_1, \ldots, y_t) \left[M_{t+1}(x_1, \ldots, x_{t+1}) - M_t(x_1, \ldots, x_t) \right] d\gamma = 0.$$

Causal Duality

Theorem (Backhoff, Beiglböck, Lin, Z.)

Let $c : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ *is lower-semicontinuous, bounded from below. Then*

$$\inf_{\gamma \in \Pi_c(\mu,\nu)} \int c d\gamma = \sup_{\substack{\Phi \in C_b(\mathbb{R}^N), \Psi \in C_b(\mathbb{R}^N), F \in \mathbb{F} \\ \Phi + \Psi \leq c + F}} \left[\int \Phi d\mu + \int \Psi d\nu \right]$$

Moreover, the infimum in the l.h.s. is attained.

Characterization of causal plans - kernels

Proposition (Part 2)

Write

$$\gamma(dx_1, \dots, dx_N, dy_1, \dots, dy_N) \\ = \bar{\gamma}(dx_1, dy_1)\gamma^{x_1,y_1}(dx_2, dy_2) \dots \gamma^{x_1,\dots,x_{N-1},y_1,\dots,y_{N-1}}(dx_N, dy_N). \tag{1}$$

Then $\gamma \in \Pi_c(\mu, \nu)$ *iff* $\bar{\gamma} \in \Pi(p_1\#\mu, p_1\#\nu)$*,*
γ*-almost every* $x_1, \dots, x_t, y_1, \dots, y_t$

$$p_1\#\gamma^{x_1,\dots,x_t,y_1,\dots,y_t} = \mu^{x_1,\dots,x_t},$$

and for ν*-almost all* $y_1, \dots, y_t$

$$\gamma^{y_1,\dots,y_t}(dy_{t+1}) = \nu^{y_1,\dots,y_t}(dy_{t+1})$$

Recursive formulation

Theorem (Backhoff, Beiglböck, Lin, Z.)

Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^N)$. If $\gamma \in \Pi_{bc}(\mu, \nu)$ is decomposed as in (1), then it holds:

$$(Pbc) = \inf_{\gamma^1 \in \Pi(p_1 \# \mu, p_1 \# \nu)} \int \gamma^1(dx_1, dy_1) \inf_{\gamma^2 \in \Pi(\mu^{x_1}, \nu^{y_1})} \int \gamma^2(dx_2, dy_2) \dots \inf_{\gamma^N \in \Pi(\mu^{x_1,\dots,x_{N-1}}, \nu^{y_1,\dots,y_{N-1}})} \int \gamma^N(dx_N, dy_N) \, c(x_1, \dots, x_N, y_1, \dots, y_N).$$

Knothe-Rosenblatt rearrangement

Definition

The increasing N-dimensional Knothe-Rosenblatt rearrangement of μ and ν is defined as the law of the random vector $(X_1^*, \dots, X_N^*, Y_1^*, \dots, Y_N^*)$ where

$$X_1^* = F_{\mu_1}^{-1}(U_1), \qquad Y_1^* = F_{\nu_1}^{-1}(U_1), \qquad \text{and inductively}$$

$$X_n^* = F_{\mu^{X_1^*, \dots, X_{n-1}^*}}^{-1}(U_n), \quad Y_n^* = F_{\nu^{Y_1^*, \dots, Y_{n-1}^*}}^{-1}(U_n), \text{ for } n = 2, \dots, N, \tag{2}$$

for $U_1, \dots, U_N$ independent and uniformly distributed random variables on $[0, 1]$.

Main theorem

Theorem (Backhoff, Beiglböck, Lin, Z.)

Assume that c is l.s.c. bounded from below and has a separable structure

$$c(x_1, \dots, x_N, y_1, \dots, y_N) = \sum_{t \le N} c_t(x_t, y_t).$$

Suppose that μ is the product of its marginals, i.e.

$$\mu(dx_1, \dots, dx_N) = \mu_1(dx_1) \dots \mu_N(dx_N).$$

Then the values of (Pc) *and* (Pbc) *coincide.*
If $c_t(x, y) = c_t(x - y)$ and c_t is convex, then a solution to (Pc) *is given by the Knothe-Rosenblatt rearrangement*

Multistage Stochastic programming

Consider the value function $v(\eta)$ of stochastic optimization problem

$$v(\eta) := \tag{3}$$

$$\inf_{u_1(),\ldots,u_N()} \int H(x_1,\ldots,x_N,u_1(x_1),u_2(x_1,x_2),\ldots,u_N(x_1,\ldots,x_N))d\eta, \tag{4}$$

Theorem (Pflug, Pichler, 2012)

Let the objective function H be r-Lipschitz in its first argument and convex in its second, then

$$|v(\mu)-v(\nu)| \le r \inf_{\gamma\in\Pi_{bc}(\mu,\nu)} \int ||x-y||_1 d\gamma(x,y)$$

Further bounds

Theorem (Backhoff, Beiglböck, Lin, Z.)

Let $\mu \in \mathcal{P}(\mathbb{R}^N)$ *be a measure with bounded support, then for all* $\nu \in \mathcal{P}(\mathbb{R}^N)$

$$|v(\mu) - v(\nu)| \leq rK\sqrt{Ent(\nu|\mu)},$$

with $v(\cdot)$ *defined as in* (3).

Further bounds

Theorem (Backhoff, Beiglböck, Lin, Z.)

Let $\mu \in \mathcal{P}(\mathbb{R}^N)$ *be a measure with bounded support, then for all* $\nu \in \mathcal{P}(\mathbb{R}^N)$

$$|v(\mu) - v(\nu)| \leq rK\sqrt{Ent(\nu|\mu)},$$

with $v(\cdot)$ *defined as in* (3).

Thank you!

References I

R. Lassalle, *Causal transference plans and their monge-kantorovich problems*, arXiv:1303.6925, 2013.

G. Ch. Pflug, *Version-independence and nested distributions in multistage stochastic optimization*, SIAM Journal on Optimization **20** (2009), no. 3, 1406–1420.

G. Ch. Pflug and A. Pichler, *A distance for multistage stochastic optimization models*, SIAM J. Optim. **22** (2012), no. 1, 1–23. MR 2902682

L. Rüschendorf, *The Wasserstein distance and approximation theorems*, Z. Wahrsch. Verw. Gebiete **70** (1985), no. 1, 117–129. MR 795791 (86m:60004)

D. Zaev, *On the Monge–Kantorovich Problem with Additional Linear Constraints*, Mat. Zametki **98** (2015), no. 5, 664–683. MR 3438523

DPP for causal

Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^N)$ and μ law of a Markov process. If one has kernels s.t.

1. $m_{t+1}^{x_1,\dots,x_t,y_1,\dots,y_t}(dx_{t+1}, dy_{t+1}) = m_{t+1}^{x_t,y_1,\dots,y_t}(dx_{t+1}, dy_{t+1})$,
2. $m_{t+1}^{x_t,y_1,\dots,y_t}(dx_{t+1}, \mathbb{R}) = \mu^{x_t}(dx_{t+1})$,
3. $\int\limits_{x_t \in \mathbb{R}} m_{t+1}^{x_t,y_1,\dots,y_t}(\mathbb{R}, dy_{t+1}) m_{1\dots t}^{y_1,\dots,y_t}(dx_t) = \nu^{y_1,\dots,y_t}(dy_{t+1})$,

then

$$\tilde{\gamma}(dx_1, \dots, dx_N, dy_1, \dots, dy_N) = m_1(dx_1, dy_1) m_2^{x_1,y_1}(dx_2, dy_2) \dots m_N^{x_{N-1},y_1,\dots,y_{N-1}}(dx_N, dy_N)$$

is a causal plan between μ and ν.

DPP for causal

Theorem

$\mu, \nu \in \mathcal{P}(\mathbb{R}^3)$, μ law of a Markov process. Then

$$(Pc) = \inf_{m_1} \int \nu(dy_1) \inf_{m_2^{x_1}} \int m_1^{y_1}(dx_1) \int m_2^{x_1}(dy_2)$$

$$\inf_{m_3^{x_2}} \int m_2^{x_1,y_2}(dx_2) \int m_3^{x_2}(dx_3, dy_3) c(x_1, x_2, x_3, y_1, y_2, y_3),$$