Numerical Optimal Transport

http://optimaltransport.github.io

Semi-discrete Methods

Gabriel Peyré

www.numerical-tours.com

ÉCOLE NORMALE SUPÉRIEURE

Overview

- **Semi-dual Problem**
- Entropic Semi-dual
- Stochastic Optimization

C-Transform

$$\mathrm{L}_{\mathbf{C}}(\mathbf{a},\mathbf{b}) = \max_{(\mathbf{f},\mathbf{g})\in\mathbf{R}(\mathbf{C})} \langle \mathbf{f},\, \mathbf{a}\rangle + \langle \mathbf{g},\, \mathbf{b}\rangle$$

$$\mathbf{R}(\mathbf{C}) \overset{\text{def.}}{=} \{(\mathbf{f},\mathbf{g}) \in \mathbb{R}^n \times \mathbb{R}^m \;:\; \forall\,(i,j) \in [\![n]\!] \times [\![m]\!], \mathbf{f} \oplus \mathbf{g} \leq \mathbf{C}\}$$

$$\mathcal{L}_c(\alpha,\beta) = \sup_{(f,g)\in\mathcal{R}(c)} \int_{\mathcal{X}} f(x)\mathrm{d}\alpha(x) + \int_{\mathcal{Y}} g(y)\mathrm{d}\beta(y),$$

$$\mathcal{R}(c) \overset{\text{def.}}{=} \{(f,g) \in \mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{Y}) \;:\; \forall(x,y), f(x) + g(y) \leq c(x,y)\}$$

$$\mathrm{Supp}(\pi) \subset \{(x,y) \in \mathcal{X} \times \mathcal{Y} \;:\; f(x) + g(y) = c(x,y)\}$$

C-Transform

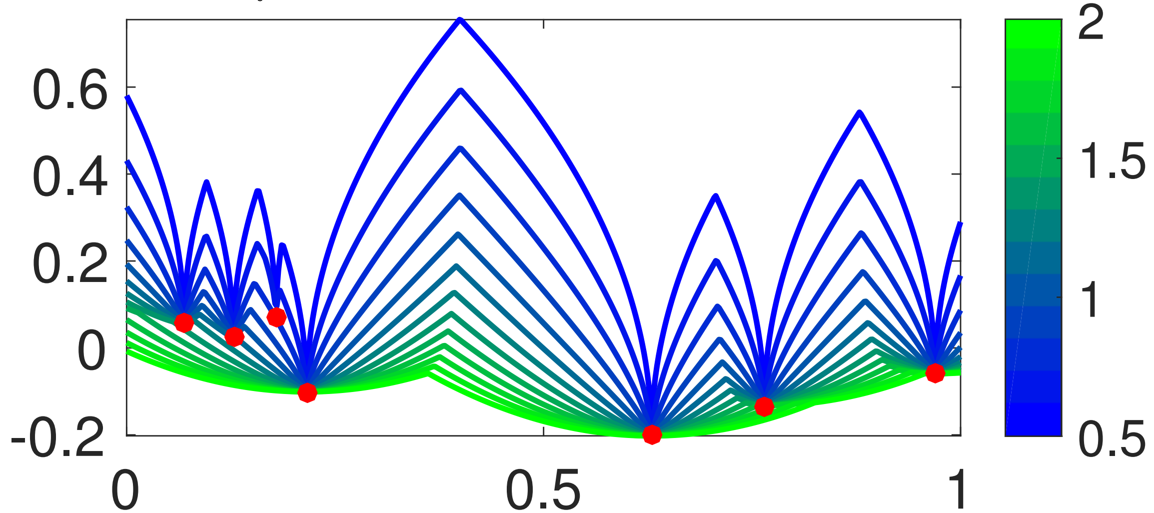
$$\forall\, y \in \mathcal{Y}, \quad f^c(y) \overset{\text{def.}}{=} \inf_{x\in\mathcal{X}} c(x,y) - f(x),$$

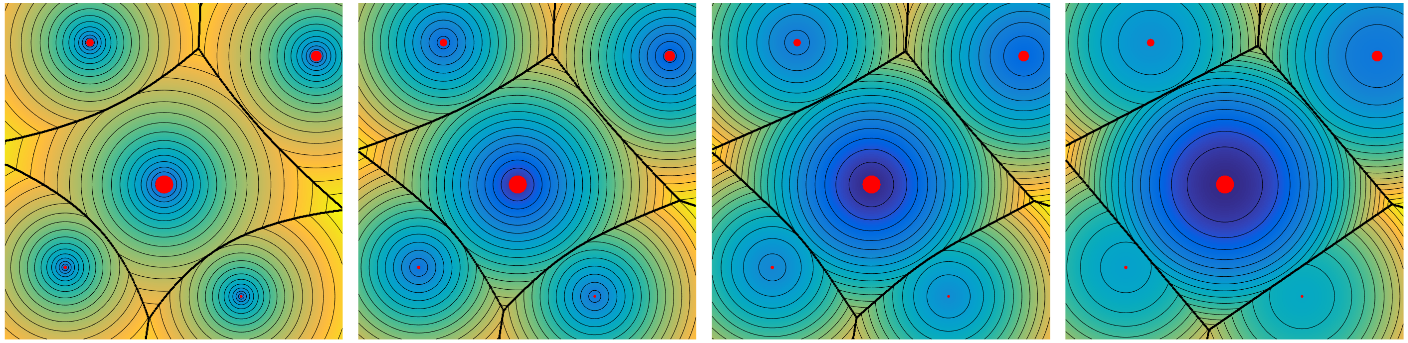
$$\forall\, x \in \mathcal{X}, \quad g^{\bar{c}}(x) \overset{\text{def.}}{=} \inf_{y\in\mathcal{Y}} c(x,y) - g(y),$$

$$\sup_{(f,g)} \mathcal{E}(f,g) \overset{\text{def.}}{=} \int_{\mathcal{X}} f(x)\mathrm{d}\alpha(x) + \int_{\mathcal{Y}} g(y)\mathrm{d}\beta(y) + \iota_{\mathcal{R}(c)}(f,g)$$

$$f^c \in \underset{g}{\operatorname{argmax}}\ \mathcal{E}(f,g) \quad \text{and} \quad g^{\bar{c}} \in \underset{f}{\operatorname{argmax}}\ \mathcal{E}(f,g)$$

$$f^{c\bar{c}c} = f^c \quad \text{and} \quad g^{\bar{c}c\bar{c}} = g^{\bar{c}}$$

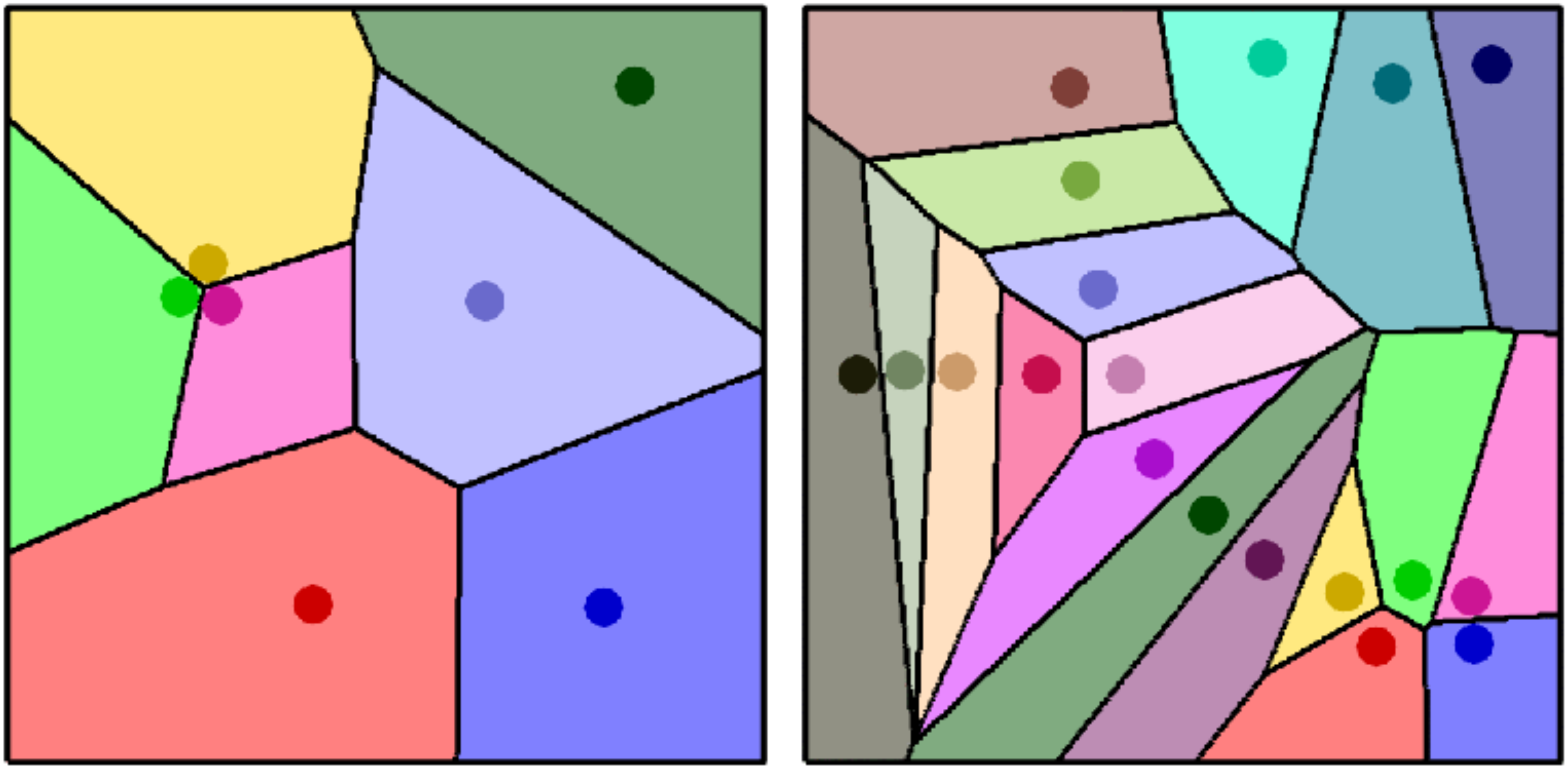




$p = 1/2$ $\quad$ $p = 1$ $\quad$ $p = 3/2$ $\quad$ $p = 2$

Laguerre Cells

Laguerre cells $$\mathbb{L}_{\mathbf{g}}(y_j) \stackrel{\text{def.}}{=} \left\{ x \in \mathcal{X} \;:\; \forall\, j' \neq j, c(x, y_j) - \mathbf{g}_j \leq c(x, y_{j'}) - \mathbf{g}_{j'} \right\}$$



Semi-dual Problem

$$\begin{aligned}\mathcal{L}_c(\alpha,\beta) &= \max_{f\in\mathcal{C}(\mathcal{X})} \int_{\mathcal{X}} f(x)\mathrm{d}\alpha(x) + \int_{\mathcal{Y}} f^c(y)\mathrm{d}\beta(y),\\ &= \max_{g\in\mathcal{C}(\mathcal{Y})} \int_{\mathcal{X}} g^{\bar{c}}(x)\mathrm{d}\alpha(x) + \int_{\mathcal{Y}} g(y)\mathrm{d}\beta(y).\end{aligned}$$

$$\forall\, \mathbf{g}\in\mathbb{R}^m,\ \forall\, x\in\mathcal{X},\quad \mathbf{g}^{\bar{c}}(x) \overset{\text{def.}}{=} \min_{j\in[\![m]\!]} c(x,y_j) - \mathbf{g}_j$$

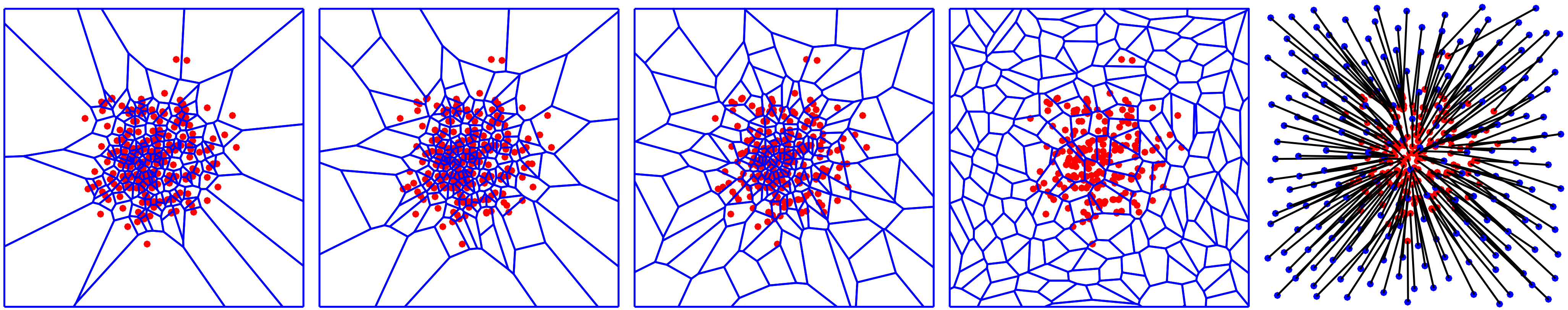
$$\mathcal{L}_c(\alpha,\beta) = \max_{\mathbf{g}\in\mathbb{R}^m} \mathcal{E}(\mathbf{g}) \overset{\text{def.}}{=} \int_{\mathcal{X}} \mathbf{g}^{\bar{c}}(x)\mathrm{d}\alpha(x) + \sum \mathbf{g}_y \mathbf{b}_j$$

Laguerre cells $\quad \mathbb{L}_{\mathbf{g}}(y_j) \overset{\text{def.}}{=} \left\{ x\in\mathcal{X} \ :\ \forall\, j'\neq j, c(x,y_j) - \mathbf{g}_j \leq c(x,y_{j'}) - \mathbf{g}_{j'} \right\}$

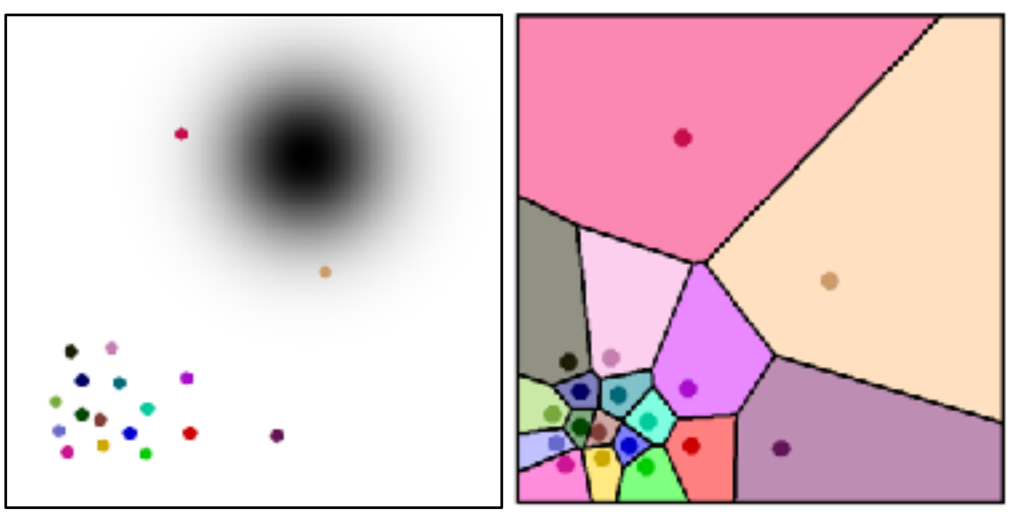
$$\mathcal{E}(\mathbf{g}) = \sum_{j=1}^m \int_{\mathbb{L}_{\mathbf{g}}(y_j)} \left(c(x,y_j) - \mathbf{g}_j \right) \mathrm{d}\alpha(x) + \langle \mathbf{g},\, \mathbf{b} \rangle$$

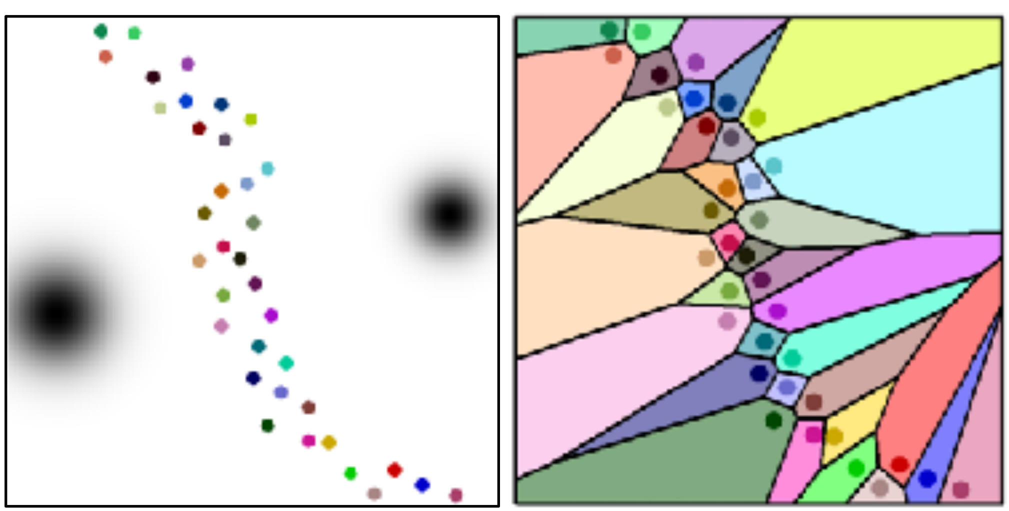
$$\forall\, j\in[\![m]\!],\quad \nabla\mathcal{E}(\mathbf{g})_j = -\int_{\mathbb{L}_{\mathbf{g}}(y_j)} \mathrm{d}\alpha(x) + \mathbf{b}_j$$

Evolution of the Semi-Discrete Optimization



$\ell = 1$ $\ell = 3$ $\ell = 50$ $\ell = 100$ Matching





Overview

- Semi-dual Problem
- **Entropic Semi-dual**
- Stochastic Optimization

Entropic Regularized Dual

$$\mathrm{L}_{\mathbf{C}}^{\varepsilon}(\mathbf{a},\mathbf{b}) \overset{\text{def.}}{=} \min_{\mathbf{P}\in\mathbf{U}(\mathbf{a},\mathbf{b})} \langle \mathbf{P},\ \mathbf{C}\rangle - \varepsilon \mathbf{H}(\mathbf{P})$$

$$\mathrm{L}_{\mathbf{C}}^{\varepsilon}(\mathbf{a},\mathbf{b}) = \max_{\mathbf{f}\in\mathbb{R}^n,\mathbf{g}\in\mathbb{R}^m} \langle \mathbf{f},\ \mathbf{a}\rangle + \langle \mathbf{g},\ \mathbf{b}\rangle - \varepsilon\langle e^{\mathbf{f}/\varepsilon},\ \mathbf{K}e^{\mathbf{g}/\varepsilon}\rangle$$

$$(\mathbf{u},\mathbf{v}) = (e^{\mathbf{f}/\varepsilon}, e^{\mathbf{g}/\varepsilon})$$

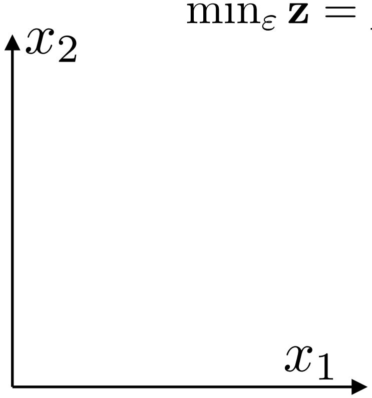
$$\mathbf{P}_{i,j} = e^{\mathbf{f}_i/\varepsilon} e^{-\mathbf{C}_{i,j}/\varepsilon} e^{\mathbf{g}_j/\varepsilon}$$

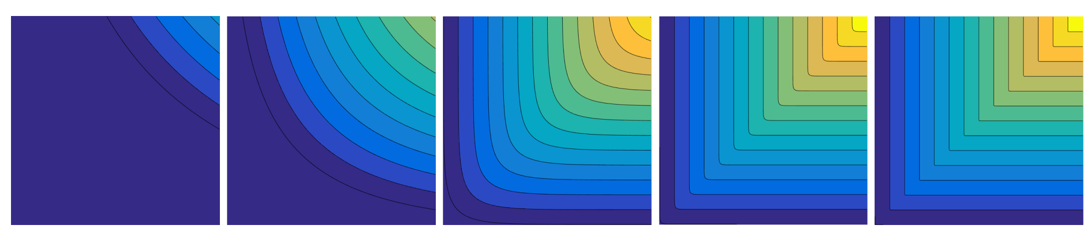
$$\min\nolimits_{\varepsilon} \mathbf{z} = -\varepsilon \log \sum_i e^{-\mathbf{z}_i/\varepsilon}$$

$$\min\nolimits_{\varepsilon} \mathbf{z} = \underline{z} - \varepsilon \log \sum_i e^{-(\mathbf{z}_i - \underline{z})/\varepsilon}.$$

$$(\mathbf{f}^{(\ell+1)})_i = \min\nolimits_{\varepsilon} (\mathbf{C}_{ij} - \mathbf{g}_j^{(\ell)})_j + \varepsilon \log \mathbf{a}_i$$

$$(\mathbf{g}^{(\ell+1)})_j = \min\nolimits_{\varepsilon} (\mathbf{C}_{ij} - \mathbf{f}_i^{(\ell)})_i + \varepsilon \log \mathbf{b}_j.$$





$\varepsilon = 1$ $\varepsilon = 0.5$ $\varepsilon = 10^{-1}$ $\varepsilon = 10^{-2}$ $\varepsilon = 10^{-3}$

General Measures

$$\mathcal{L}_c^\varepsilon(\alpha,\beta) \overset{\text{def.}}{=} \sup_{(f,g)\in\mathcal{C}(\mathcal{X})\times\mathcal{C}(\mathcal{Y})} \int_\mathcal{X} f\mathrm{d}\alpha + \int_\mathcal{Y} g\mathrm{d}\beta - \varepsilon\int_{\mathcal{X}\times\mathcal{Y}} e^{\frac{-c+f\oplus g}{\varepsilon}}\mathrm{d}\alpha\mathrm{d}\beta$$

$$\forall\, y\in\mathcal{Y},\quad f^{c,\varepsilon}(y) \overset{\text{def.}}{=} -\varepsilon\log\left(\int_\mathcal{X} e^{\frac{-c(x,y)+f(x)}{\varepsilon}}\mathrm{d}\alpha(x)\right)$$

$$\forall\, x\in\mathcal{X},\quad g^{\bar{c},\varepsilon}(x) \overset{\text{def.}}{=} -\varepsilon\log\left(\int_\mathcal{Y} e^{\frac{-c(x,y)+g(y)}{\varepsilon}}\mathrm{d}\beta(y)\right)$$

$$\forall\, x\in\mathcal{X},\quad \mathbf{g}^{\bar{c},\varepsilon}(x) \overset{\text{def.}}{=} -\varepsilon\log\left(\sum_{j=1}^m e^{\frac{-c(x,y_j)+\mathbf{g}_j}{\varepsilon}}\mathbf{b}_j\right)$$

Sinkhorn $\quad \mathbf{f}_i^{(\ell+1)} = \mathbf{g}^{\bar{c},\varepsilon}(x_i) \quad \text{and} \quad \mathbf{g}_j^{(\ell+1)} = \mathbf{f}^{c,\varepsilon}(y_j)$

Entropic Semi-Dual

$$\max_{\mathbf{g}\in\mathbb{R}^n} \mathcal{E}^\varepsilon(\mathbf{g}) \stackrel{\text{def.}}{=} \int_{\mathcal{X}} \mathbf{g}^{\bar{c},\varepsilon}(x)\mathrm{d}\alpha(x) + \langle \mathbf{g},\, \mathbf{b}\rangle \qquad \forall\, j \in [\![m]\!], \quad \nabla\mathcal{E}^\varepsilon(\mathbf{g})_j = -\int_{\mathcal{X}} \chi_j^\varepsilon(x)\mathrm{d}\alpha(x) + \mathbf{b}_j$$

$$\chi_j^\varepsilon(x) = \frac{e^{\frac{-c(x,y_j)+\mathbf{g}_j}{\varepsilon}}}{\sum_\ell e^{\frac{-c(x,y_\ell)+\mathbf{g}_\ell}{\varepsilon}}}$$

$$\mathcal{E}^\varepsilon(\mathbf{g}) = \int_{\mathcal{X}} E^\varepsilon(\mathbf{g},x)\mathrm{d}\alpha(x) = \mathbb{E}_X(E^\varepsilon(\mathbf{g},X))$$

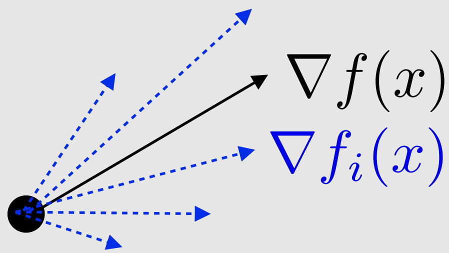
$$E^\varepsilon(\mathbf{g},x) \stackrel{\text{def.}}{=} \mathbf{g}^{\bar{c},\varepsilon}(x) - \langle \mathbf{g},\, \mathbf{b}\rangle$$

Overview

- Semi-dual Problem
- Entropic Semi-dual
- **Stochastic Optimization**

Stochastic Gradient Descent

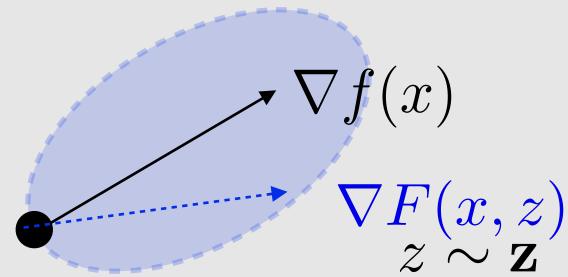
$$f(x) \overset{\text{def.}}{=} \frac{1}{n}\sum_{i=1}^n f_i(x)$$
$$\nabla f(x) = \frac{1}{n}\sum_i \nabla f_i(x)$$



Draw $i \in \{1, \ldots, n\}$ uniformly.

$$x_{k+1} = x_k - \tau_k \nabla f_i(x_k)$$

$$f(x) \overset{\text{def.}}{=} \mathbb{E}_{\mathbf{z}}(f(x, \mathbf{z}))$$
$$\nabla f(x) \overset{\text{def.}}{=} \mathbb{E}_{\mathbf{z}}(\nabla F(x, \mathbf{z}))$$



Draw $z \sim \mathbf{z}$

$$x_{k+1} = x_k - \tau_k \nabla F(x, z)$$

Theorem: If $\mu > 0$ and $\|\nabla f_i(x)\| \leqslant C$, then for $\tau_k = \frac{1}{\mu(k+1)}$,

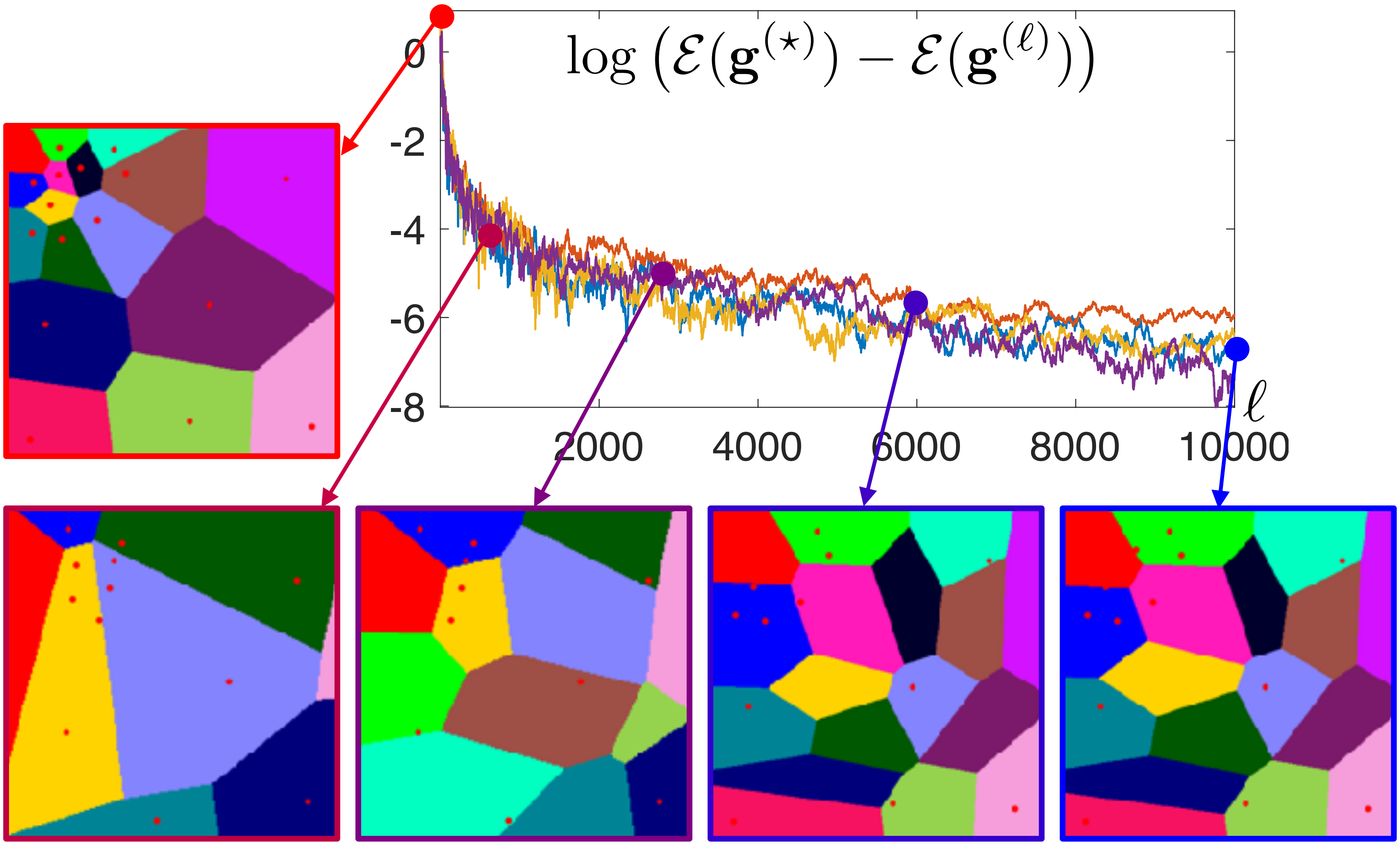
$$\mathbb{E}(\|x_k - x^\star\|^2) \leqslant \frac{R}{k+1} \qquad \text{where} \quad R \overset{\text{def.}}{=} \max(\|x_0 - x^\star\|^2, C^2/\mu^2)$$

$\tau_k \to 0$ to cancel gradient noise.

No benefit from strong convexity.

$\longrightarrow$ Only useful when n is *very* large.

Semi-discrete Stochastic Descent



Stochastic gradient descent for the semi-discrete Optimal Transport,
illustration of convergence and corresponding Laguerre cells.
https://arxiv.org/abs/1605.08527