Entropic Regularization and Iterative Scaling for Unbalanced Optimal Transport – A Reprise

Sinkhole Algorithm

Max von Renesse

Joint work with

Bernd Sturmfels, Simon Telen, and François-Xavier Vialard

IG4DS Conference
Sept 2022 TU Hamburg

Outline

*'Sinkhole Algorithm'

Monge-Kantorovich-Problem

Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$

$$T_2(\mu,\nu) = \inf \left\{ \iint_{\mathbb{R}^d \times \mathbb{R}^d} d^2(x,y)\, \gamma(dx,dy) \,|\gamma \in \Gamma_{\mu,\nu} \right\},$$

$$\Gamma_{\mu,\nu} = \{\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) \,|\, (\pi_1)_\# \gamma = \mu, (\pi_2)_\# \gamma = \nu\}$$

$$\mu \otimes \nu \in \Gamma_{\mu,\nu} \neq \emptyset$$

Definition ('Unbalanced' case)

$\rightsquigarrow$ What if $\langle \mu \rangle \neq \langle \nu \rangle$?

Unbalanced Optimal Transport – References

[1] S. Kondratyev, L. Monsaingeon, D. Vorotnikov, A new optimal transport distance on the space of finite radon measures, 2015.

[2] M. Liero, A. Mielke, G. Savaré, Optimal entropy-transport problems and the Hellinger-Kantorovich distance, 2015

[3] L. Chizat, G. Peyré, B. Schmitzer, F.-X. Vialard, An interpolating distance between optimal transport and Fisher-Rao metrics, 2015.

[4] L. Chizat, G. Peyré, B. Schmitzer, F.-X. Vialard, Unbalanced optimal transport: geometry and Kantorovich formulation, 2015.

[5] L. Chizat, G. Peyré, B. Schmitzer, F.-X. Vialard, Scaling algorithms for unbalanced transport problems, 2016.

[6] A. Baradat, H. Lavenant, Regularized unbalanced optimal transport as entropy minimization with respect to branching Brownian motion, 2021

Generalized ('Conic') Coulings

Let $C\mathbb{R}^d := \mathbb{R}^d \times \mathbb{R}_+$ and let $X : C\mathbb{R}^d \mapsto \mathcal{M}_+(\mathbb{R}^d)$,

$$X((x,m)) = m\,\delta_x.$$

and let $X : C\mathbb{R}^d \times C\mathbb{R}^d \mapsto \mathcal{M}_+(\mathbb{R}^d) \times \mathcal{M}_+(\mathbb{R}^d)$,

$$X[(x,m),(y,n)] \longrightarrow (m\,\delta_x, n\,\delta_y) =: (X_1(z), X_2(z)),$$

with $z = [(x,m)),(y,n)] \in C\mathbb{R}^d \times C\mathbb{R}^d$.

For $\mu, \nu \in \mathcal{M}_+(\mathbb{R}^d)$ define

$$\Gamma_{\mu,\nu} := \{\gamma \in \mathcal{P}(C\mathbb{R}^d \times C\mathbb{R}^d) \big| \mathbb{E}_\gamma[X_1(Z)] = \mu, \mathbb{E}_\gamma[X_2(Z)] = \nu\}$$

Remark

$$\gamma \in \Gamma_{\mu,\nu} \Leftrightarrow \begin{cases} \int_{C\mathbb{R}^d} \int_{C\mathbb{R}^d} m f(x)\, \gamma(d(x,m), d(y,n)) = \langle f, \mu \rangle \\ \int_{C\mathbb{R}^d} \int_{C\mathbb{R}^d} n f(y)\, \gamma(d(x,m), d(y,n)) = \langle f, \nu \rangle \end{cases}$$

for all bounded measurable $f : \mathbb{R}^d \mapsto \mathbb{R}$*,* $\gamma \in \mathcal{P}(C\mathbb{R}^d \times C\mathbb{R}^d)$*.*

.

Definition (Unbalanced Conic Optimal Transport†)

Let d be a metric on $C\mathbb{R}^d$.

$$\tilde{T}_2(\mu, \nu) = \inf \left\{ \int_{C\mathbb{R}^d} \int_{C\mathbb{R}^d} d^2(\overline{x}, \overline{y}) \gamma(d\overline{x}, d\overline{y}) \mid \gamma \in \Gamma(\mu, \nu) \right\} \quad \text{(CUBOT)}$$

† cf.[1]–[3]

The Reference Measure on Path Space

Let g a Riemannian metric on $C\mathbb{R}^d$, associated gradient ∇^g and volume measure vol_g, intrinsic distance d.

For smooth $h : C\mathbb{R}^d \mapsto \mathbb{R}_{\geq 0}$ let $\rho(d\overline{z}) = e^{-h(\overline{z})}\mathrm{vol}_g(d\overline{z})$.

For $\overline{z} \in C\mathbb{R}^d$ let $l_{\overline{z}} \in \mathcal{P}(\Omega)$ denote the law on $\Omega = C_{\mathbb{R}_{\geq 0}}(C\mathbb{R}^d)$ of the diffusion process on $C\mathbb{R}^d$ generated by the Dirichlet form

$$\mathcal{E}(f,f) = \frac{1}{2}\int_{C\mathbb{R}^d} \|\nabla^g f\|^2 \, d\rho \text{ on } L^2(C\mathbb{R}^d, \rho).$$

starting in $\overline{z}$ and set

$$\mathcal{W} := \int_{C\mathbb{R}^d} l_{\overline{z}} \, \rho(d\overline{z}) \in \mathcal{P}(C_{\mathbb{R}_{\geq 0}}(C\mathbb{R}^d))$$

Definition (Conic Unbalanced Schrödinger Problem)

Let $\mu, \nu \in \mathcal{M}_+(\mathbb{R}^d)$

$$\min \left\{ \mathsf{Ent}(\Lambda, \mathcal{W}) \,\middle|\, \Lambda \in \mathcal{P}(C_{\mathbb{R}_{\geq 0}}(C\mathbb{R}^d)), (\pi_0, \pi_\epsilon)_\# \Lambda \in \Gamma(\mu, \nu) \right\}, \quad (\mathsf{CS}_\epsilon)$$

where $(\pi_0, \pi_\epsilon) : C_{\mathbb{R}_{\geq 0}}(C\mathbb{R}^d)) \mapsto C\mathbb{R}^d \times C\mathbb{R}^d$ is the joint evaluation map at $t = 0$ and $t = \epsilon$, respectively.

Föllmer's Lemma[‡]

For two measures μ, ν on some measurable space $(\Omega, \mathcal{F})$ and a measurable map $T : \Omega \mapsto (S, \Sigma)$ it holds that

$$\mathsf{Ent}(\mu, \nu) = \mathsf{Ent}(T_{\#}\mu, T_{\#}\nu) + \int_{\mathcal{S}} \mathsf{Ent}(\mu_s^T, \nu_s^T) T_{\#}\mu(ds),$$

where μ_s^T, μ_s^T denotes the conditional distribution of μ and μ in Ω given $\{T = s\}$.

[‡]cf. [Leonard, '13]

Föllmer Lemma on Unbalanced Schrödinger

For $T = (\pi_0, \pi_\epsilon)$ in (CS$_\epsilon$)

$$\mathsf{Ent}(\Lambda, \mathcal{W}) = \mathsf{Ent}(\gamma, \sigma) + \int_{C\mathbb{R}^d \times C\mathbb{R}^d} \mathsf{Ent}(\Lambda^T_{\overline{x},\overline{y}}, \mathcal{W}^T_{\overline{x},\overline{y}})\gamma(d(\overline{x}, \overline{y}))$$

with $\gamma := T_\# \Lambda$ and $\sigma := T_\# \mathcal{W}$.

For fixed γ is minimal when $\Lambda^T_{\overline{x},\overline{y}} = l^\epsilon_{\overline{x},\overline{y}} = \mathcal{W}^T_{\overline{x},\overline{y}}$ is the law of the bridge process associated to the Dirichlet form $\mathcal{E}$ starting in $\overline{x}$ and conditioned to be in $\overline{y}$ at $t = \epsilon$.

Hence, (CS$_\epsilon$) takes the form

$$\min \{ \mathsf{Ent}(\gamma, \sigma) \mid \gamma \in \Gamma(\mu, \nu) \} . \qquad (1)$$

Asymptotics for $\epsilon \to 0$: Unbalanced Transport

By construction

$$\sigma(d\overline{x}d\overline{y}) = p_\epsilon(\overline{x},\overline{y})\rho(dx)\rho(dy) = p_\epsilon(\overline{x},\overline{y})e^{-h(\overline{x})}e^{-h(\overline{y})}\mathrm{vol}_g(d\overline{x})\mathrm{vol}_g(d\overline{y}),$$

where p_ϵ denote the heat kernel on $C\mathbb{R}^d$ associated to $\mathcal{E}$. Hence

$$\begin{aligned}\mathsf{Ent}(\gamma,\sigma) &= \int_{C\mathbb{R}^d}\int_{C\mathbb{R}^d}\log\frac{\gamma(\overline{x},\overline{y})}{p_\epsilon(\overline{x},\overline{y})}\gamma(\overline{x},\overline{y})\rho(d\overline{x})\rho(d\overline{y}) \\ &= \mathsf{Ent}(\gamma,\rho\otimes\rho) - \int_{C\mathbb{R}^d}\int_{C\mathbb{R}^d}\log(p_\epsilon(\overline{x},\overline{y}))\gamma(\overline{x},\overline{y})\rho(d\overline{x})\rho(d\overline{y})\end{aligned}$$

'Varadhan formula'

$$\epsilon\ln p_\epsilon(\overline{x},\overline{y}) \longrightarrow -\frac{1}{2}d^2(\overline{x},\overline{y}) \text{ for } \epsilon\to 0$$

where d is the intrinsic metric on $C\mathbb{R}^d$ of $\mathcal{E}$. Hence,

$$\lim_{\epsilon\longrightarrow 0}\epsilon\,\mathsf{Ent}(\gamma,\sigma_\epsilon) = \frac{1}{2}\int_{C\mathbb{R}^d}\int_{C\mathbb{R}^d}d^2(\overline{x},\overline{y})\gamma(d\overline{x},d\overline{y}).$$

Recap

Conic formulation of unbalanced optimal transport

$$\inf\left\{\int_{C\mathbb{R}^d}\int_{C\mathbb{R}^d} d^2(\overline{x},\overline{y})\gamma(d\overline{x},d\overline{y}) \mid \gamma\in\Gamma(\mu,\nu)\right\}. \qquad \text{(CUBOT)}$$

Entropic regularization

$$\inf\left\{\frac{1}{2}\int_{C\mathbb{R}^d}\int_{C\mathbb{R}^d} d^2(\overline{x},\overline{y})\gamma(d\overline{x},d\overline{y}) + \epsilon\,\mathsf{Ent}(\gamma,\rho\otimes\rho) \mid \gamma\in\Gamma(\mu,\nu)\right\}. \qquad \text{(R-CBUOT)}$$

Entropic Regularization of Linear Programs

Standard LP: Given $A \in \mathbb{R}^{d \times n}$, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^d$

$$\text{Minimize } \ c \cdot x \ \text{ subject to } \ Ax = b \text{ and } x \geq 0. \qquad \text{(LP)}$$

(LP) is feasible iff $b \in \operatorname{pos}(A)$

Interior Point Method: Choose a *barrier function*[§] H on $\mathbb{R}_{\geq 0}$,

$$\text{Minimize } \ c \cdot x + \epsilon \sum_{i=1}^{n} H(x_i) \ \text{ subject to } \ Ax = b \text{ and } x \geq 0. \qquad (\text{LP}_\epsilon)$$

Entropic Regularization: Choose $H(t) = t \log t - t$

[§] H convex, $\lim_{t \to 0} H(t) \overset{!}{=} \infty$

LP for OT

Given probability distributions $\mu \in \mathbb{R}^{d_1}_{\geq 0}$ and $\nu \in \mathbb{R}^{d_2}_{\geq 0}$ on the finite sets $[d_1] = \{1, \ldots, d_1\}$ and $[d_2] = \{1, \ldots, d_2\}$, and a cost matrix $c = (c_{\kappa,\lambda})_{\kappa \in [d_1], \lambda \in [d_2]} \in \mathbb{R}^{d_1 \times d_2}$, we aim to

$$\text{Minimize} \sum_{(\kappa,\lambda) \in [d_1] \times [d_2]} c_{\kappa,\lambda} \cdot x_{\kappa,\lambda} \quad \text{subject to} \quad x \geq 0 \quad \text{and}$$

$$\sum_{\lambda \in [d_2]} x_{\kappa,\lambda} = \mu_\kappa \text{ for all } \kappa \in [d_1] \quad \text{and} \quad \sum_{\kappa \in [d_1]} x_{\kappa,\lambda} = \nu_\lambda \text{ for all } \lambda \in [d_2].$$

The matrix $A \in \{0,1\}^{d \times n}$ with $d = d_1 + d_2 - 1$, $n = d_1 d_2$. This LP is feasible (choose $x_{k,\lambda} = \mu_k \nu_\lambda$).

LP for Unbalanced Conic OT

Assume wlog the entries of μ and ν are integers.

Let $e_1, e_2 \in \mathbb{N}$ such that

$$\mu_\kappa \in [e_1] \text{ for all } \kappa \in [d_1]$$

and

$$\nu_\lambda \in [e_2] \text{ for all } \lambda \in [d_2]$$

A probability vector $x = (x_{\kappa,i,\lambda,j})$ on $([d_1]\times[e_1]) \times ([d_2]\times[e_2])$ is a *conic coupling* for μ and ν if

$$\begin{aligned} \sum_{\lambda=1}^{d_2}\sum_{i=1}^{e_1}\sum_{j=1}^{e_2} i\, x_{\kappa,i,\lambda,j} &= \mu_\kappa \quad \text{for } \kappa \in [d_1] \\ \sum_{\kappa=1}^{d_1}\sum_{i=1}^{e_1}\sum_{j=1}^{e_2} j\, x_{\kappa,i,\lambda,j} &= \nu_\lambda \quad \text{for } \lambda \in [d_2]. \end{aligned} \tag{2}$$

LP for Unbalanced Conic OT (cont.'d)

Normalization constraint

$$\sum_{\substack{(\kappa,i,\lambda,j)\ \in \\ [d_1]\times[e_1]\times[d_2]\times[e_2]}} x_{\kappa,i,\lambda,j} \quad = \quad 1. \tag{3}$$

Cost function $c : ([d_1]\times[e_1]) \times ([d_2]\times[e_2]) \to \mathbb{R}$.

$\rightsquigarrow$ LP for UOT

Minimize $\displaystyle\sum_{\substack{(\kappa,i,\lambda,j)\ \in \\ [d_1]\times[e_1]\times[d_2]\times[e_2]}} c_{\kappa,i,\lambda,j} \cdot x_{\kappa,i,\lambda,j}$ subject to $x \geq 0,\ (2),$ and (3).

Choosing $x = (x_{\kappa,i,\lambda,j}) = (\overline{\mu}_\kappa \cdot \delta_{\|\mu\|_1,i} \cdot \overline{\nu}_\lambda \cdot \delta_{\|\nu\|_1,j})$ shows feasibility.

Darroch-Ratcliff Algorithm: Preparations

Assume w.l.o.g. $(1, \ldots, 1) \in L_A$.

Let $a = \max\{\|A^i\|_1, i = 1, \cdots, n\}$ and let

$$\mathcal{A} \;=\; \begin{bmatrix} 0 & & A & \\ a & a_{d+1,1} & \cdots & a_{d+1,n} \end{bmatrix} \;\in\; \mathbb{N}^{(d+1)\times(n+1)}.$$

$$s_c := 1 + \sum_{i=1}^{n} \exp(-c_i/\epsilon)$$

$$\beta \;=\; \left(\frac{b}{2}\,,\; a - \frac{\|b\|_1}{2} \right)^{\top}$$

$$\gamma \;=\; \big(\epsilon \log(s_c),\, c_1 + \epsilon \log(s_c),\, \ldots,\, c_n + \epsilon \log(s_c)\big).$$

$\rightsquigarrow$ Modification of (LP$_\epsilon$)

$$\text{Minimize } \gamma \cdot y \,+\, \epsilon \sum_{i=0}^{n} H(y_i) \;\text{ subject to }\; \mathcal{A}\, y = \beta \text{ and } y \geq 0. \qquad (4)$$

Remark: Column sums of $\mathcal{A}$ equal a.

DR Iterative Scaling

Theorem (DR '72[¶])

If (LP$_\epsilon$) *is feasible, then the solution* $x^*(\epsilon)$ *is given by* $(y_1/y_0, \ldots, y_n/y_0)$, *where* $y = y^*(\epsilon) \in \mathbb{R}^{n+1}_{\geq 0}$ *is the unique solution to* (4). *It satisfies* $\sum_{i=0}^n y_i = 1$ *and is obtained as the unique limit point of the iteration*

$$y^{(0)} = \exp(-\gamma/\epsilon),$$

$$y_i^{(k+1)} = y_i^{(k)} \left(\frac{\beta^{a_i}}{(\mathcal{A}\, y^{(k)})^{a_i}} \right)^{\frac{1}{a}}.$$

[¶] J. N. Darroch, D. Ratcliff: Generalized Iterative Scaling for Log-Linear Models, Ann. Math. Statist. 43(5): 1470-1480

Conclusion

Summary.

1. Relaxation of OT to CUBOT by 'Conic Couplings'
2. Unbalanced Conic Schrödinger Problem

   and pointwise convergence to CUBOT
3. (Gen. Sinkhorn) DR-Iterative Scaling for Entropic Regularization of Conic UBOT

Remarks/Questions.

i DR ok for other constrained OT problems: Multi-Marginal OT, Martingale OT etc. $\surd$

ii Linear Convergence estimates for DR?

iii Γ-convergence of R-CUBOT to CUBOT?

iv Benamou-Brenier, Otto-Calculus etc. for CUBOT?

v ...

vi Thank you!