Numerical Optimal Transport

http://optimaltransport.github.io

Dual and Semi-discrete

Gabriel Peyré

www.numerical-tours.com

Overview

- **Dual Problem**
- W_1
- Semi-discrete Problem
- Optimal Quantization

Discrete Dual Problem

$$\mathrm{W}_c(\mathbf{a}, \mathbf{b}) = \min_{\mathbf{P} \in \mathbb{R}_+^{n \times m}} \left\{ \langle \mathbf{C}, \mathbf{P} \rangle \ : \ \mathbf{P}\mathbb{1} = \mathbf{a}, \mathbf{P}^\top \mathbb{1} = \mathbf{b} \right\}$$

Discrete Dual Problem

$$
\begin{aligned}
\mathrm{W}_c(\mathbf{a}, \mathbf{b}) &= \min_{\mathbf{P} \in \mathbb{R}_+^{n \times m}} \left\{ \langle \mathbf{C}, \mathbf{P} \rangle \ : \ \mathbf{P}\mathbb{1} = \mathbf{a}, \mathbf{P}^\top \mathbb{1} = \mathbf{b} \right\} \\
&= \min_{\mathbf{P} \in \mathbb{R}_+^{n \times m}} \max_{\mathbf{f} \in \mathbb{R}^n, \mathbf{g} \in \mathbb{R}^m} \langle \mathbf{C}, \mathbf{P} \rangle + \langle \mathbf{f}, \mathbf{a} - \mathbf{P}\mathbb{1} \rangle + \langle \mathbf{g}, \mathbf{b} - \mathbf{P}^\top \mathbb{1} \rangle
\end{aligned}
$$

Discrete Dual Problem

$$
\begin{aligned}
\mathrm{W}_c(\mathbf{a}, \mathbf{b}) &= \min_{\mathbf{P} \in \mathbb{R}_+^{n \times m}} \left\{ \langle \mathbf{C}, \mathbf{P} \rangle \;:\; \mathbf{P}\mathbb{1} = \mathbf{a}, \mathbf{P}^\top \mathbb{1} = \mathbf{b} \right\} \\
&= \min_{\mathbf{P} \in \mathbb{R}_+^{n \times m}} \max_{\mathbf{f} \in \mathbb{R}^n, \mathbf{g} \in \mathbb{R}^m} \langle \mathbf{C}, \mathbf{P} \rangle + \langle \mathbf{f}, \mathbf{a} - \mathbf{P}\mathbb{1} \rangle + \langle \mathbf{g}, \mathbf{b} - \mathbf{P}^\top \mathbb{1} \rangle \\
&= \max_{\mathbf{f} \in \mathbb{R}^n, \mathbf{g} \in \mathbb{R}^m} \langle \mathbf{f}, \mathbf{a} \rangle + \langle \mathbf{g}, \mathbf{b} \rangle + \min_{\mathbf{P} \in \mathbb{R}_+^{n \times m}} \langle \mathbf{C}, \mathbf{P} \rangle - \langle \mathbf{f}, \mathbf{P}\mathbb{1} \rangle - \langle \mathbf{g}, \mathbf{P}^\top \mathbb{1} \rangle
\end{aligned}
$$

Discrete Dual Problem

$$\begin{aligned}
\mathrm{W}_c(\mathbf{a},\mathbf{b}) &= \min_{\mathbf{P}\in\mathbb{R}_+^{n\times m}} \left\{ \langle \mathbf{C},\, \mathbf{P}\rangle \;:\; \mathbf{P}\mathbb{1} = \mathbf{a}, \mathbf{P}^\top \mathbb{1} = \mathbf{b} \right\} \\
&= \min_{\mathbf{P}\in\mathbb{R}_+^{n\times m}} \max_{\mathbf{f}\in\mathbb{R}^n, \mathbf{g}\in\mathbb{R}^m} \langle \mathbf{C},\, \mathbf{P}\rangle + \langle \mathbf{f},\, \mathbf{a} - \mathbf{P}\mathbb{1}\rangle + \langle \mathbf{g},\, \mathbf{b} - \mathbf{P}^\top\mathbb{1}\rangle \\
&= \max_{\mathbf{f}\in\mathbb{R}^n, \mathbf{g}\in\mathbb{R}^m} \langle \mathbf{f},\, \mathbf{a}\rangle + \langle \mathbf{g},\, \mathbf{b}\rangle + \min_{\mathbf{P}\in\mathbb{R}_+^{n\times m}} \langle \mathbf{C},\, \mathbf{P}\rangle - \langle \mathbf{f},\, \mathbf{P}\mathbb{1}\rangle - \langle \mathbf{g},\, \mathbf{P}^\top\mathbb{1}\rangle \\
&= \max_{\mathbf{f}\in\mathbb{R}^n, \mathbf{g}\in\mathbb{R}^m} \langle \mathbf{f},\, \mathbf{a}\rangle + \langle \mathbf{g},\, \mathbf{b}\rangle + \min_{\mathbf{P}\geqslant 0} \langle \mathbf{C} - (\mathbf{f}\mathbb{1}^\top + \mathbb{1}\mathbf{g}^\top),\, \mathbf{P}\rangle
\end{aligned}$$

Discrete Dual Problem

$$\begin{aligned}
W_c(\mathbf{a},\mathbf{b}) &= \min_{\mathbf{P}\in\mathbb{R}_+^{n\times m}} \left\{\langle \mathbf{C},\, \mathbf{P}\rangle \ : \ \mathbf{P}\mathbb{1} = \mathbf{a}, \mathbf{P}^\top\mathbb{1} = \mathbf{b}\right\} \\
&= \min_{\mathbf{P}\in\mathbb{R}_+^{n\times m}} \max_{\mathbf{f}\in\mathbb{R}^n,\mathbf{g}\in\mathbb{R}^m} \langle \mathbf{C},\, \mathbf{P}\rangle + \langle \mathbf{f},\, \mathbf{a} - \mathbf{P}\mathbb{1}\rangle + \langle \mathbf{g},\, \mathbf{b} - \mathbf{P}^\top\mathbb{1}\rangle \\
&= \max_{\mathbf{f}\in\mathbb{R}^n,\mathbf{g}\in\mathbb{R}^m} \langle \mathbf{f},\, \mathbf{a}\rangle + \langle \mathbf{g},\, \mathbf{b}\rangle + \min_{\mathbf{P}\in\mathbb{R}_+^{n\times m}} \langle \mathbf{C},\, \mathbf{P}\rangle - \langle \mathbf{f},\, \mathbf{P}\mathbb{1}\rangle - \langle \mathbf{g},\, \mathbf{P}^\top\mathbb{1}\rangle \\
&= \max_{\mathbf{f}\in\mathbb{R}^n,\mathbf{g}\in\mathbb{R}^m} \langle \mathbf{f},\, \mathbf{a}\rangle + \langle \mathbf{g},\, \mathbf{b}\rangle + \boxed{\min_{\mathbf{P}\geqslant 0} \langle \mathbf{C} - (\mathbf{f}\mathbb{1}^\top + \mathbb{1}\mathbf{g}^\top),\, \mathbf{P}\rangle}
\end{aligned}$$

$$\boxed{\min_{\mathbf{P}\geqslant 0} \langle \mathbf{Q},\, \mathbf{P}\rangle = \begin{cases} 0 & \text{if} \quad \mathbf{Q} \geqslant 0, \\ -\infty & \text{otherwise.} \end{cases}}$$

Theorem:

$$W_c(\mathbf{a},\mathbf{b}) = \max_{\mathbf{f}\in\mathbb{R}^n,\mathbf{g}\in\mathbb{R}^m} \{\langle \mathbf{f},\, \mathbf{a}\rangle + \langle \mathbf{g},\, \mathbf{b}\rangle \ : \boxed{\mathbf{f} \oplus \mathbf{g} \leqslant \mathbf{C}}\}$$

Discrete Dual Problem

$$\mathrm{W}_c(\mathbf{a},\mathbf{b}) = \min_{\mathbf{P}\in\mathbb{R}_+^{n\times m}} \left\{\langle \mathbf{C}, \mathbf{P}\rangle \ : \ \mathbf{P}\mathbb{1} = \mathbf{a}, \mathbf{P}^\top\mathbb{1} = \mathbf{b}\right\}$$

$$= \min_{\mathbf{P}\in\mathbb{R}_+^{n\times m}} \max_{\mathbf{f}\in\mathbb{R}^n, \mathbf{g}\in\mathbb{R}^m} \langle \mathbf{C}, \mathbf{P}\rangle + \langle \mathbf{f}, \mathbf{a} - \mathbf{P}\mathbb{1}\rangle + \langle \mathbf{g}, \mathbf{b} - \mathbf{P}^\top\mathbb{1}\rangle$$

$$= \max_{\mathbf{f}\in\mathbb{R}^n, \mathbf{g}\in\mathbb{R}^m} \langle \mathbf{f}, \mathbf{a}\rangle + \langle \mathbf{g}, \mathbf{b}\rangle + \min_{\mathbf{P}\in\mathbb{R}_+^{n\times m}} \langle \mathbf{C}, \mathbf{P}\rangle - \langle \mathbf{f}, \mathbf{P}\mathbb{1}\rangle - \langle \mathbf{g}, \mathbf{P}^\top\mathbb{1}\rangle$$

$$= \max_{\mathbf{f}\in\mathbb{R}^n, \mathbf{g}\in\mathbb{R}^m} \langle \mathbf{f}, \mathbf{a}\rangle + \langle \mathbf{g}, \mathbf{b}\rangle + \boxed{\min_{\mathbf{P}\geqslant 0} \langle \mathbf{C} - (\mathbf{f}\mathbb{1}^\top + \mathbb{1}\mathbf{g}^\top), \mathbf{P}\rangle}$$

$$\boxed{\min_{\mathbf{P}\geqslant 0} \langle \mathbf{Q}, \mathbf{P}\rangle = \begin{cases} 0 & \text{if} \quad \mathbf{Q} \geqslant 0, \\ -\infty & \text{otherwise.} \end{cases}}$$

Theorem:

$$\mathrm{W}_c(\mathbf{a},\mathbf{b}) = \max_{\mathbf{f}\in\mathbb{R}^n, \mathbf{g}\in\mathbb{R}^m} \{\langle \mathbf{f}, \mathbf{a}\rangle + \langle \mathbf{g}, \mathbf{b}\rangle \ : \boxed{\mathbf{f} \oplus \mathbf{g} \leqslant \mathbf{C}}\}$$

Primal-dual relations: $\{(i,j) \ : \ \mathbf{P}_{i,j} \neq 0\} \subset \{(i,j) \ : \ \mathbf{f}_i + \mathbf{g}_j = \mathbf{C}_{i,j}\}$

Continuous Dual Problem

$$\mathrm{W}_c(\alpha, \beta) = \min_{\pi \in \mathcal{M}_+(\mathcal{X} \times \mathcal{Y})} \{\langle c, \pi \rangle \;:\; \pi_1 = \alpha, \pi_2 = \beta\}$$

$$\langle f, \alpha \rangle \overset{\text{def.}}{=} \int_{\mathcal{X}} f(x) \mathrm{d}\alpha(x) \xrightarrow[\mathbf{f} = (f(x_i))_i]{\alpha = \sum_i \mathbf{a}_i \delta_{x_i}} \langle \mathbf{f}, \mathbf{a} \rangle \overset{\text{def.}}{=} \sum_i \mathbf{f}_i \mathbf{a}_i$$

Continuous Dual Problem

$$\mathrm{W}_c(\alpha,\beta) = \min_{\pi \in \mathcal{M}_+(\mathcal{X}\times\mathcal{Y})} \{\langle c,\, \pi\rangle \;:\; \pi_1 = \alpha, \pi_2 = \beta\}$$

$$\langle f,\, \alpha\rangle \stackrel{\text{def.}}{=} \int_{\mathcal{X}} f(x)\mathrm{d}\alpha(x) \xrightarrow[\mathbf{f} = (f(x_i))_i]{\alpha = \sum_i \mathbf{a}_i \delta_{x_i}} \langle \mathbf{f},\, \mathbf{a}\rangle \stackrel{\text{def.}}{=} \sum_i \mathbf{f}_i \mathbf{a}_i$$

Theorem:

$$\mathrm{W}_c(\alpha,\beta) = \max_{f\in\mathcal{C}(\mathcal{X}),g\in\mathcal{C}(\mathcal{Y})} \{\langle f,\, \alpha\rangle + \langle g,\, \beta\rangle \;:\; f \oplus g \leqslant c\}$$

$$\max_{f\in\mathcal{C}(\mathcal{X}),g\in\mathcal{C}(\mathcal{Y})} \left\{ \int_{\mathcal{X}} f(x)\mathrm{d}\alpha(x) + \int_{\mathcal{Y}} g(y)\mathrm{d}\beta(y) \;:\; \forall (x,y),\; f(x)+g(y) \leqslant c(x,y) \right\}$$

Continuous Dual Problem

$$\mathrm{W}_c(\alpha, \beta) = \min_{\pi \in \mathcal{M}_+(\mathcal{X} \times \mathcal{Y})} \{\langle c, \pi \rangle \ : \ \pi_1 = \alpha, \pi_2 = \beta\}$$

$$\langle f, \alpha \rangle \overset{\text{def.}}{=} \int_{\mathcal{X}} f(x) \mathrm{d}\alpha(x) \xrightarrow[\mathbf{f} = (f(x_i))_i]{\alpha = \sum_i \mathbf{a}_i \delta_{x_i}} \langle \mathbf{f}, \mathbf{a} \rangle \overset{\text{def.}}{=} \sum_i \mathbf{f}_i \mathbf{a}_i$$

Theorem:

$$\mathrm{W}_c(\alpha, \beta) = \max_{f \in \mathcal{C}(\mathcal{X}), g \in \mathcal{C}(\mathcal{Y})} \{\langle f, \alpha \rangle + \langle g, \beta \rangle \ : \ f \oplus g \leqslant c\}$$

$$\max_{f \in \mathcal{C}(\mathcal{X}), g \in \mathcal{C}(\mathcal{Y})} \left\{ \int_{\mathcal{X}} f(x) \mathrm{d}\alpha(x) + \int_{\mathcal{Y}} g(y) \mathrm{d}\beta(y) \ : \ \forall (x, y), \ f(x) + g(y) \leqslant c(x, y) \right\}$$

Primal-dual relations: $\mathrm{Supp}(\pi) \subset \{(x, y) \ : \ f(x) + g(y) = c(x, y)\}$

C-transforms

Fixing f, solve the dual with respect to g:

$$\max_{g\in\mathcal{C}(\mathcal{Y})}\left\{\int_{\mathcal{Y}} g(y)\mathrm{d}\beta(y) \ : \ \forall(x,y),\ f(x)+g(y)\leqslant c(x,y)\right\}$$

$$\max_{g\in\mathcal{C}(\mathcal{Y})}\left\{\int_{\mathcal{Y}} g(y)\mathrm{d}\beta(y) \ : \ \forall y,\ g(y)\leqslant \min_x c(x,y)-f(x)\right\}$$

C-transforms

Fixing f, solve the dual with respect to g:

$$\max_{g \in \mathcal{C}(\mathcal{Y})} \left\{ \int_{\mathcal{Y}} g(y) \mathrm{d}\beta(y) \;:\; \forall (x,y),\; f(x) + g(y) \leqslant c(x,y) \right\}$$

$$\max_{g \in \mathcal{C}(\mathcal{Y})} \left\{ \int_{\mathcal{Y}} g(y) \mathrm{d}\beta(y) \;:\; \forall y,\; g(y) \leqslant \min_{x} c(x,y) - f(x) \right\}$$

c-transforms:

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

$$g^c(x) \overset{\text{def.}}{=} \min_{y \in \mathcal{Y}} c(x,y) - g(y)$$

C-transforms

Fixing f, solve the dual with respect to g:

$$\max_{g \in \mathcal{C}(\mathcal{Y})} \left\{ \int_{\mathcal{Y}} g(y) \mathrm{d}\beta(y) \;:\; \forall (x,y),\; f(x) + g(y) \leqslant c(x,y) \right\}$$

$$\max_{g \in \mathcal{C}(\mathcal{Y})} \left\{ \int_{\mathcal{Y}} g(y) \mathrm{d}\beta(y) \;:\; \forall y,\; g(y) \leqslant \min_{x} c(x,y) - f(x) \right\}$$

c-transforms:

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

$$g^c(x) \overset{\text{def.}}{=} \min_{y \in \mathcal{Y}} c(x,y) - g(y)$$

Proposition: The dual solution on g is equal α-a.e to f^c.

C-transforms

Fixing f, solve the dual with respect to g:

$$\max_{g \in \mathcal{C}(\mathcal{Y})} \left\{ \int_{\mathcal{Y}} g(y) \mathrm{d}\beta(y) \ : \ \forall (x,y),\ f(x) + g(y) \leqslant c(x,y) \right\}$$

$$\max_{g \in \mathcal{C}(\mathcal{Y})} \left\{ \int_{\mathcal{Y}} g(y) \mathrm{d}\beta(y) \ : \ \forall y,\ g(y) \leqslant \min_{x} c(x,y) - f(x) \right\}$$

c-transforms:

$$f^c(y) \stackrel{\text{def.}}{=} \min_{x \in \mathcal{X}} c(x,y) - f(x)$$

$$g^c(x) \stackrel{\text{def.}}{=} \min_{y \in \mathcal{Y}} c(x,y) - g(y)$$

Proposition: The dual solution on g is equal α-a.e to f^c.

Cost $c(x,y) = -\langle x, y \rangle$*:*

$$\int \|x-y\|^2 \mathrm{d}\pi(x,y) = -2 \int \langle x, y \rangle \mathrm{d}\pi(x,y) + \underbrace{\boxed{\int \|x\|^2 \mathrm{d}\alpha(x) + \int \|y\|^2 \mathrm{d}\beta(y)}}_{\text{constant}}$$

$$f^c = -(-f)^* \quad \text{where} \quad h^*(y) \stackrel{\text{def.}}{=} \sup_x \langle x, y \rangle - f(x)$$

Overview

- Dual Problem
- W_1
- Semi-discrete Problem
- Optimal Quantization

Alternate Minimization and W1

$$(f, g) \longmapsto (f, f^c)$$

Alternate Minimization and W1

$$(f, g) \longmapsto (f, f^c) \longmapsto (f^{cc}, f^c) \longmapsto (f^{cc}, f^{ccc})$$

Alternate Minimization and W1

$$(f, g) \longmapsto (f, f^c) \longmapsto \boxed{(f^{cc}, f^c)} \longmapsto (f^{cc}, f^{ccc}) = \boxed{(f^{cc}, f^c)}$$

Proposition: $f^{ccc} = f^c$.

Alternate Minimization and W1

$$(f, g) \longmapsto (f, f^c) \longmapsto \boxed{(f^{cc}, f^c)} \longmapsto (f^{cc}, f^{ccc}) = \boxed{(f^{cc}, f^c)}$$

Proposition: $f^{ccc} = f^c$.

W_1 *case:* $c(x, y) = d(x, y)$

Proposition: $f^{cc} = -f^c$

$\exists f$ s.t. $g = f^c \iff g$ is 1-Lipschitz.

Alternate Minimization and W1

$$(f, g) \longmapsto (f, f^c) \longmapsto \boxed{(f^{cc}, f^c)} \longmapsto (f^{cc}, f^{ccc}) = \boxed{(f^{cc}, f^c)}$$

Proposition: $f^{ccc} = f^c$.

W_1 *case:* $c(x, y) = d(x, y)$

Proposition: $f^{cc} = -f^c$

$\exists f$ s.t. $g = f^c \iff g$ is 1-Lipschitz.

$$\begin{aligned} \mathrm{W}_1(\alpha, \beta) &= \max_{f, g} \{\langle f, \alpha \rangle + \langle g, \beta \rangle \ : \ f \oplus g \leqslant c\} \\ &= \max_{f} \langle f, \alpha \rangle + \langle f^c, \beta \rangle \end{aligned}$$

Alternate Minimization and W1

$$(f, g) \longmapsto (f, f^c) \longmapsto \boxed{(f^{cc}, f^c)} \longmapsto (f^{cc}, f^{ccc}) = \boxed{(f^{cc}, f^c)}$$

Proposition: $f^{ccc} = f^c$.

W_1 *case:* $c(x, y) = d(x, y)$

Proposition: $f^{cc} = -f^c$

$\exists f$ s.t. $g = f^c \iff g$ is 1-Lipschitz.

$$\mathrm{W}_1(\alpha, \beta) = \max_{f,g} \{\langle f, \alpha\rangle + \langle g, \beta\rangle \ : \ f \oplus g \leqslant c\}$$

$$= \max_{f} \langle f, \alpha\rangle + \langle f^c, \beta\rangle \ = \max_{f} \langle f^c, \beta - \alpha\rangle$$

Alternate Minimization and W1

$$(f, g) \longmapsto (f, f^c) \longmapsto \boxed{(f^{cc}, f^c)} \longmapsto (f^{cc}, f^{ccc}) = \boxed{(f^{cc}, f^c)}$$

Proposition: $f^{ccc} = f^c$.

W_1 *case:* $c(x,y) = d(x,y)$

Proposition: $f^{cc} = -f^c$

$\exists f$ s.t. $g = f^c \iff g$ is 1-Lipschitz.

$$\mathrm{W}_1(\alpha, \beta) = \max_{f,g} \{\langle f, \alpha \rangle + \langle g, \beta \rangle \; : \; f \oplus g \leqslant c\}$$

$$= \max_{f} \langle f, \alpha \rangle + \langle f^c, \beta \rangle \; = \max_{f} \langle f^c, \beta - \alpha \rangle$$

$$= \max_{\mathrm{Lip}(f) \leqslant 1} \langle f, \alpha - \beta \rangle = \|\alpha - \beta\|_{\mathrm{W}_1}$$

Euclidean and Graphs W1

Case $d(x,y) = \|x-y\|$*:*

$$\|\alpha - \beta\|_{\mathrm{W}_1} = \max_{\|\nabla f(x)\|_\infty \leqslant 1} \langle f,\, \alpha - \beta \rangle = \min_{\mathrm{div}(u)=\alpha-\beta} \int \|u(x)\| \mathrm{d}x$$

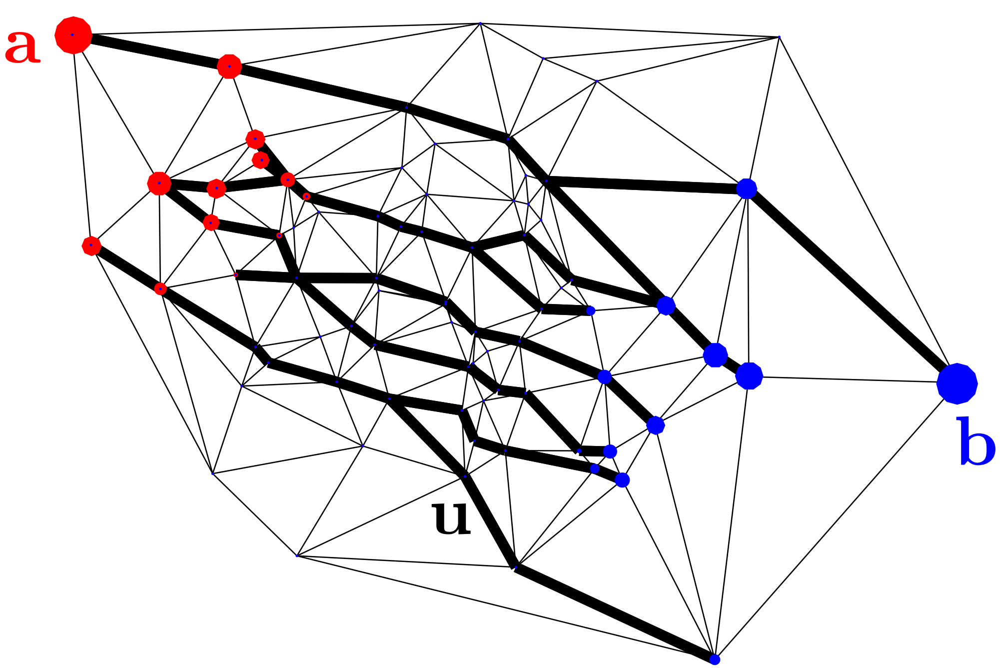
Euclidean and Graphs W1

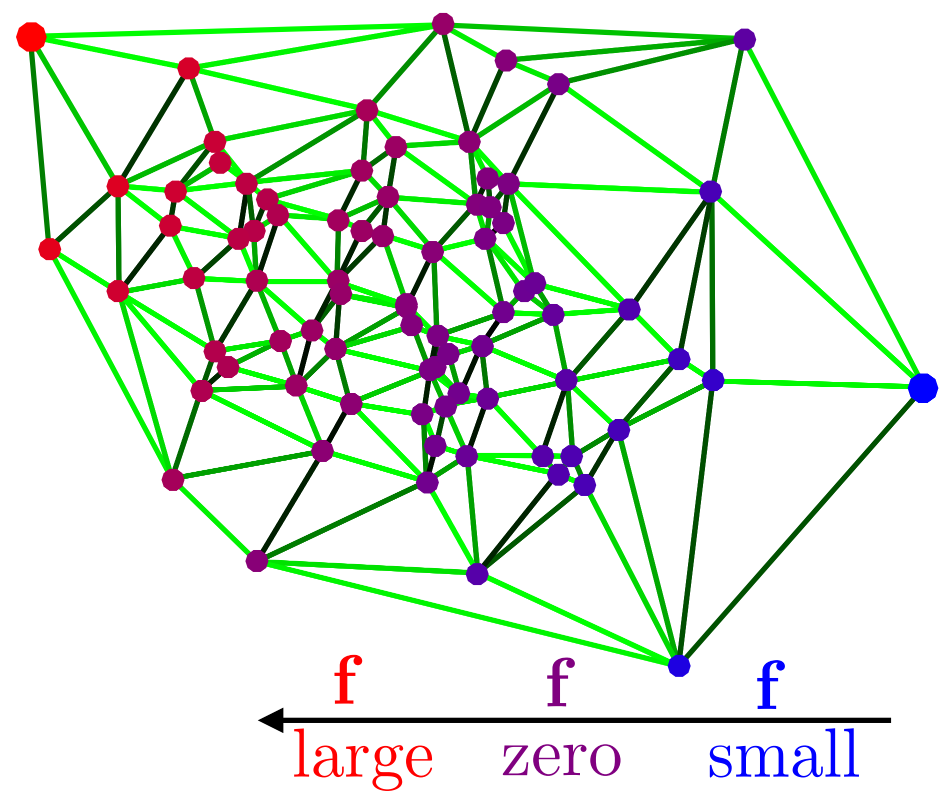
Case $d(x,y) = \|x-y\|$:

$$\|\alpha - \beta\|_{W_1} = \max_{\|\nabla f(x)\|_\infty \leqslant 1} \langle f,\, \alpha - \beta \rangle = \min_{\mathrm{div}(u)=\alpha-\beta} \int \|u(x)\| \mathrm{d}x$$

On graph: $d(x,y) = \mathrm{Geod}_w(x,y)$ $\qquad \nabla_{i,j}\mathbf{f} \overset{\text{def.}}{=} w_{i,j}^{-1}(\mathbf{f}_i - \mathbf{f}_j)$

$$\|\mathbf{a} - \mathbf{b}\|_{W_1} = \max_{|\nabla_{i,j}\mathbf{f}| \leqslant 1} \langle \mathbf{f},\, \mathbf{a} - \mathbf{b} \rangle = \min_{\mathrm{div}(\mathbf{u})=\mathbf{a}-\mathbf{b}} \sum_{i,j} w_{i,j} |\mathbf{u}_{i,j}|$$

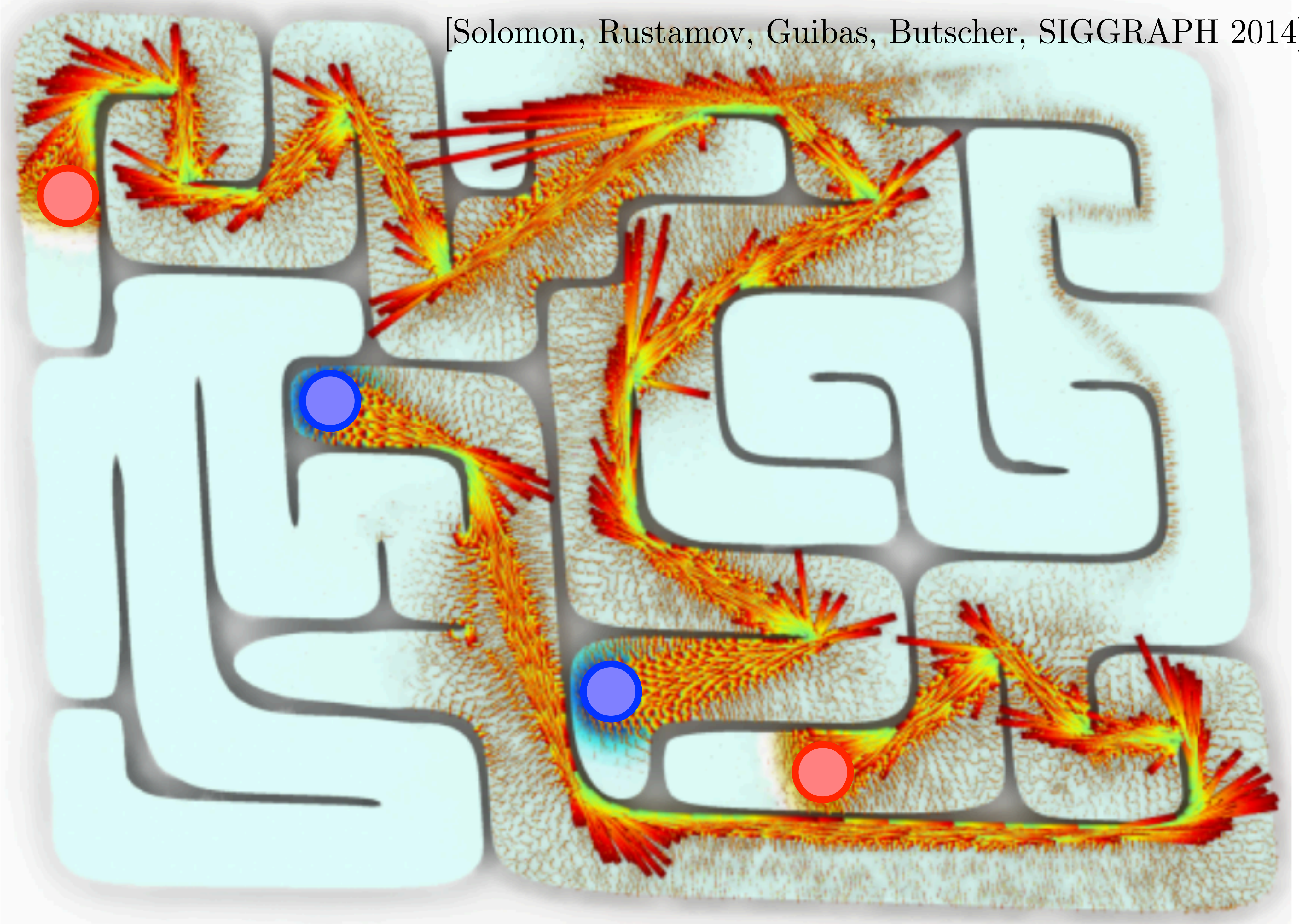




W_1 On Surfaces

[Solomon, Rustamov, Guibas, Butscher, SIGGRAPH 2014]

W_1 On Sub-domains



Dual Norms

$$\|\alpha - \beta\|_B \stackrel{\text{def.}}{=} \sup_{f \in B} \int f(x)\mathrm{d}(\alpha - \beta)(x)$$

$B = \{f \ : \ \|f\|_\infty \leqslant 1\}$ $\longrightarrow$ Total variation

$B = \{f \ : \ \|\nabla f\|_\infty \leqslant 1\}$ $\longrightarrow$ W_1

$B = \{f \ : \ \|\nabla f\|_\infty + \|f\|_\infty \leqslant 1\}$ $\longrightarrow$ Flat norm (unbalanced)

$B = \left\{f \ : \ \|\nabla^{(k)} f\|_2 \leqslant 1\right\}$ $\longrightarrow$ Sobolev $\dot{H}^{-k}$ $\quad (k > \frac{d}{2})$

$B = \{\text{neural networks}\}$ $\longrightarrow$?? (Non-convex)

Dual Norms

$$\|\alpha - \beta\|_B \overset{\text{def.}}{=} \sup_{f \in B} \int f(x)\mathrm{d}(\alpha - \beta)(x)$$

$B = \{f \ : \ \|f\|_\infty \leqslant 1\} \longrightarrow$ Total variation

$B = \{f \ : \ \|\nabla f\|_\infty \leqslant 1\} \longrightarrow$ W_1

$B = \{f \ : \ \|\nabla f\|_\infty + \|f\|_\infty \leqslant 1\} \longrightarrow$ Flat norm (unbalanced)

$B = \left\{f \ : \ \|\nabla^{(k)} f\|_2 \leqslant 1\right\} \longrightarrow$ Sobolev $\dot{H}^{-k}$ $\quad (k > \frac{d}{2})$

$B = \{\text{neural networks}\} \longrightarrow$?? (Non-convex)

Proposition:

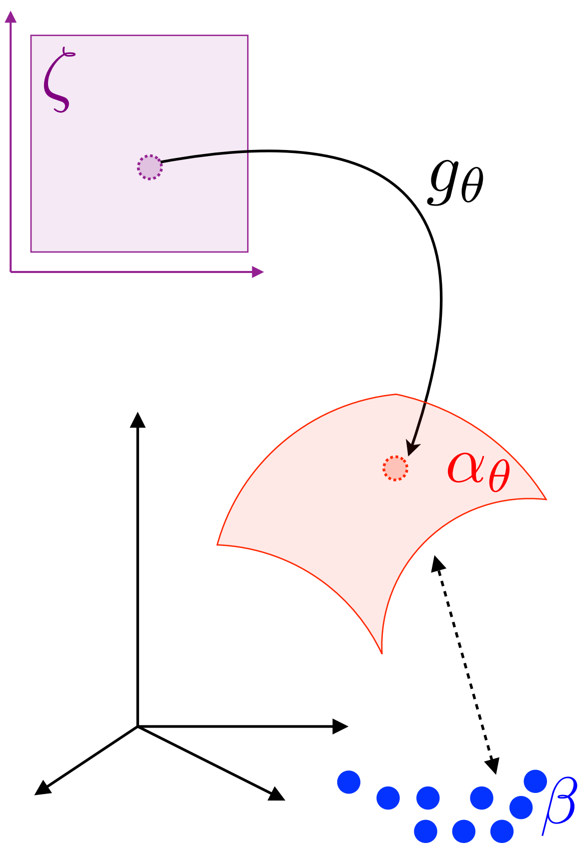
If $\mathcal{C}(\mathcal{X}) \subset \overline{\mathrm{Span}(B)}$, then $\|\cdot\|_B$ is stronger than weak convergence.

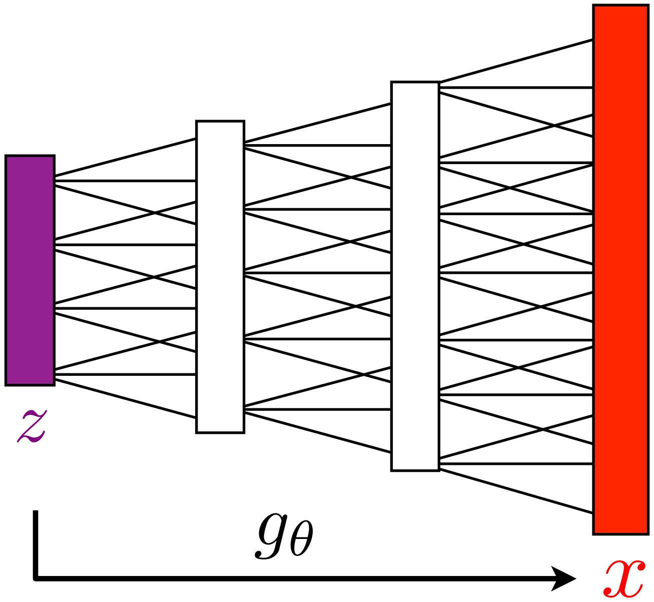
If $B \subset \mathcal{C}(\mathcal{X})$ is compact, then $\|\cdot\|_B$ is weaker than weak convergence.

$\|\cdot\|_B$ metrizes weak convergence.

Wasserstein GANs

Generative model fitting: $\min_{\theta} \|\alpha_\theta - \beta\|_B$

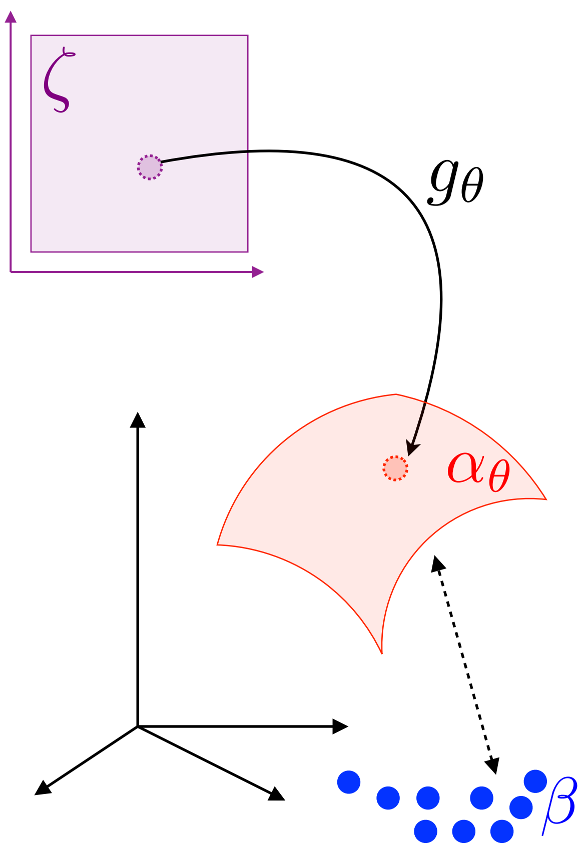


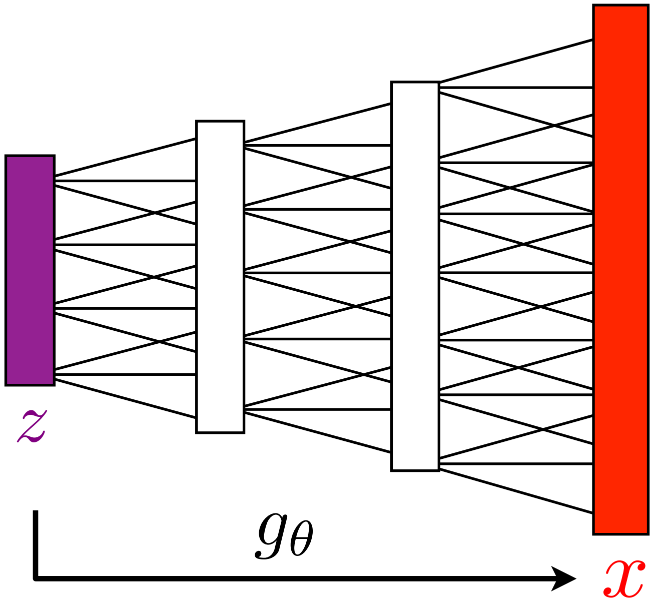


Wasserstein GANs

Generative model fitting: $\min_\theta \|\alpha_\theta - \beta\|_B$

$$= \min_\theta \sup_{f \in B} \int f \mathrm{d}\alpha_\theta - \frac{1}{m} \sum_j f(y_j)$$



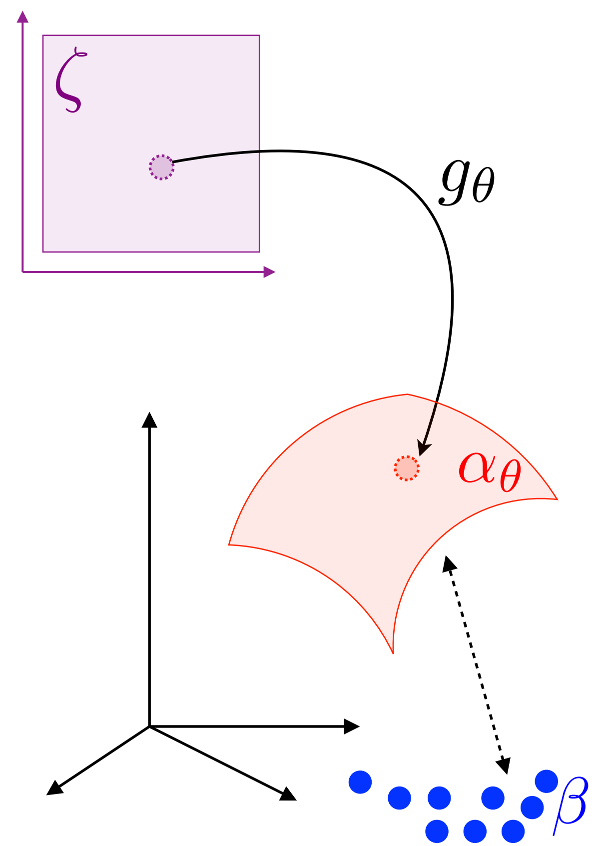


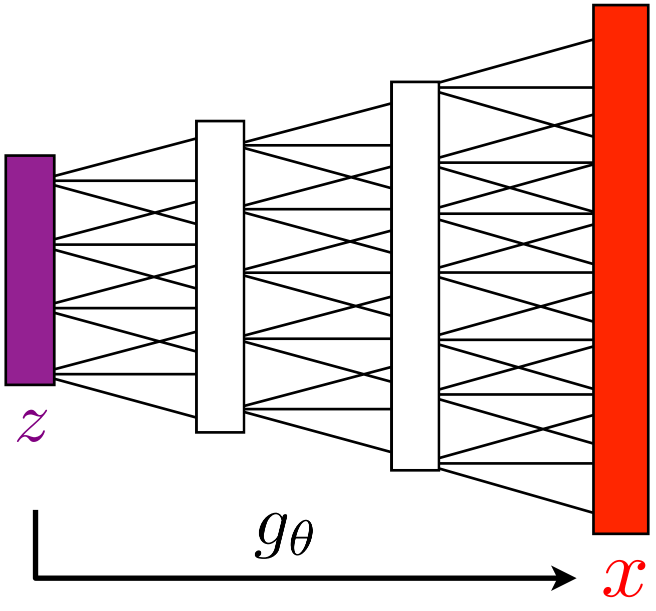
Wasserstein GANs

Generative model fitting: $\min_\theta \|\alpha_\theta - \beta\|_B$

$$= \min_\theta \sup_{f \in B} \int f \mathrm{d}\alpha_\theta - \frac{1}{m} \sum_j f(y_j)$$

$$= \min_\theta \sup_{f \in B} \int f(g_\theta(z)) \mathrm{d}\zeta - \frac{1}{m} \sum_j f(y_j)$$





Wasserstein GANs

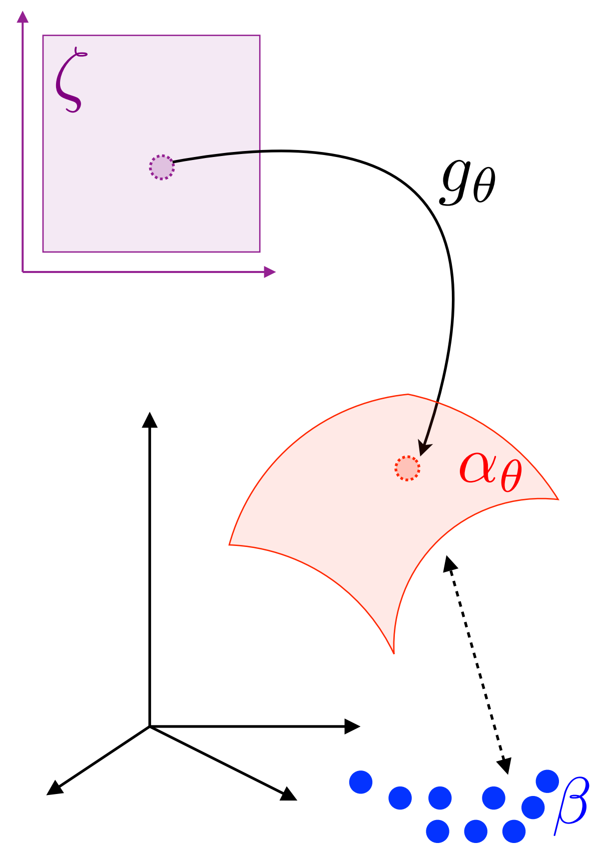
Generative model fitting: $\min_\theta \|\alpha_\theta - \beta\|_B$

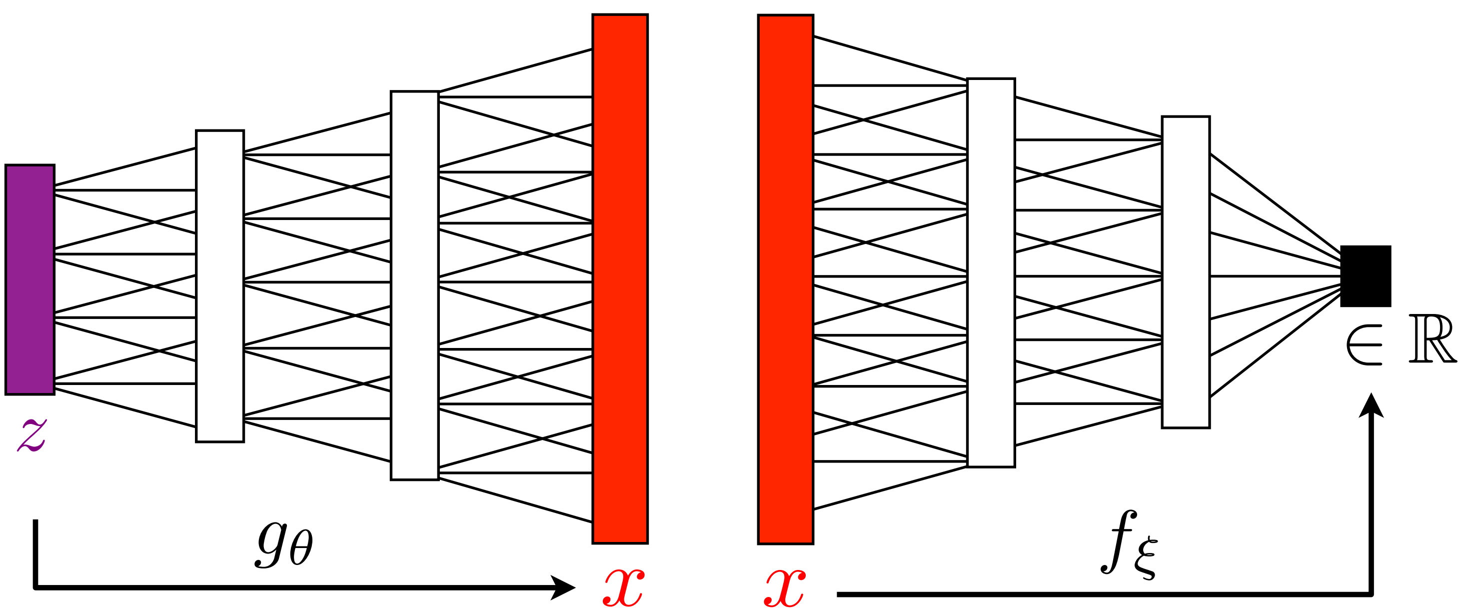
$$= \min_\theta \sup_{f \in B} \int f \mathrm{d}\alpha_\theta - \frac{1}{m}\sum_j f(y_j)$$

$$= \min_\theta \sup_{f \in B} \int f(g_\theta(z)) \mathrm{d}\zeta - \frac{1}{m}\sum_j f(y_j)$$

$$\Big\downarrow \{f_\xi\}_\xi \subset B$$

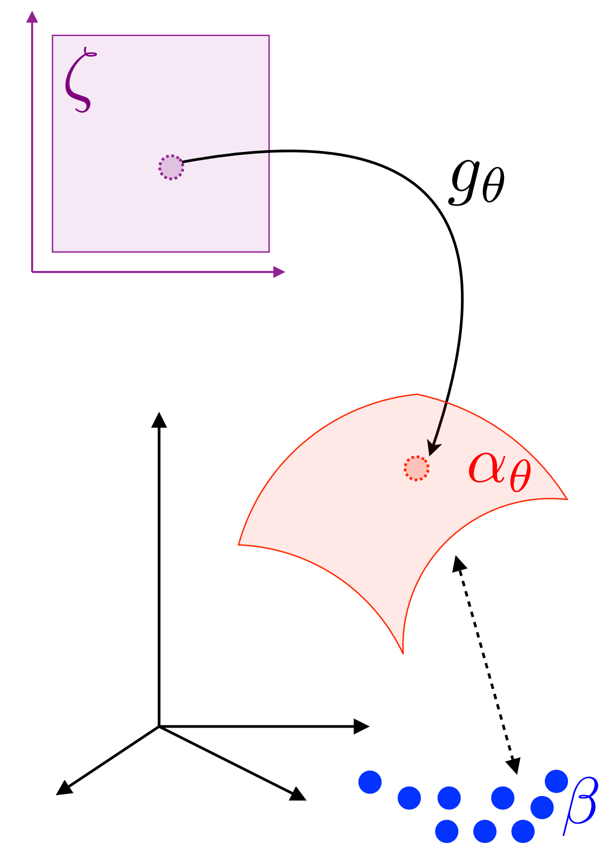
$$\neq \min_\theta \sup_\xi \int f_\xi(g_\theta(z)) \mathrm{d}\zeta - \sum_j f_\xi(y_j)$$





Wasserstein GANs

Generative model fitting: $\min_\theta \|\alpha_\theta - \beta\|_B$

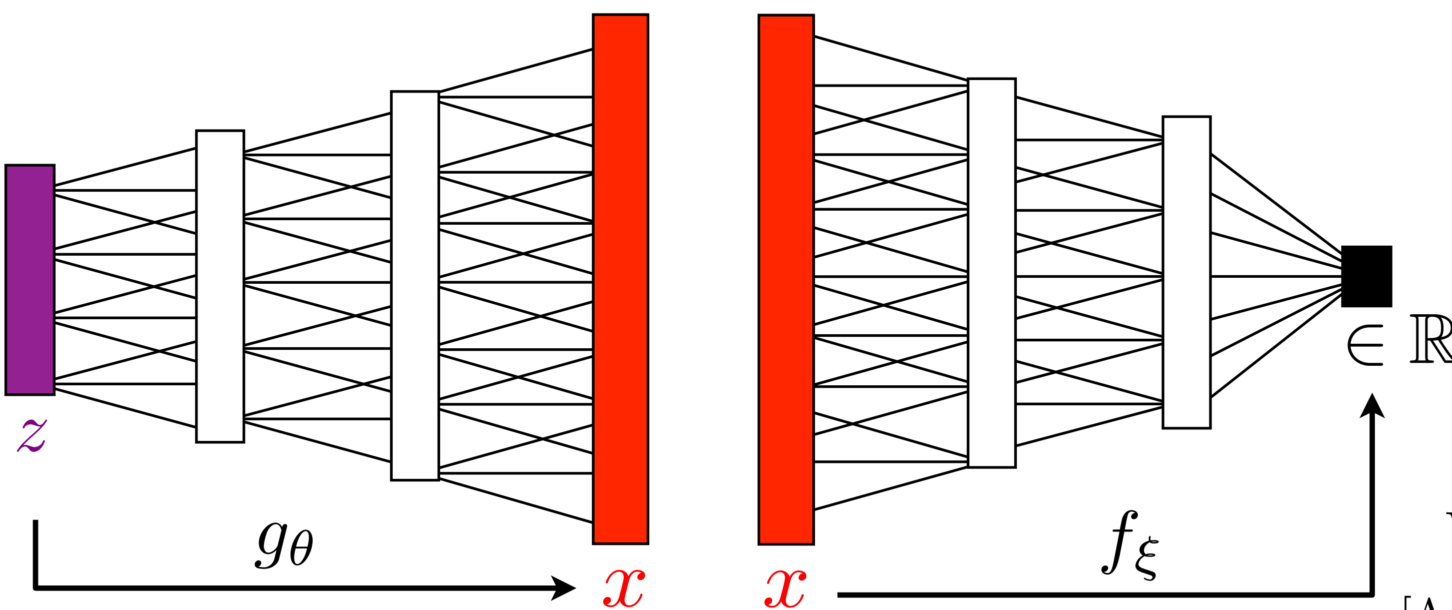


$$= \min_\theta \sup_{f \in B} \int f \mathrm{d}\alpha_\theta - \frac{1}{m} \sum_j f(y_j)$$

$$= \min_\theta \sup_{f \in B} \int f(g_\theta(z)) \mathrm{d}\zeta - \frac{1}{m} \sum_j f(y_j)$$

$$\Big\downarrow \; \{f_\xi\}_\xi \subset B$$

$$\neq \min_\theta \sup_\xi \int f_\xi(g_\theta(z)) \mathrm{d}\zeta - \sum_j f_\xi(y_j)$$



Wasserstein GANs,

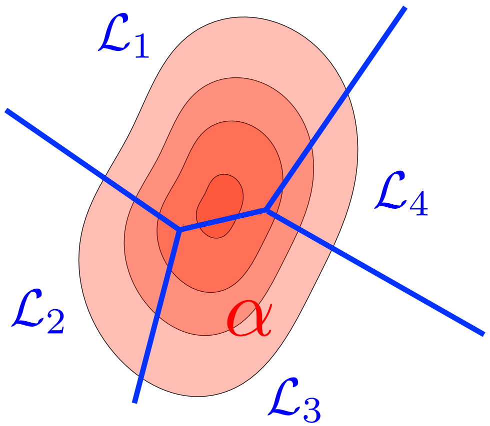
Weight clipping $\|\xi\|_\infty \leqslant 1$.

[Arjovsky, Chintala,Bottou, 2017]:

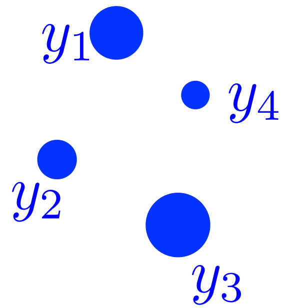
Overview

- Dual Problem
- W_1
- **Semi-discrete Problem**
- Optimal Quantization

Semi-discrete Problem



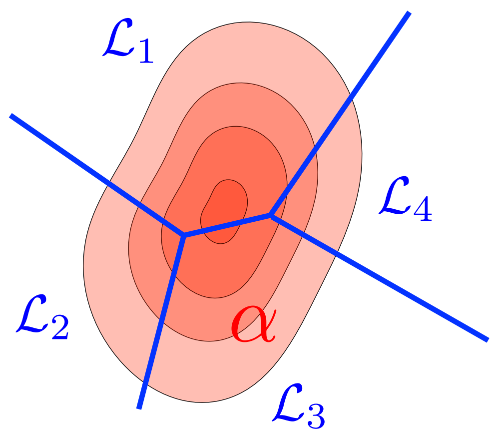
$\beta = \sum_{j=1}^{m} \mathbf{b}_j \delta_{y_j}$



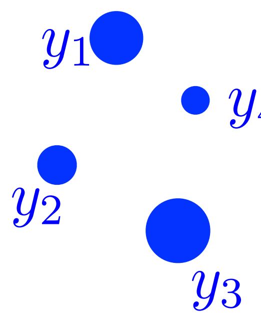
Optimal Monge map:

$$T : \mathcal{L}_j \longmapsto y_j$$

Semi-discrete Problem



$\beta = \sum_{j=1}^{m} \mathbf{b}_j \delta_{y_j}$

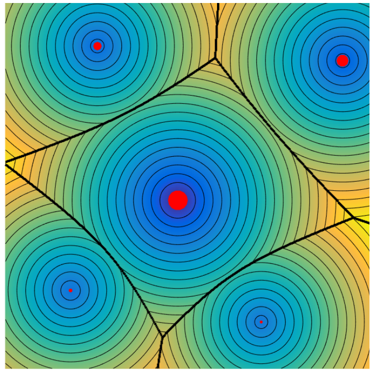


Optimal Monge map:

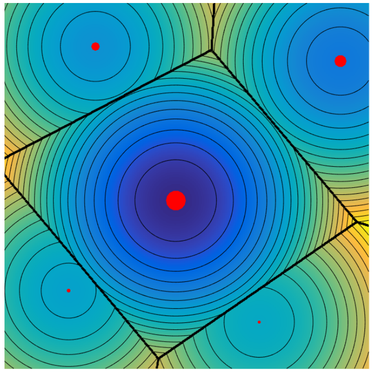
$$T : \mathcal{L}_j \longmapsto y_j$$

c-transform:

$$\mathbf{g}^c(x) \overset{\text{def.}}{=} \min_{1 \leqslant j \leqslant m} c(x, y_j) - \mathbf{g}_j$$

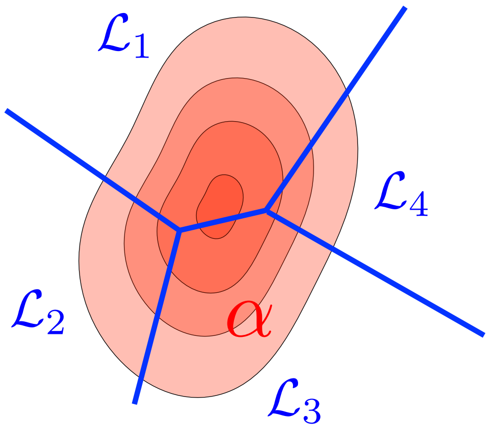


$c(x, y) = \|x - y\|$

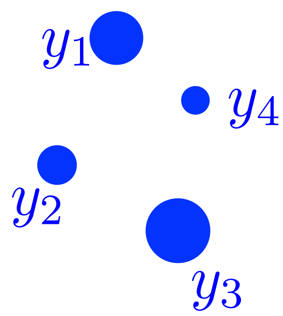


$c(x, y) = \|x - y\|^2$

Semi-discrete Problem



$\beta = \sum_{j=1}^m \mathbf{b}_j \delta_{y_j}$

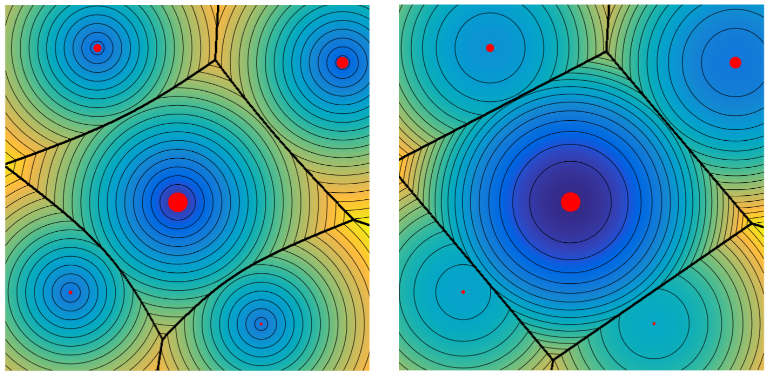


Optimal Monge map:

$$T : \mathcal{L}_j \longmapsto y_j$$

c-transform:

$$\mathbf{g}^c(x) \stackrel{\text{def.}}{=} \min_{1 \leqslant j \leqslant m} c(x, y_j) - \mathbf{g}_j$$

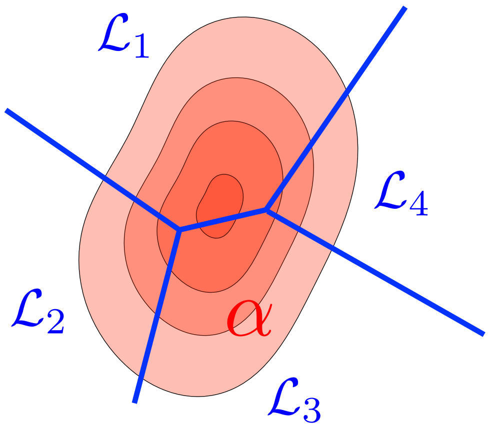


$c(x,y) = \|x-y\|$ $\quad$ $c(x,y) = \|x-y\|^2$

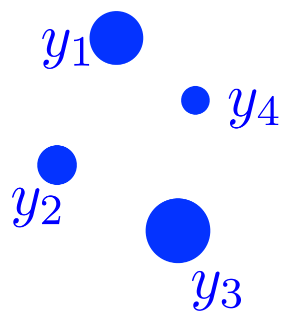
Semi-discrete OT:

$$\max_{\mathbf{g} \in \mathbb{R}^m} \mathcal{E}(\mathbf{g}) \stackrel{\text{def.}}{=} \int \mathbf{g}^c(x)(x)\mathrm{d}\alpha(x) + \sum_j \mathbf{g}_j \mathbf{b}_j$$

Semi-discrete Problem



$\beta = \sum_{j=1}^m \mathbf{b}_j \delta_{y_j}$



Optimal Monge map:

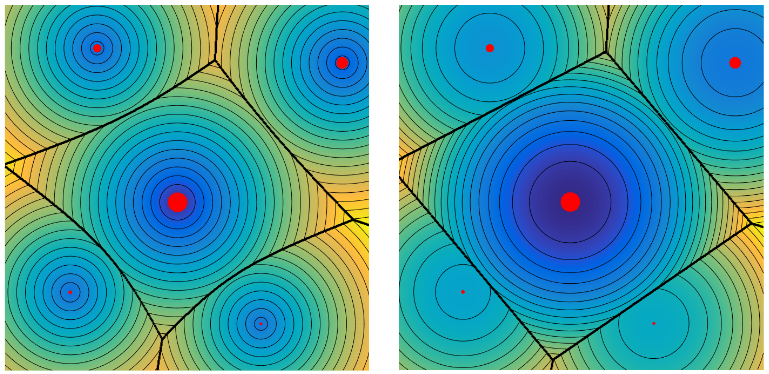
$$T : \mathcal{L}_j \longmapsto y_j$$

c-transform:

$$\mathbf{g}^c(x) \stackrel{\text{def.}}{=} \min_{1 \leqslant j \leqslant m} c(x, y_j) - \mathbf{g}_j$$

Semi-discrete OT:

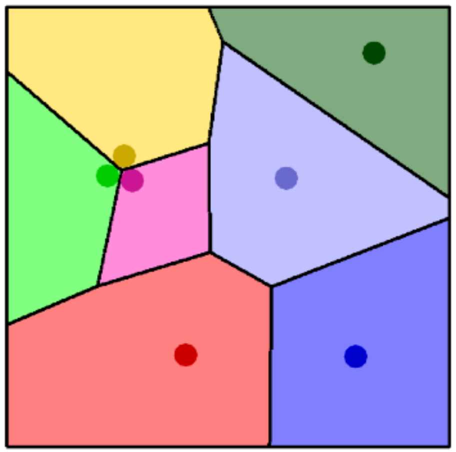
$$\max_{\mathbf{g} \in \mathbb{R}^m} \mathcal{E}(\mathbf{g}) \stackrel{\text{def.}}{=} \int \mathbf{g}^c(x)(x) \mathrm{d}\alpha(x) + \sum_j \mathbf{g}_j \mathbf{b}_j$$



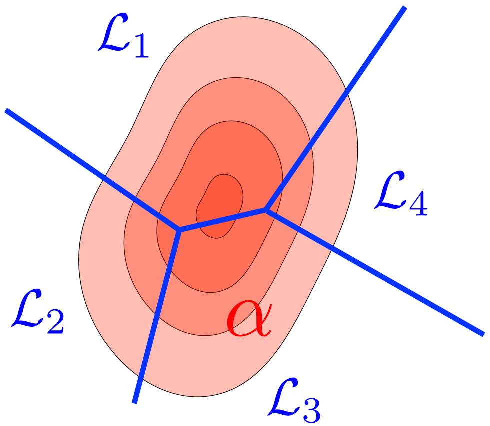
$c(x,y) = \|x-y\|$ $\qquad$ $c(x,y) = \|x-y\|^2$

Laguerre cells:

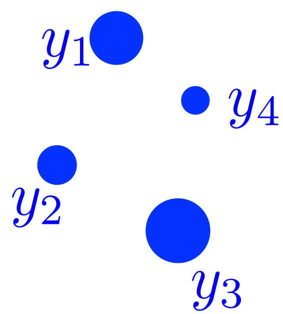
$$\mathcal{L}_j(\mathbf{g}) \stackrel{\text{def.}}{=} \{x \; ; \; \forall \ell, \; c(x, y_j) - \mathbf{g}_j \leqslant c(x, y_\ell) - \mathbf{g}_\ell\}$$



Semi-discrete Problem



$\beta = \sum_{j=1}^m \mathbf{b}_j \delta_{y_j}$

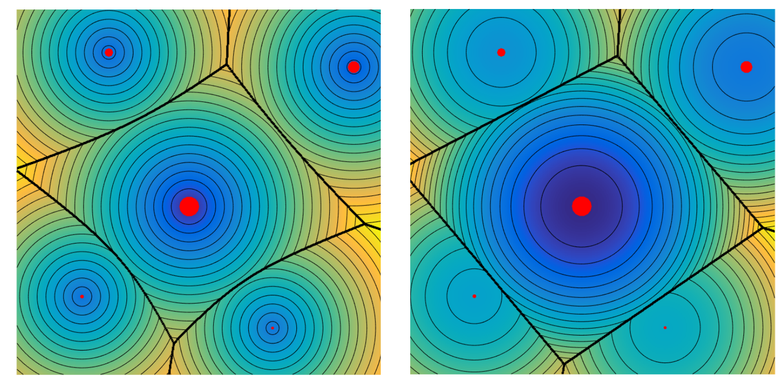


Optimal Monge map:

$$T : \mathcal{L}_j \longmapsto y_j$$

c-transform:

$$\mathbf{g}^c(x) \overset{\text{def.}}{=} \min_{1 \leqslant j \leqslant m} c(x, y_j) - \mathbf{g}_j$$



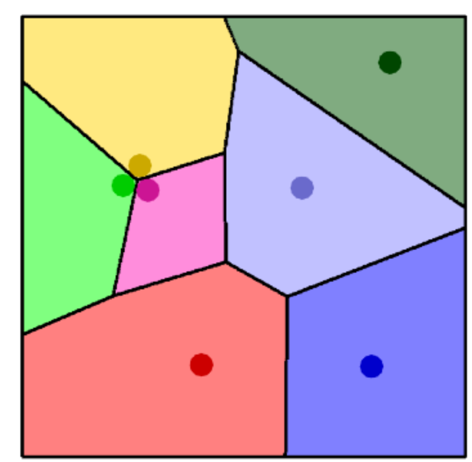
$c(x,y) = \|x-y\|$ $\quad$ $c(x,y) = \|x-y\|^2$

Semi-discrete OT:

$$\max_{\mathbf{g} \in \mathbb{R}^m} \mathcal{E}(\mathbf{g}) \overset{\text{def.}}{=} \int \mathbf{g}^c(x)(x) \mathrm{d}\alpha(x) + \sum_j \mathbf{g}_j \mathbf{b}_j$$

Laguerre cells:

$$\mathcal{L}_j(\mathbf{g}) \overset{\text{def.}}{=} \{ x \; ; \; \forall \ell, \; c(x, y_j) - \mathbf{g}_j \leqslant c(x, y_\ell) - \mathbf{g}_\ell \}$$

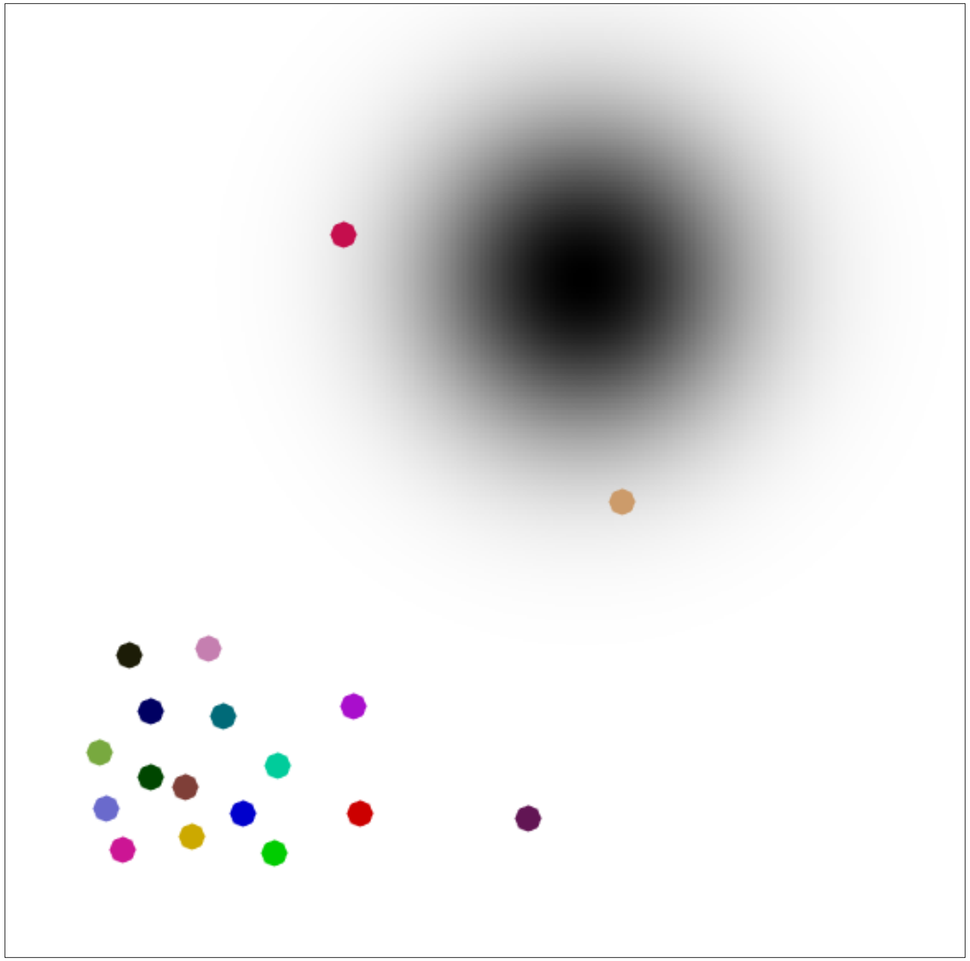


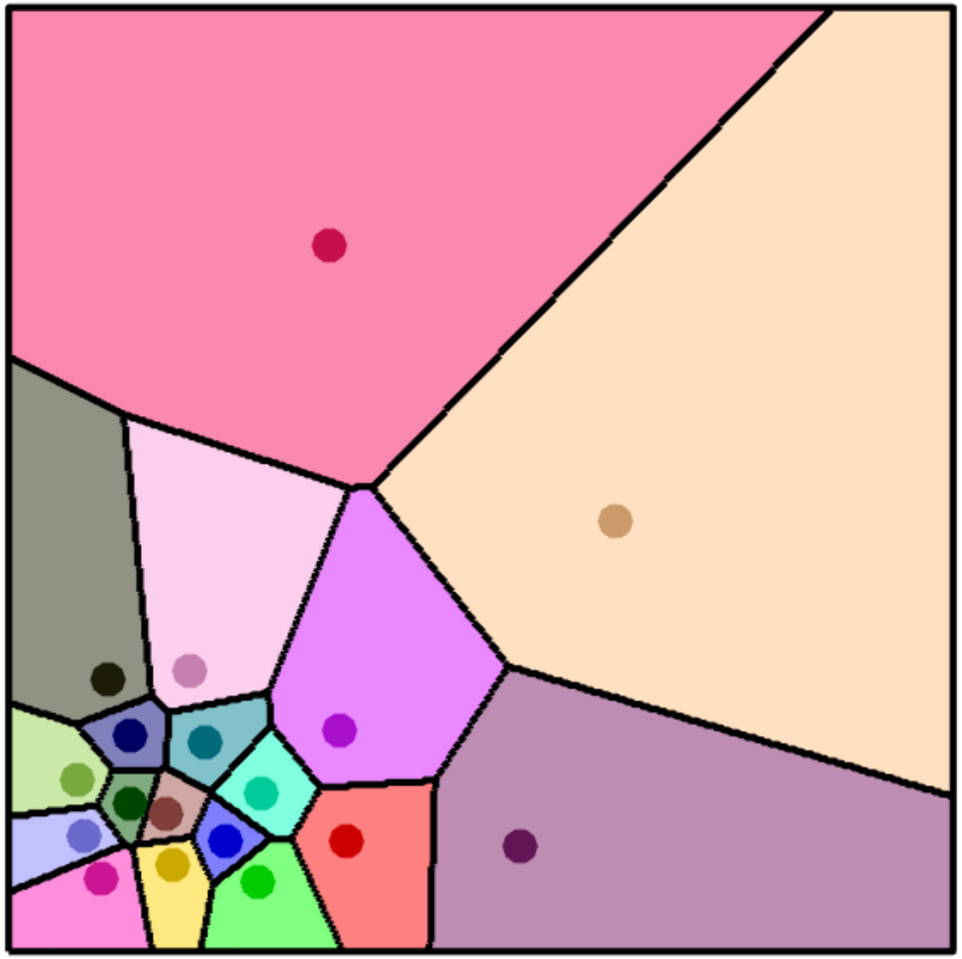
Proposition: $\nabla \mathcal{E}(\mathbf{g})_j = - \int_{\mathcal{L}_j(\mathbf{g})} \mathrm{d}\alpha + \mathbf{b}_j$

Semi-discrete Optimization

Gradient descent:

$$\mathbf{g} \leftarrow (1-\tau)\mathbf{g} + \tau \int_{\mathcal{L}_j(\mathbf{g})} \mathrm{d}\alpha$$



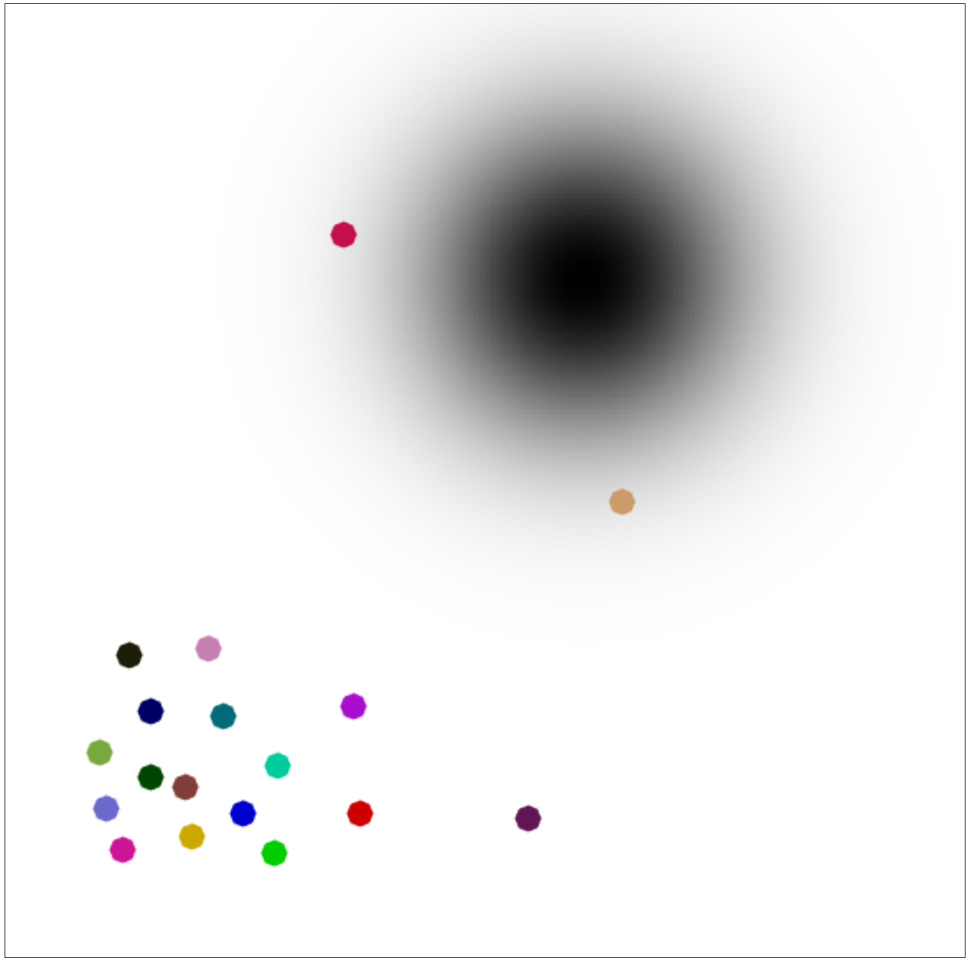


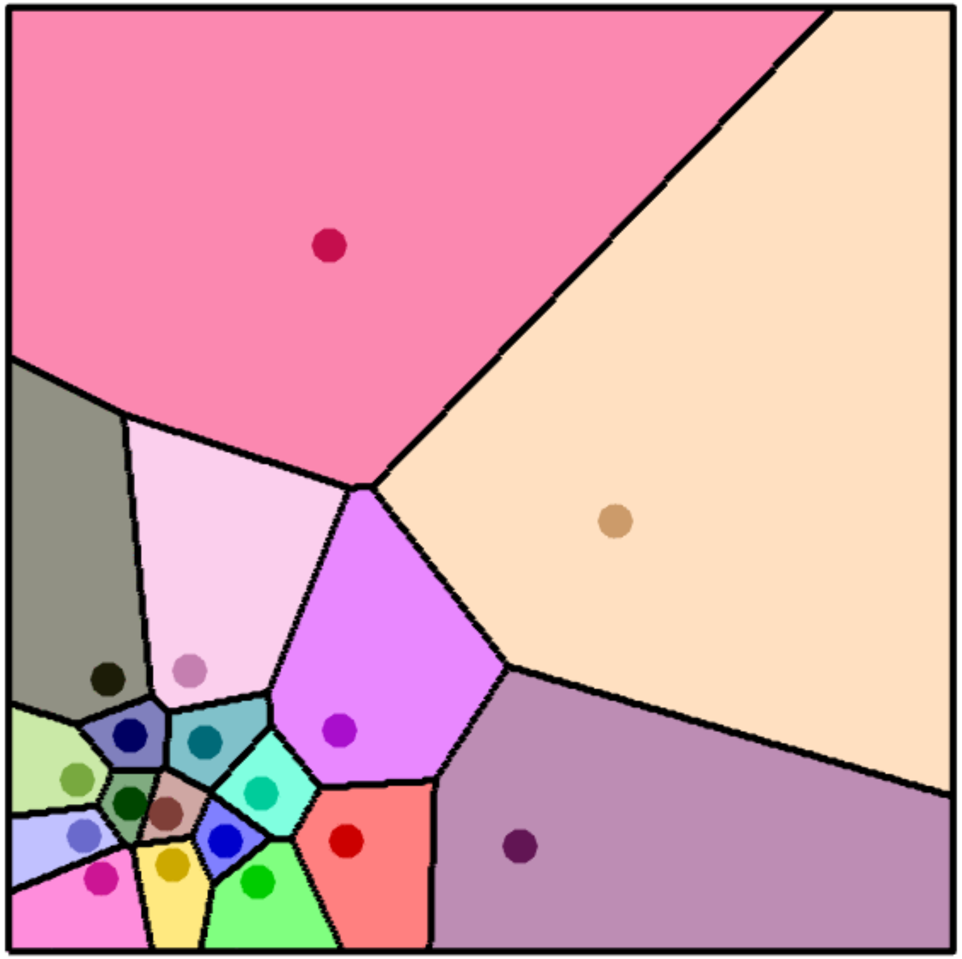
→ *In practice:* use quasi-Newton (L-BFGS).

Semi-discrete Optimization

Gradient descent:

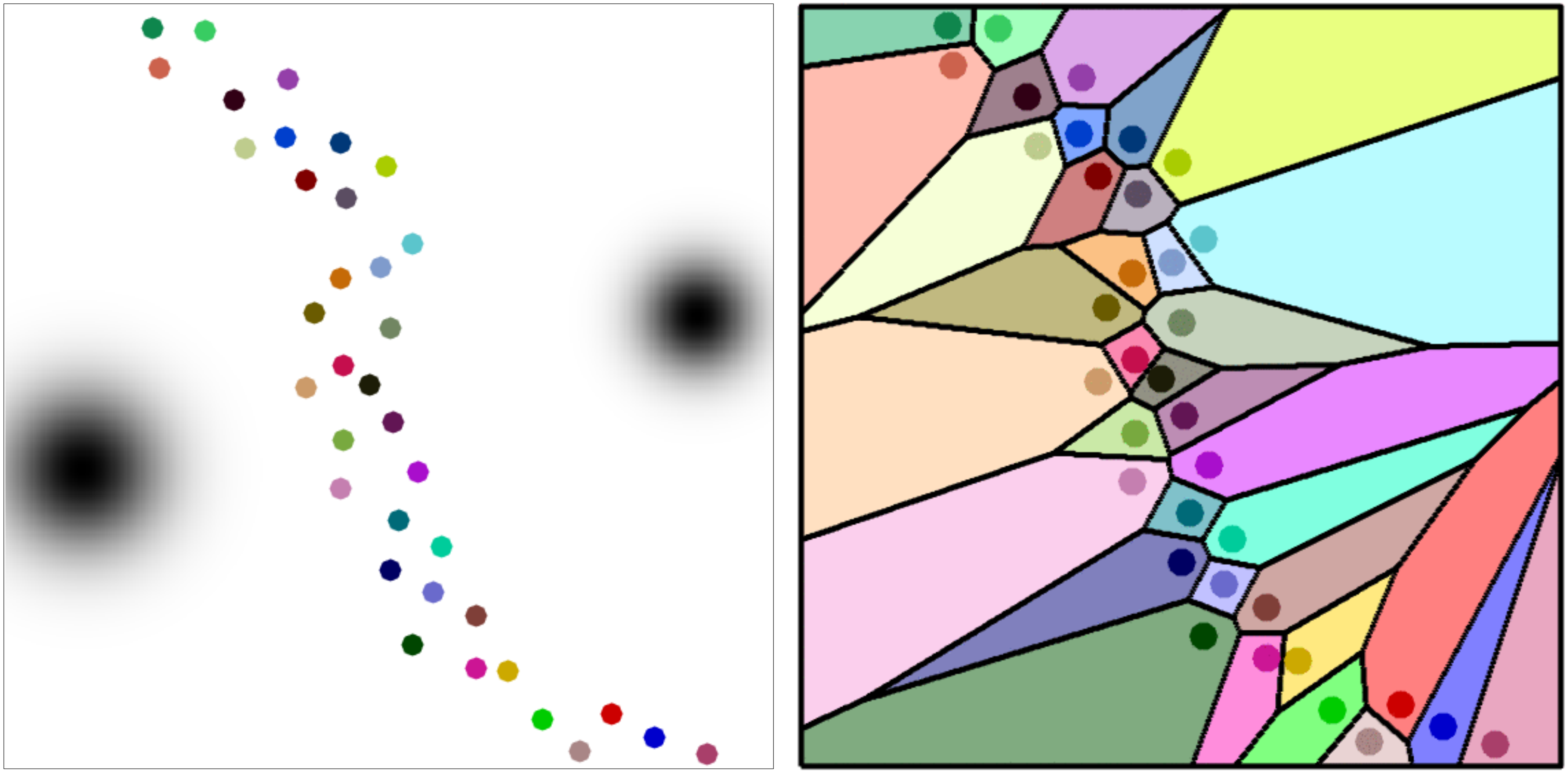
$$\mathbf{g} \leftarrow (1-\tau)\mathbf{g} + \tau \int_{\mathcal{L}_j(\mathbf{g})} \mathrm{d}\alpha$$



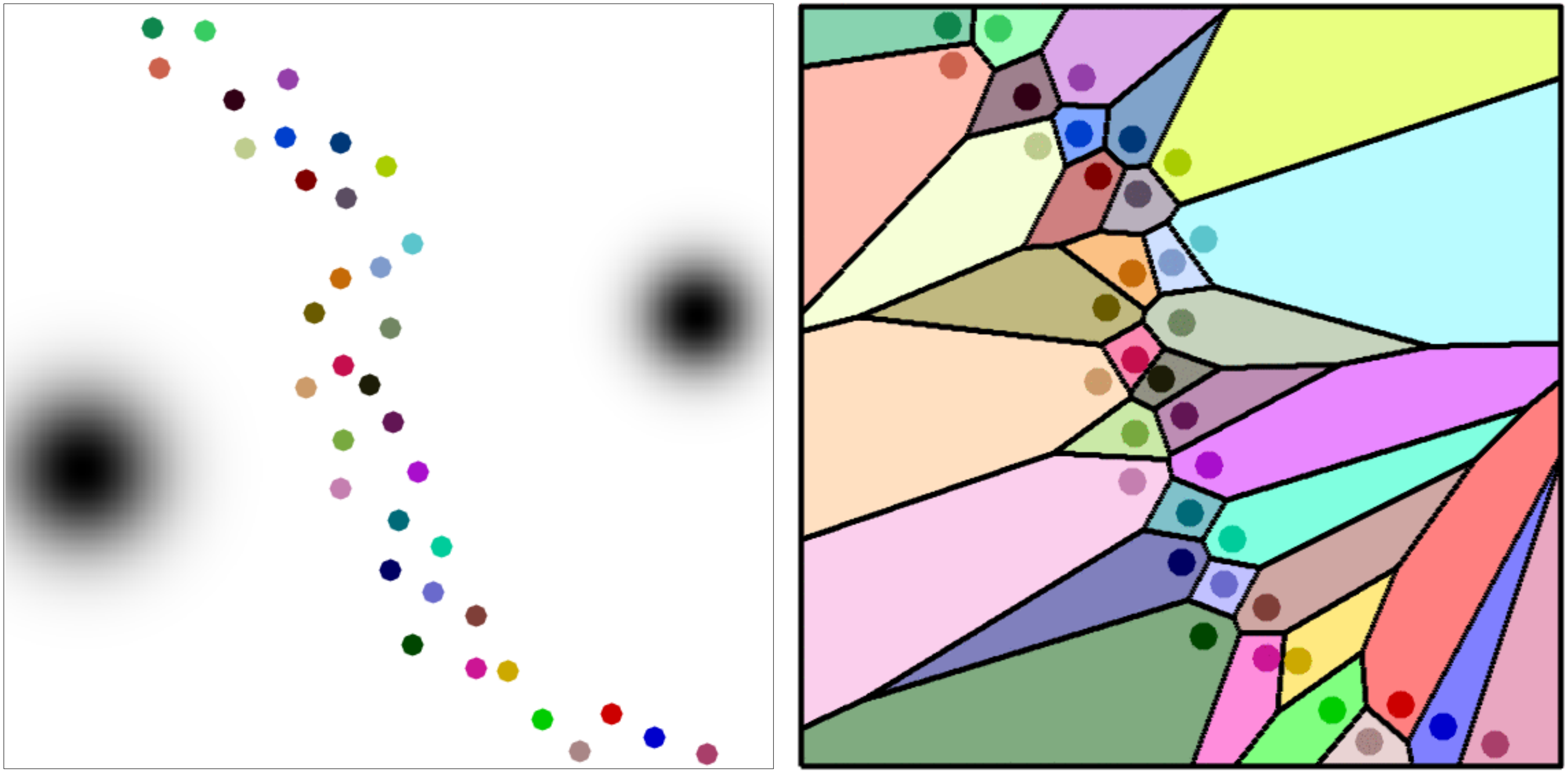


$\rightarrow$ *In practice:* use quasi-Newton (L-BFGS).

Semi-discrete Optimization



Semi-discrete Optimization



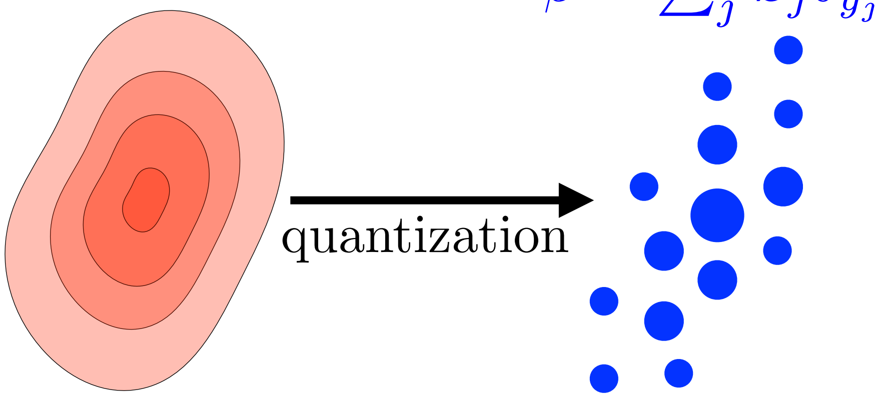
Overview

- Dual Problem
- W_1
- Semi-discrete Problem
- **Optimal Quantization**

Optimal Quantization

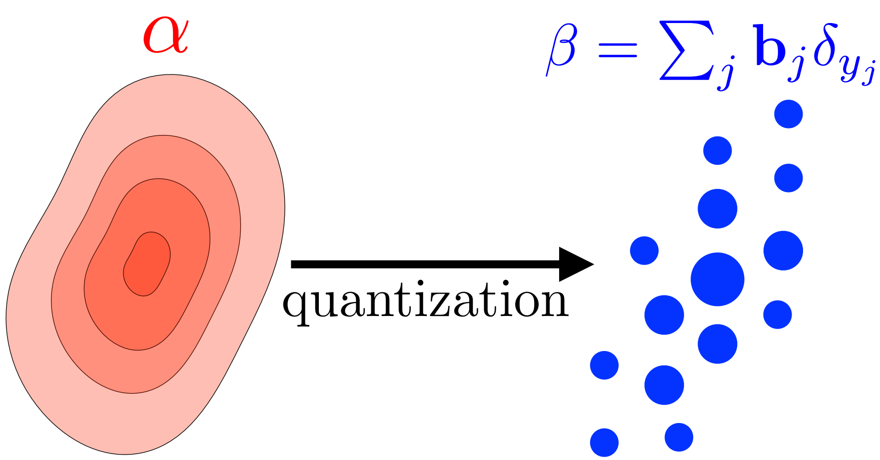
α

$\beta = \sum_j \mathbf{b}_j \delta_{y_j}$



$$\mathcal{Q}_m(\alpha) \stackrel{\text{def.}}{=} \min_{\mathbf{b},\, Y} \mathrm{W}_p(\alpha, \textstyle\sum_j \mathbf{b}_j \delta_{y_j})$$

Optimal Quantization

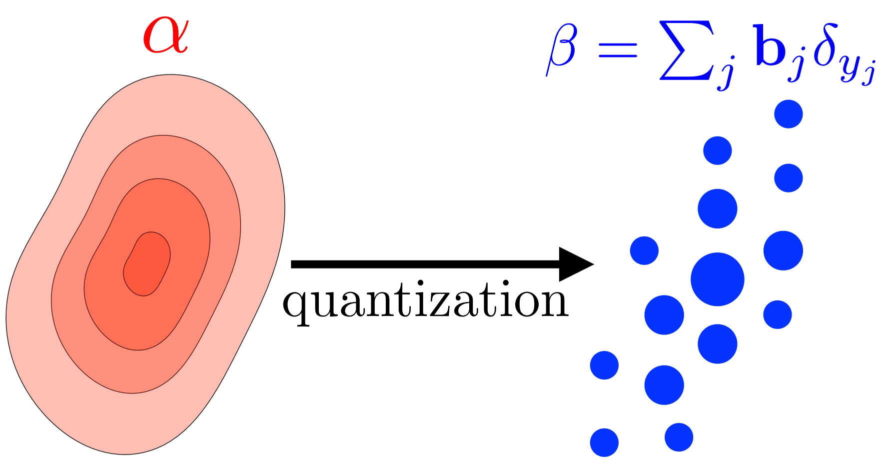


$$\mathcal{Q}_m(\alpha) \stackrel{\text{def.}}{=} \min_{\mathbf{b},\, Y} \mathrm{W}_p\big(\alpha, \textstyle\sum_j \mathbf{b}_j \delta_{y_j}\big)$$

convex

non-convex

Optimal Quantization

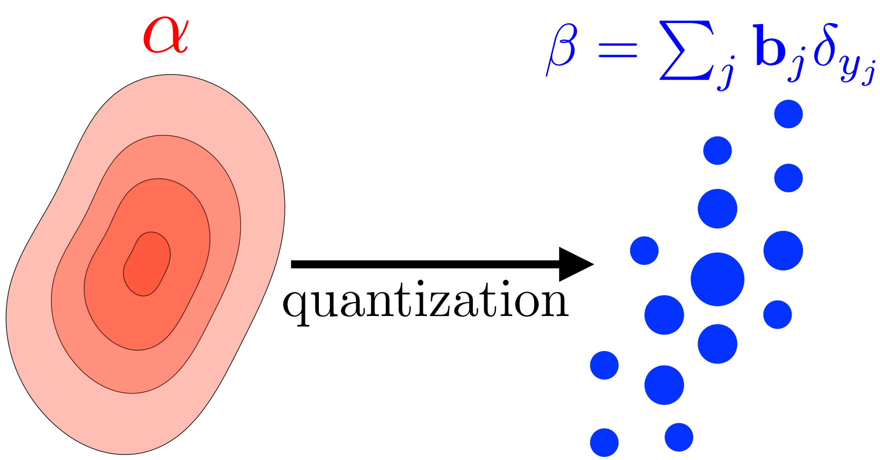


$$\mathcal{Q}_m(\alpha) \stackrel{\text{def.}}{=} \min_{\mathbf{b},\, Y} \mathrm{W}_p(\alpha, \textstyle\sum_j \mathbf{b}_j \delta_{y_j})$$

convex non-convex

In general: $\mathcal{Q}_m(\alpha) \sim 1/m^{1/d}$.

Optimal Quantization



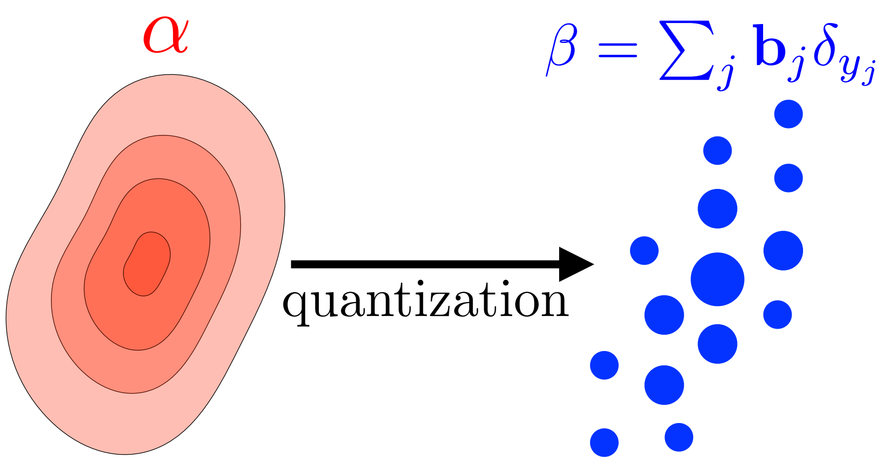
$$\mathcal{Q}_m(\alpha) \overset{\text{def.}}{=} \min_{\mathbf{b},\, Y} \mathrm{W}_p(\alpha, \textstyle\sum_j \mathbf{b}_j \delta_{y_j})$$

convex — non-convex

In general: $\mathcal{Q}_m(\alpha) \sim 1/m^{1/d}$.

$$\mathcal{Q}_m(\alpha) = \min_{Y} \max_{\mathbf{g} \in \mathbb{R}^m} \int \mathbf{g}^c(x)(x) \mathrm{d}\alpha(x) + \min_{\mathbf{b}} \textstyle\sum_j \mathbf{g}_j \mathbf{b}_j$$

Optimal Quantization



$$\mathcal{Q}_m(\alpha) \overset{\text{def.}}{=} \min_{\mathbf{b},\, Y} \mathrm{W}_p(\alpha, \textstyle\sum_j \mathbf{b}_j \delta_{y_j})$$

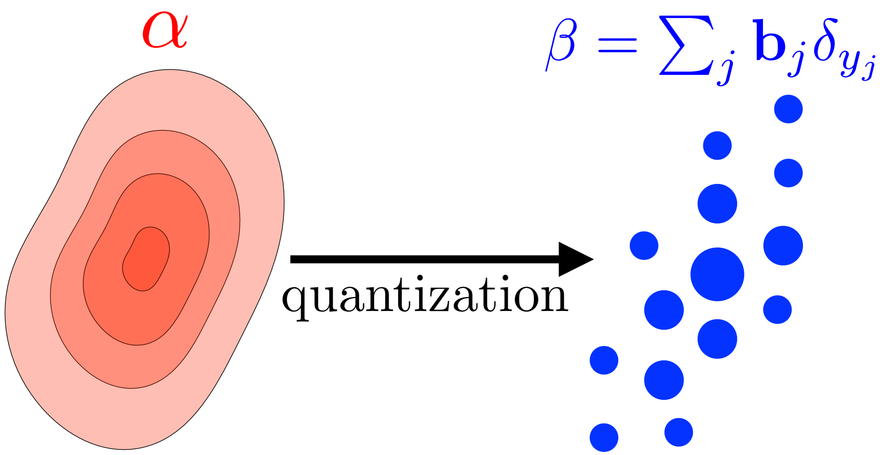
convex | non-convex

In general: $\mathcal{Q}_m(\alpha) \sim 1/m^{1/d}$.

$$\mathcal{Q}_m(\alpha) = \min_{Y} \max_{\mathbf{g} \in \mathbb{R}^m} \int \mathbf{g}^c(x)(x) \mathrm{d}\alpha(x) + \boxed{\min_{\mathbf{b}} \textstyle\sum_j \mathbf{g}_j \mathbf{b}_j} \implies \boxed{\mathbf{g} = 0}$$

Voronoi cells: $\mathcal{L}_j(\mathbf{0}) = \mathbb{V}_j(Y) \overset{\text{def.}}{=} \{x \,;\, \forall \ell,\ \|x - y_j\| \leqslant \|x - y_\ell\|\}$

Optimal Quantization



$$\mathcal{Q}_m(\alpha) \overset{\text{def.}}{=} \min_{\mathbf{b},\, Y} \mathrm{W}_p(\alpha, \textstyle\sum_j \mathbf{b}_j \delta_{y_j})$$

convex non-convex

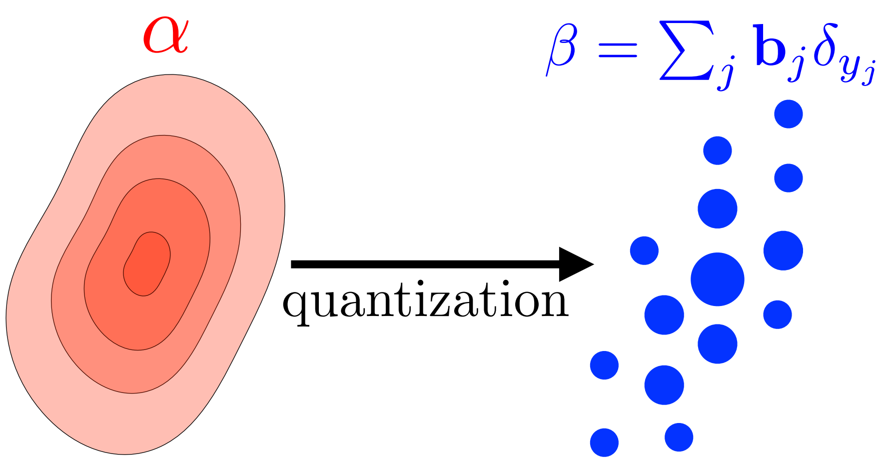
In general: $\mathcal{Q}_m(\alpha) \sim 1/m^{1/d}$.

$$\mathcal{Q}_m(\alpha) = \min_{Y} \max_{\mathbf{g}\in\mathbb{R}^m} \int \mathbf{g}^c(x)(x)\mathrm{d}\alpha(x) + \boxed{\min_{\mathbf{b}} \textstyle\sum_j \mathbf{g}_j \mathbf{b}_j} \implies \boxed{\mathbf{g} = 0}$$

Voronoi cells: $\mathcal{L}_j(\mathbf{0}) = \mathbb{V}_j(Y) \overset{\text{def.}}{=} \{x \,;\, \forall \ell,\ \|x - y_j\| \leqslant \|x - y_\ell\|\}$

Proposition: $\mathcal{Q}_m(\alpha) = \min_{Y} \int \min_{1\leqslant j\leqslant m} c(x, y_j)\mathrm{d}\alpha(x)$

Optimal Quantization



$$\mathcal{Q}_m(\alpha) \overset{\text{def.}}{=} \min_{\mathbf{b},\, Y} \mathrm{W}_p\Big(\alpha, \textstyle\sum_j \mathbf{b}_j \delta_{y_j}\Big)$$

convex — non-convex

In general: $\mathcal{Q}_m(\alpha) \sim 1/m^{1/d}$.

$$\mathcal{Q}_m(\alpha) = \min_{Y} \max_{\mathbf{g}\in\mathbb{R}^m} \int \mathbf{g}^c(x)(x)\mathrm{d}\alpha(x) + \boxed{\min_{\mathbf{b}} \textstyle\sum_j \mathbf{g}_j \mathbf{b}_j} \implies \boxed{\mathbf{g} = 0}$$

Voronoi cells: $\mathcal{L}_j(\mathbf{0}) = \mathbb{V}_j(Y) \overset{\text{def.}}{=} \{x \,;\, \forall \ell,\ \|x - y_j\| \leqslant \|x - y_\ell\|\}$

Proposition: $\mathcal{Q}_m(\alpha) = \min\limits_{Y} \displaystyle\int \min_{1\leqslant j\leqslant m} c(x, y_j)\mathrm{d}\alpha(x)$

Local minimizers: $y_j = \operatorname*{argmin}\limits_{y} \displaystyle\int_{\mathbb{V}_j(Y)} c(x, y)\mathrm{d}\alpha(x)$

Lloyd's Algorithm

$$\text{Cost } c(x,y) = \|x-y\|^2$$

$$\underset{y}{\operatorname{argmin}} \int_{\mathbb{V}_j} \|x-y\|^2 \mathrm{d}\alpha(x) = \frac{1}{|\mathbb{V}_j|}\int_{\mathbb{V}_i} x \mathrm{d}\alpha(x)$$

Lloyd's Algorithm

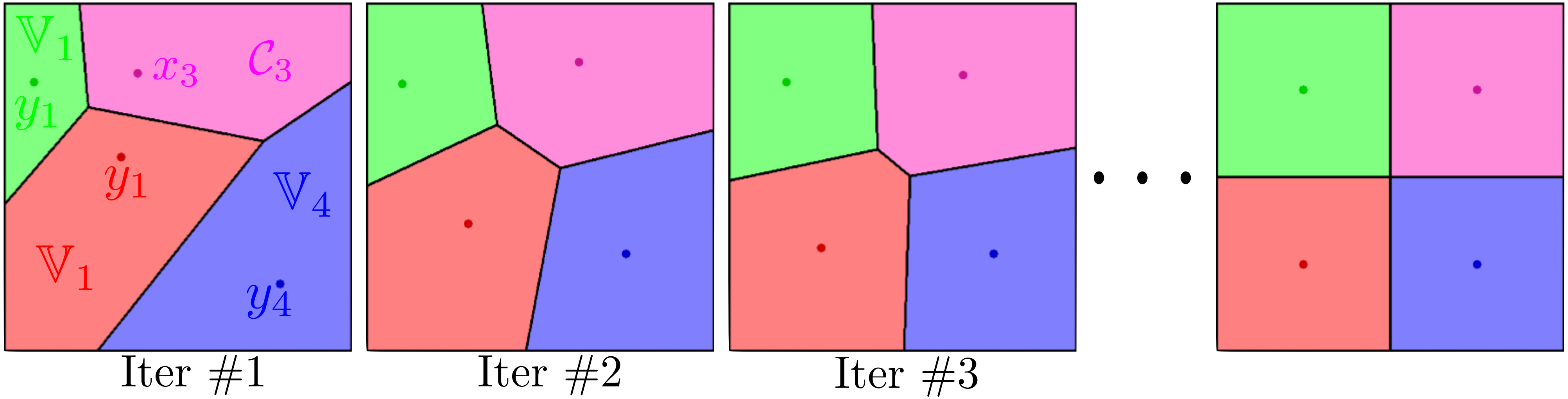
$$\text{Cost } c(x,y) = \|x-y\|^2$$

$$\underset{y}{\operatorname{argmin}} \int_{\mathbb{V}_j} \|x-y\|^2 \mathrm{d}\alpha(x) = \frac{1}{|\mathbb{V}_j|}\int_{\mathbb{V}_i} x\mathrm{d}\alpha(x)$$

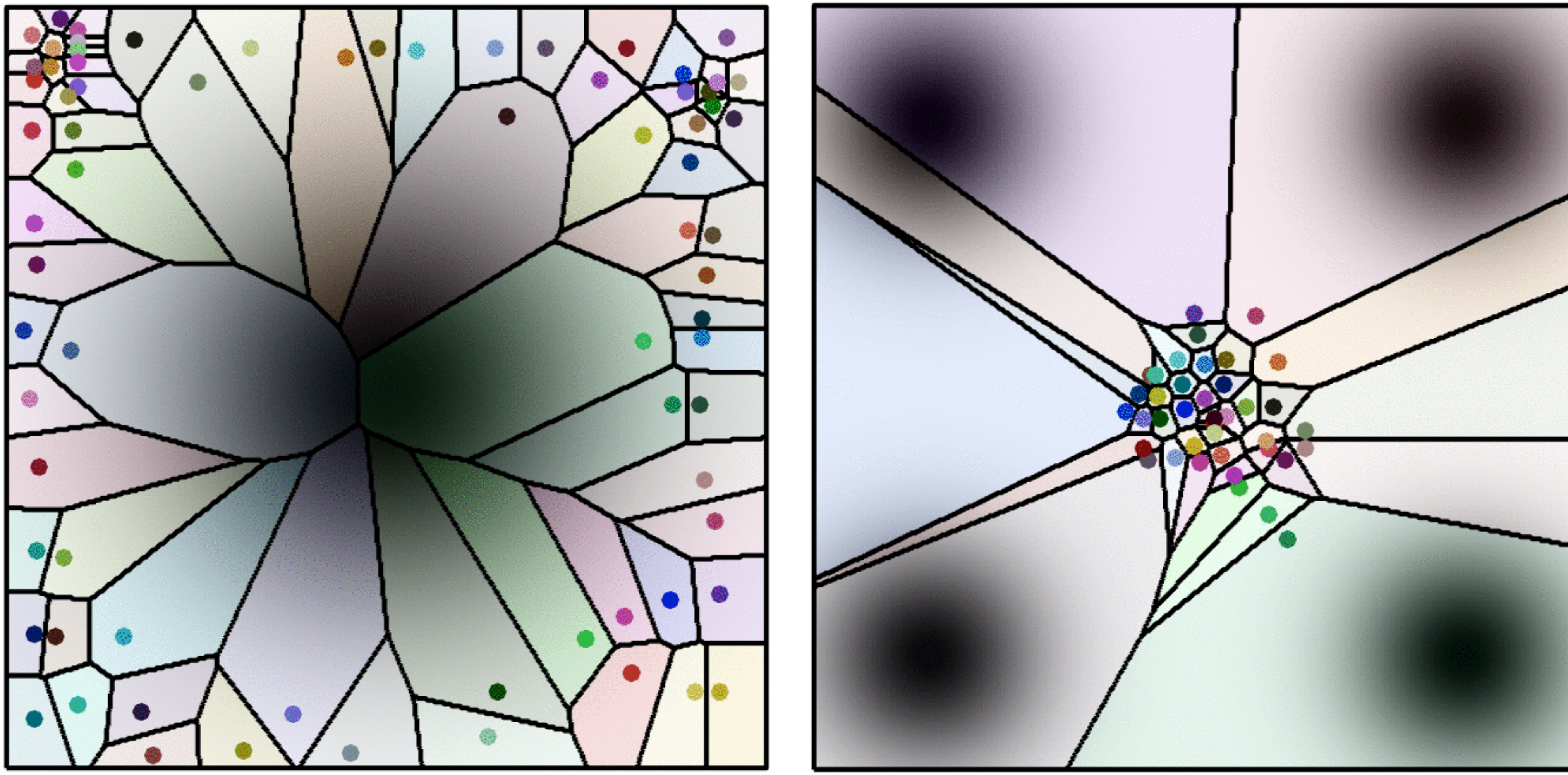
Lloyd algorithm:

$$\forall j, \quad \mathbb{V}_j \leftarrow \{x \; ; \; \forall \ell \neq j, \|x - y_j\| \leqslant \|x - \ell\|\}$$

$$\forall j, \quad y_j \leftarrow \frac{1}{|\mathbb{V}_j|}\int_{\mathbb{V}_i} x\mathrm{d}\alpha(x)$$



Lloyd’s Algorithm



Lloyd's Algorithm