Computational Optimal Transport

http://optimaltransport.github.io

Unbalanced OT

Gabriel Peyré

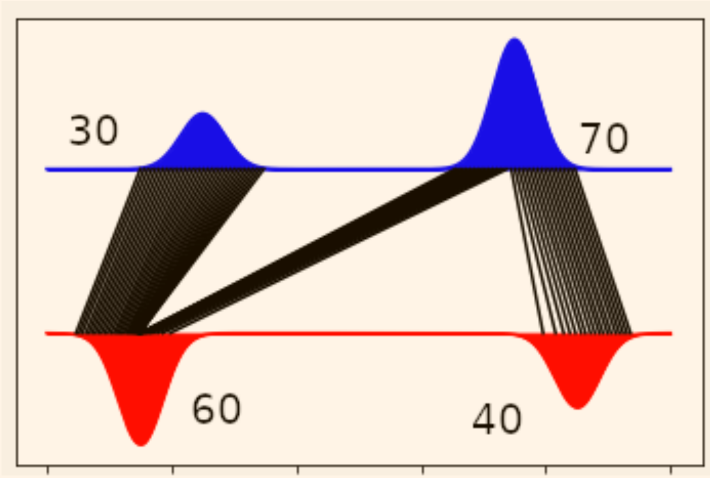
www.numerical-tours.com

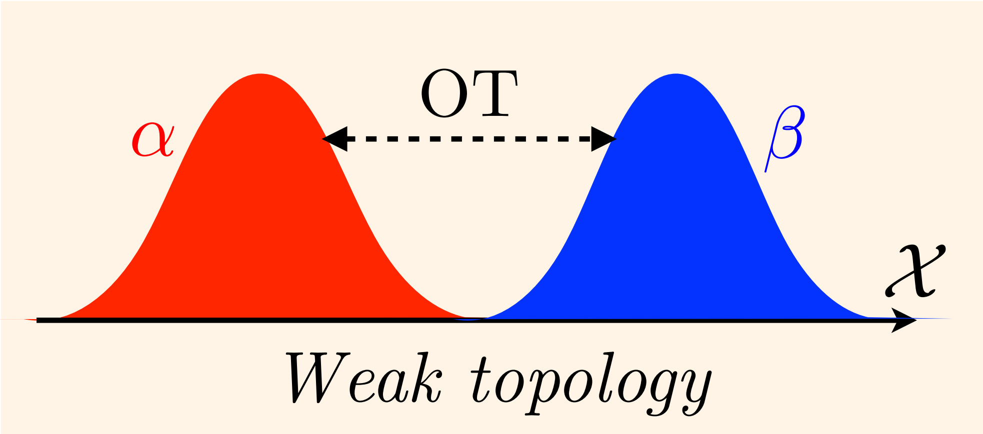
Total Variation Norm

"Balanced" OT: imposes $\alpha(\mathcal{X}) = \beta(\mathcal{X})$

Robust to support shift.

Not robust to $\rightarrow$ outliers, mode variations, missing parts



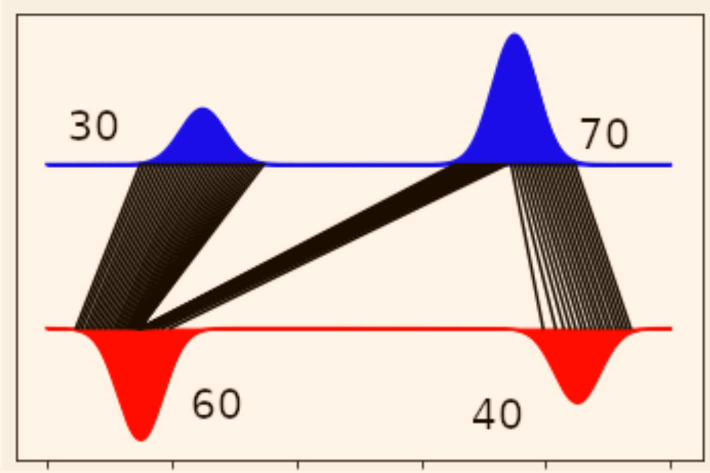


Total Variation Norm

"Balanced" OT: imposes $\alpha(\mathcal{X}) = \beta(\mathcal{X})$

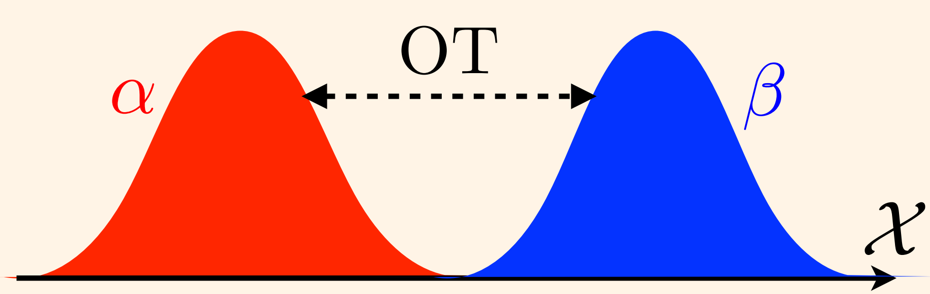
Robust to support shift.

Not robust to → outliers, mode variations, missing parts

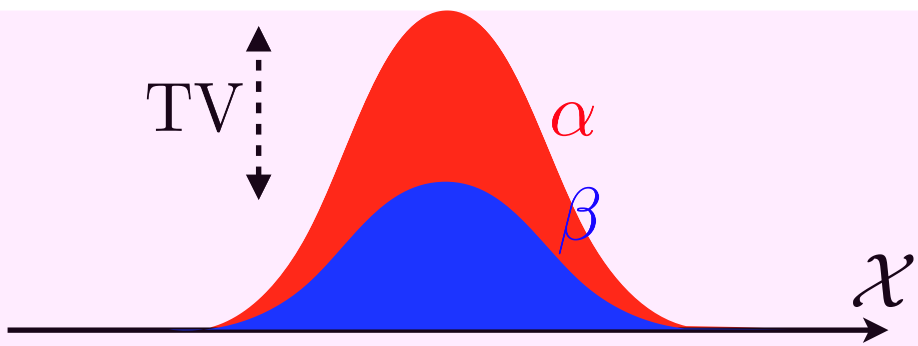


Total variation: $\|\alpha - \beta\|_{\mathrm{TV}} \triangleq \sup \left\{ \int f \mathrm{d}(\alpha - \beta) \;:\; \|f\|_\infty \leqslant 1 \right\}$

$$= |\alpha - \beta|(\mathcal{X}) = \left\| \frac{\mathrm{d}\alpha}{\mathrm{d}x} - \frac{\mathrm{d}\beta}{\mathrm{d}x} \right\|_{\mathrm{L}^1(\mathrm{d}x)}$$



Weak topology



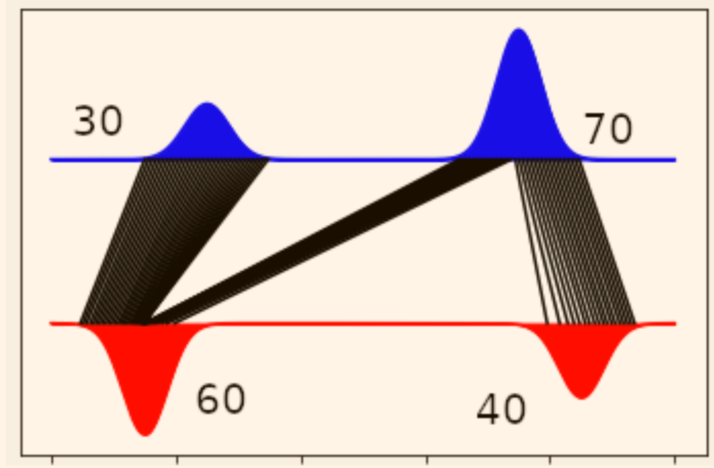
Strong topology

Total Variation Norm

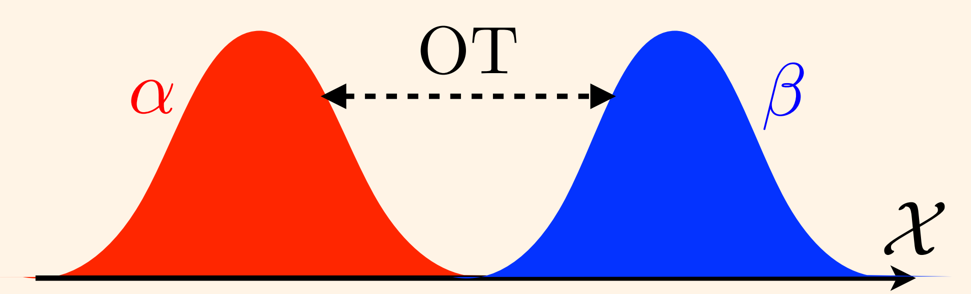
"Balanced" OT: imposes $\alpha(\mathcal{X}) = \beta(\mathcal{X})$

Robust to support shift.

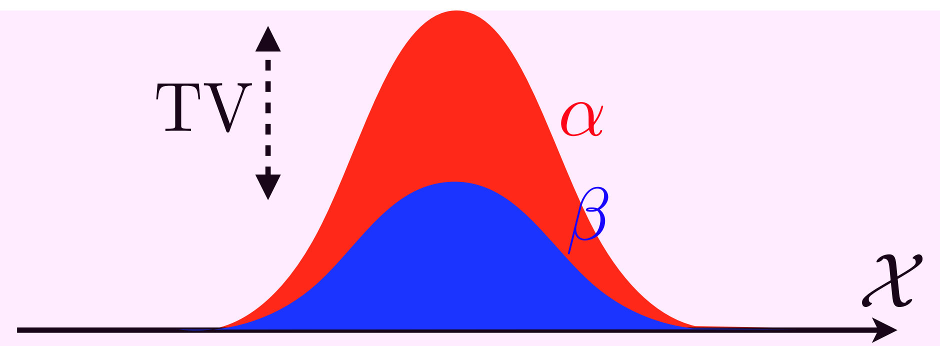
Not robust to
- outliers
- mode variations
- missing parts



$$\text{Total variation: } \|\alpha - \beta\|_{\mathrm{TV}} \triangleq \sup \left\{ \int f \mathrm{d}(\alpha - \beta) \;:\; \|f\|_\infty \leqslant 1 \right\}$$
$$= |\alpha - \beta|(\mathcal{X}) = \left\| \frac{\mathrm{d}\alpha}{\mathrm{d}x} - \frac{\mathrm{d}\beta}{\mathrm{d}x} \right\|_{\mathrm{L}^1(\mathrm{d}x)}$$



Weak topology



Strong topology

Unbalanced OT: hybridize the weak and strong topologies.

Csiszar Divergence

Comparing $$\frac{\mathrm{d}\alpha}{\mathrm{d}x} \leftrightarrow \frac{\mathrm{d}\beta}{\mathrm{d}x} \longrightarrow \frac{\mathrm{d}\alpha}{\mathrm{d}\beta} \leftrightarrow 1$$

Csiszar Divergence

Comparing $\qquad \dfrac{\mathrm{d}\alpha}{\mathrm{d}x} \leftrightarrow \dfrac{\mathrm{d}\beta}{\mathrm{d}x} \longrightarrow \dfrac{\mathrm{d}\alpha}{\mathrm{d}\beta} \leftrightarrow 1$

φ convex, $\varphi(1) = 0$, $\boxed{\varphi \geqslant 0} \longrightarrow$ Important if $\alpha(\mathcal{X}) \neq \beta(\mathcal{X})$.

Recession constant: $\quad \varphi'_\infty \triangleq \lim_{t \to +\infty} \dfrac{\varphi(t)}{t} \qquad \alpha = \dfrac{\mathrm{d}\alpha}{\mathrm{d}\beta} + \alpha^\perp$

Csiszár φ-divergence: $\mathrm{D}_\varphi(\alpha|\beta) \triangleq \int \varphi\left(\dfrac{\mathrm{d}\alpha}{\mathrm{d}\beta}(x)\right) \mathrm{d}\beta(x) + \varphi'_\infty \alpha^\perp(\mathcal{X})$

Csiszar Divergence

Comparing $\dfrac{\mathrm{d}\alpha}{\mathrm{d}x} \leftrightarrow \dfrac{\mathrm{d}\beta}{\mathrm{d}x} \longrightarrow \dfrac{\mathrm{d}\alpha}{\mathrm{d}\beta} \leftrightarrow 1$

φ convex, $\varphi(1) = 0$, $\boxed{\varphi \geqslant 0} \longrightarrow$ Important if $\alpha(\mathcal{X}) \neq \beta(\mathcal{X})$.

Recession constant: $\varphi'_\infty \triangleq \lim_{t \to +\infty} \dfrac{\varphi(t)}{t}$

$$\alpha = \frac{\mathrm{d}\alpha}{\mathrm{d}\beta} + \alpha^\perp$$

Csiszár φ-divergence: $\mathrm{D}_\varphi(\alpha|\beta) \triangleq \displaystyle\int \varphi\left(\frac{\mathrm{d}\alpha}{\mathrm{d}\beta}(x)\right) \mathrm{d}\beta(x) + \varphi'_\infty \alpha^\perp(\mathcal{X})$

Proposition: $\mathrm{D}_\varphi \geqslant 0$ is jointly convex,

$\mathrm{D}_\varphi(\alpha|\beta) = 0 \Leftrightarrow \alpha = \beta$

Example of Csiszar Divergence

$$\mathrm{D}_\varphi(\alpha|\beta) \triangleq \int \varphi\left(\frac{\mathrm{d}\alpha}{\mathrm{d}\beta}(x)\right)\mathrm{d}\beta(x) + \varphi'_\infty \alpha^\perp(\mathcal{X})$$

$$= \sum_{\mathbf{b}_i>0} \varphi\left(\frac{\mathbf{a}_i}{\mathbf{b}_i}\right)\mathbf{b}_i + \varphi'_\infty \sum_{\mathbf{b}_i=0} \mathbf{a}_i$$

Discrete measures:

$$\alpha = \sum_i \mathbf{a}_i \delta_{x_i}$$

$$\beta = \sum_i \mathbf{b}_i \delta_{x_i}$$

Example of Csiszar Divergence

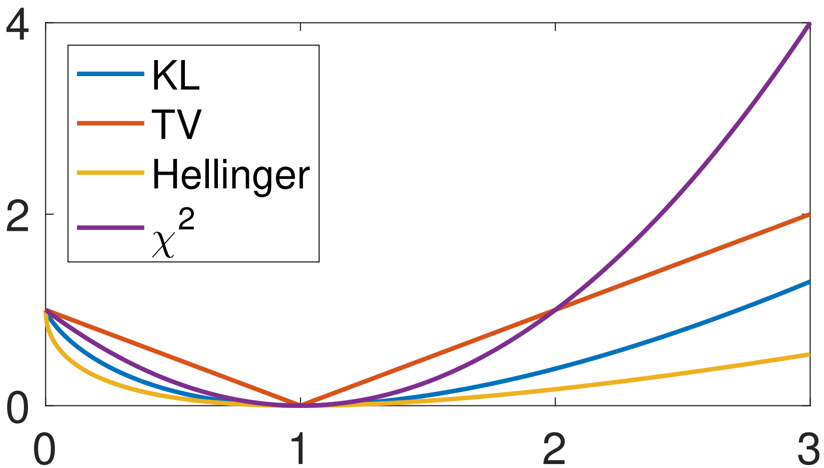
$$\mathrm{D}_\varphi(\alpha|\beta) \triangleq \int \varphi\left(\frac{\mathrm{d}\alpha}{\mathrm{d}\beta}(x)\right)\mathrm{d}\beta(x) + \varphi'_\infty \alpha^\perp(\mathcal{X})$$

$$= \sum_{\mathbf{b}_i>0} \varphi\left(\frac{\mathbf{a}_i}{\mathbf{b}_i}\right)\mathbf{b}_i + \varphi'_\infty \sum_{\mathbf{b}_i=0} \mathbf{a}_i$$

Discrete measures:

$$\alpha = \sum_i \mathbf{a}_i \delta_{x_i}$$

$$\beta = \sum_i \mathbf{b}_i \delta_{x_i}$$



$\lvert s-1\rvert^2$	χ^2	$\varphi'_\infty = +\infty$
$s\log(s)$	KL	
$s\log(s) - s + 1$	Generalized KL	

$\lvert s-1\rvert$	TV norm	$\varphi'_\infty = 1$
$\lvert\sqrt{s}-1\rvert^2$	Hellinger distance	

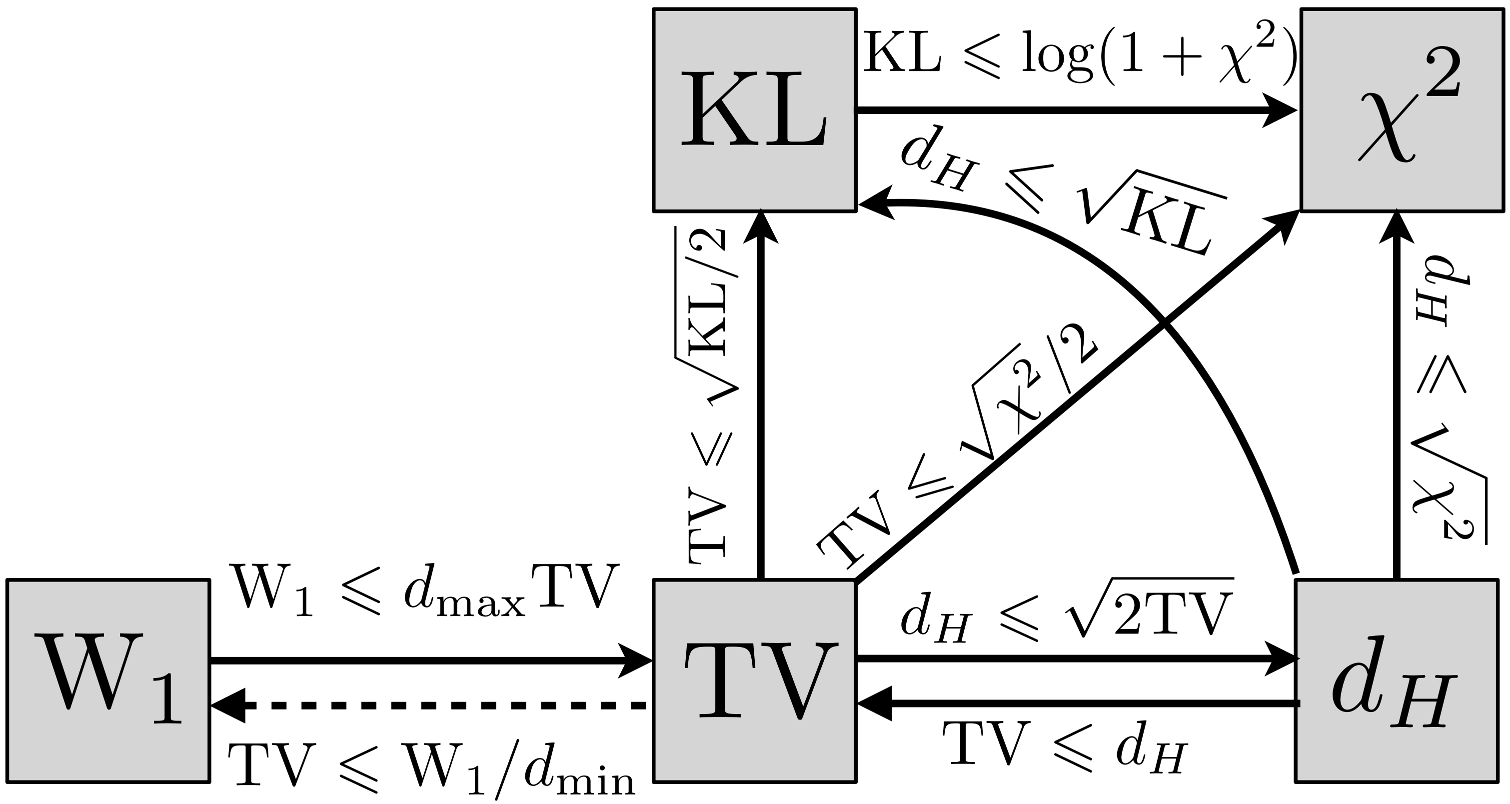
$$\|\alpha - \beta\|_{\mathrm{TV}} = \left\|\frac{\mathrm{d}\alpha}{\mathrm{d}x} - \frac{\mathrm{d}\beta}{\mathrm{d}x}\right\|_{\mathrm{L}^1(\mathrm{d}x)}$$

$$= \sum_i |\mathbf{a}_i - \mathbf{b}_i|$$

$$d_{\mathrm{H}}(\alpha,\beta)^2 = \left\|\sqrt{\frac{\mathrm{d}\alpha}{\mathrm{d}x}} - \sqrt{\frac{\mathrm{d}\beta}{\mathrm{d}x}}\right\|^2_{\mathrm{L}^2(\mathrm{d}x)}$$

$$= \sum_i (\sqrt{\mathbf{a}_i} - \sqrt{\mathbf{b}_i})^2$$

Equivalence and non-equivalence

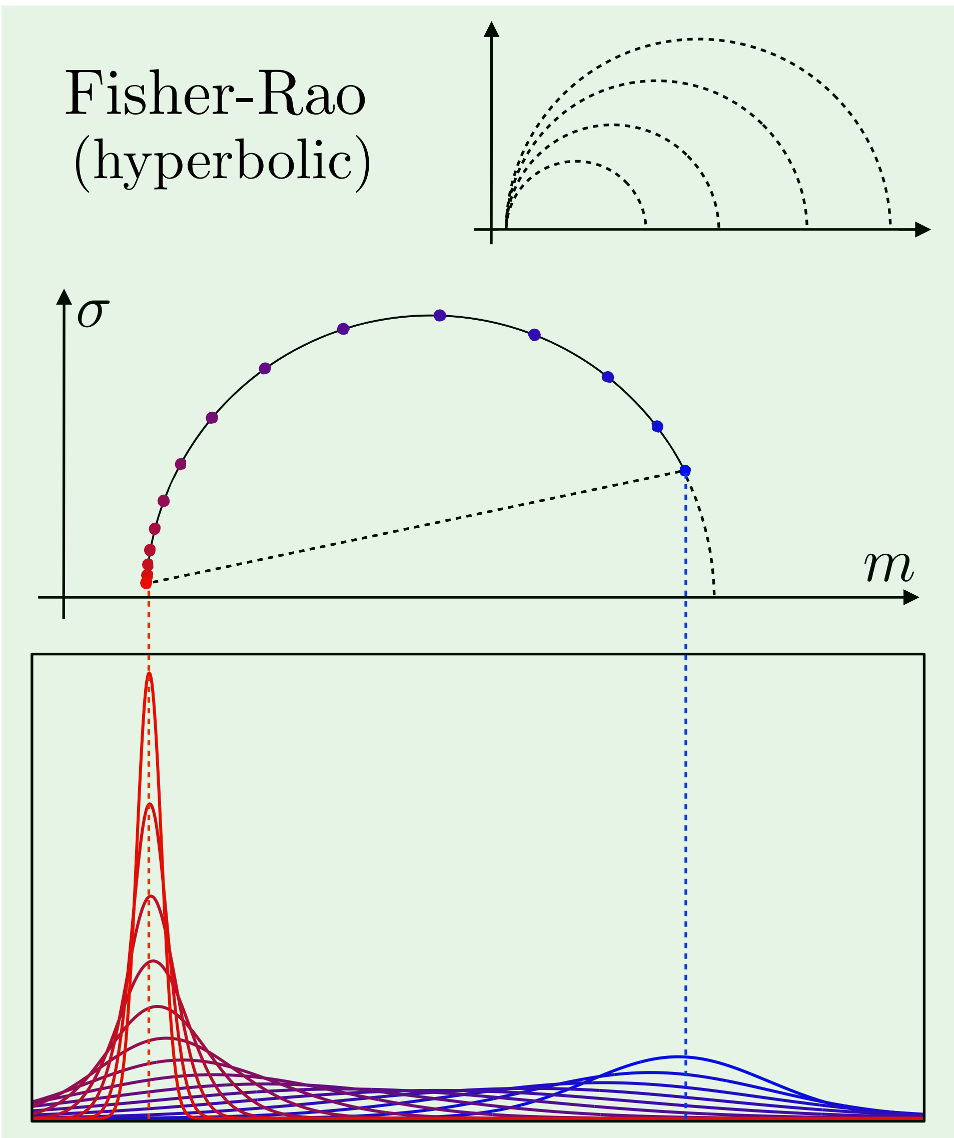


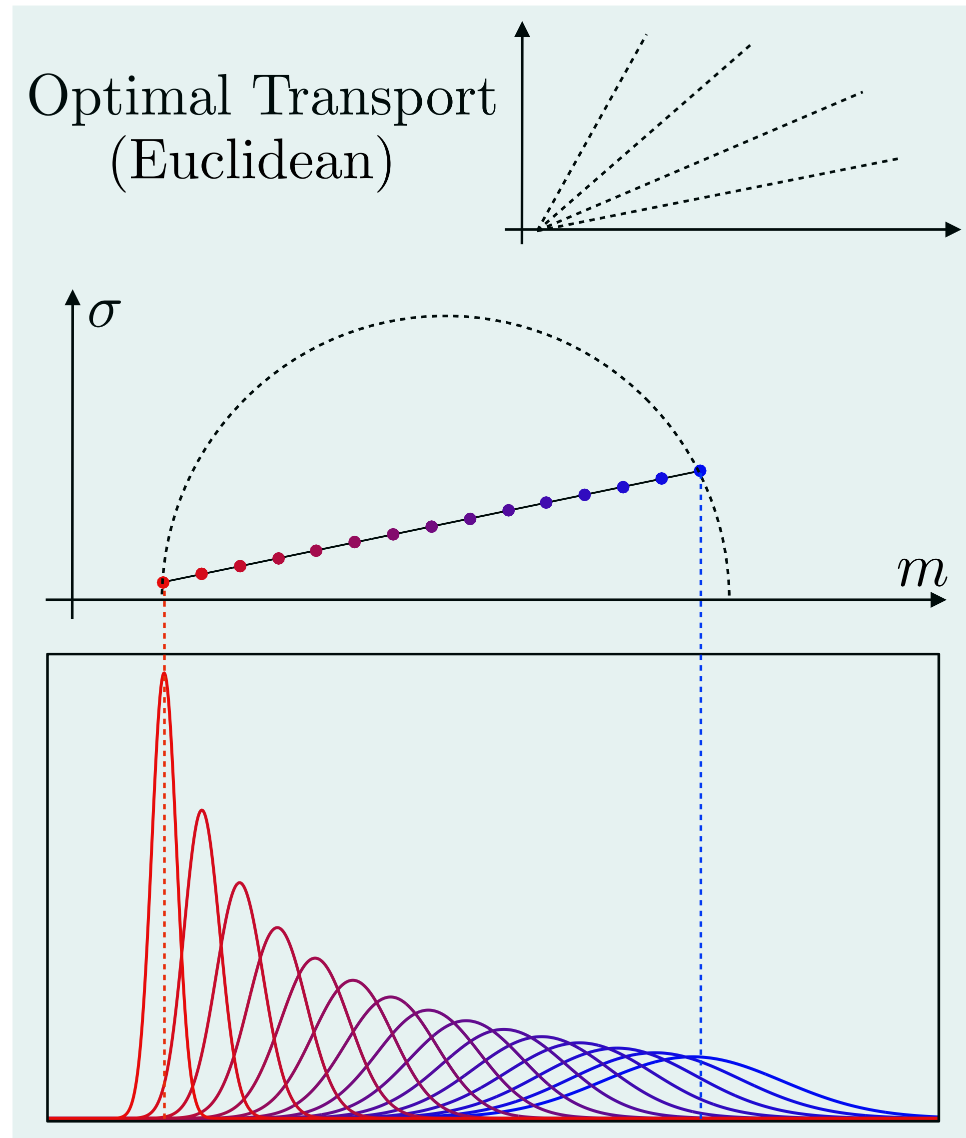
$$d_{\max} = \sup_{(x,x')} d(x,x') \qquad d_{\min} \overset{\text{def.}}{=} \min_{x\neq x'} d(x,x')$$

OT vs. KL (Fisher-Rao)

$$\mathcal{X} = \mathbb{R} \qquad \alpha = \mathcal{N}(m_\alpha, \sigma_\alpha), \quad \beta = \mathcal{N}(m_\beta, \sigma_\beta)$$

$$\mathrm{KL}(\alpha|\beta) = \frac{1}{2}\left(\frac{\sigma_\alpha^2}{\sigma_\beta^2} + \log\left(\frac{\sigma_\beta^2}{\sigma_\alpha^2}\right) + \frac{|m_\alpha - m_\beta|}{\sigma_\beta^2} - 1\right) \qquad \mathrm{W}_2^2(\alpha,\beta) = |m_\alpha - m_\beta|^2 + |\sigma_\alpha - \sigma_\beta|^2$$





Unbalanced OT

$$\mathrm{OT}_\rho(\alpha, \beta) \triangleq$$

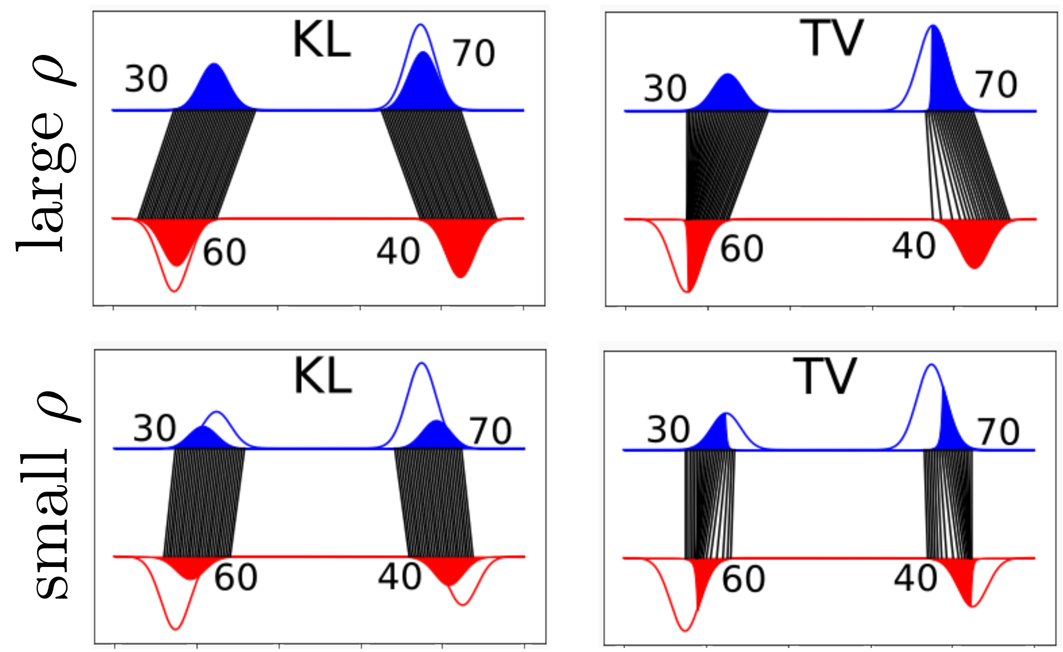
[Liero, Mielke, Savare, 2018]

$$\min_{\pi \geqslant 0} \left\{ \int c(x,y)\mathrm{d}\pi(x,y) + \rho \mathrm{D}_\varphi(\pi_1|\alpha) + \rho \mathrm{D}_\varphi(\pi_2|\alpha) \ ; \ \pi_1 = \alpha, \pi_2 = \beta \right\}$$

$\rho \approx$ tranportation radius

Balanced OT:

$\rho \to +\infty$ or $\varphi = \iota_{\{1\}}$



Unbalanced OT

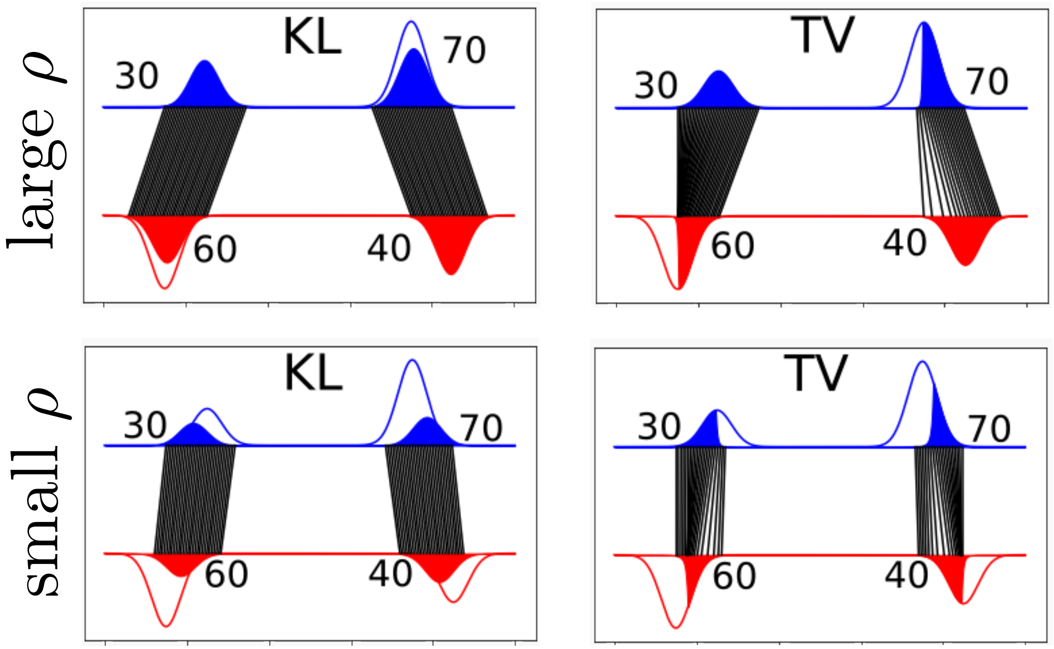
$\mathrm{OT}_\rho(\alpha, \beta) \triangleq$ [Liero, Mielke, Savare, 2018]

$$\min_{\pi \geqslant 0} \left\{ \int c(x,y)\mathrm{d}\pi(x,y) + \rho \mathrm{D}_\varphi(\pi_1|\alpha) + \rho \mathrm{D}_\varphi(\pi_2|\alpha) \; ; \; \pi_1 = \alpha, \pi_2 = \beta \right\}$$

$\rho \approx$ tranportation radius

Balanced OT:

$\rho \to +\infty$ or $\varphi = \iota_{\{1\}}$



$$\frac{1}{\rho}\mathrm{OT}_\rho(\alpha, \beta) \xrightarrow{\rho \to 0} \min_{\mu \geqslant 0} \mathrm{D}_\varphi(\mu|\alpha) + \mathrm{D}_\varphi(\mu|\beta)$$

$$\boxed{= \left\| \sqrt{\frac{\mathrm{d}\alpha}{\mathrm{d}\beta}} - 1 \right\|^2_{\mathrm{L}^2(\mathrm{d}\beta)}}$$

for KL

Unbalanced Sinkhorn

$$\min_{\pi_1=\alpha,\pi_2=\beta} \int c\,\mathrm{d}\pi + \mathrm{D}_\varphi(\pi_1|\alpha) + \mathrm{D}_\varphi(\pi_2|\alpha) + \varepsilon \mathrm{KL}(\pi|\alpha\otimes\beta)$$

$$\sup_{f,g} -\int \varphi^*(f(x))\mathrm{d}\alpha(x) - \int \varphi^*(g(x))\mathrm{d}\beta(x) - \varepsilon \int e^{\frac{f\oplus g - c}{\varepsilon}} \alpha\otimes\beta$$

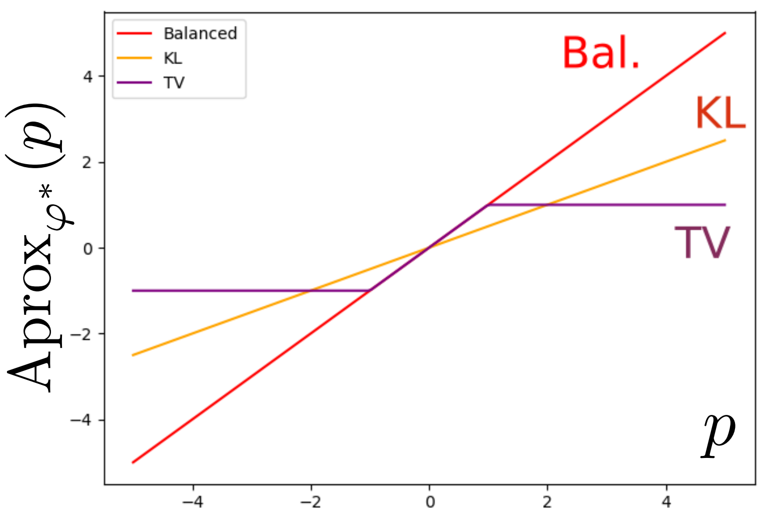
$$\pi = e^{\frac{f\oplus g-c}{\varepsilon}} \alpha\otimes\beta$$

Soft c-transforms:

$$f^{c,\varepsilon}(y) \triangleq -\varepsilon \log \int e^{\frac{c(x,y)-f(x)}{\varepsilon}} \mathrm{d}\alpha(x)$$

Anisotropic prox:

$$\mathrm{Aprox}_{\varphi^*}(p) \triangleq \underset{q\in\mathbb{R}}{\mathrm{argmin}}\ e^{p-q} + \varphi^*(q)$$



Proposition: Sinkhorn (alternate maximization)

$$f \leftarrow -\varepsilon\, \mathrm{Aprox}_{\frac{\varphi^*}{\varepsilon}}(-g^{c,\varepsilon}) \qquad g \leftarrow -\varepsilon\, \mathrm{Aprox}_{\frac{\varphi^*}{\varepsilon}}(-f^{c,\varepsilon})$$

Example: $\mathrm{D}_\varphi = \rho\mathrm{KL}$ $\qquad f \leftarrow \frac{\rho}{\varepsilon+\rho} g^{c,\varepsilon} \qquad g \leftarrow \frac{\rho}{\varepsilon+\rho} f^{c,\varepsilon}$

Distance Properties

$$\mathrm{UOT}(\alpha,\beta)^2 \triangleq \min_{\pi_1=\alpha,\pi_2=\beta} \int c(x,y)\mathrm{d}\pi(x,y) + \mathrm{KL}(\pi_1|\alpha) + \mathrm{KL}(\pi_2|\alpha)$$

Theorem: UOT is a distance on positive measure for

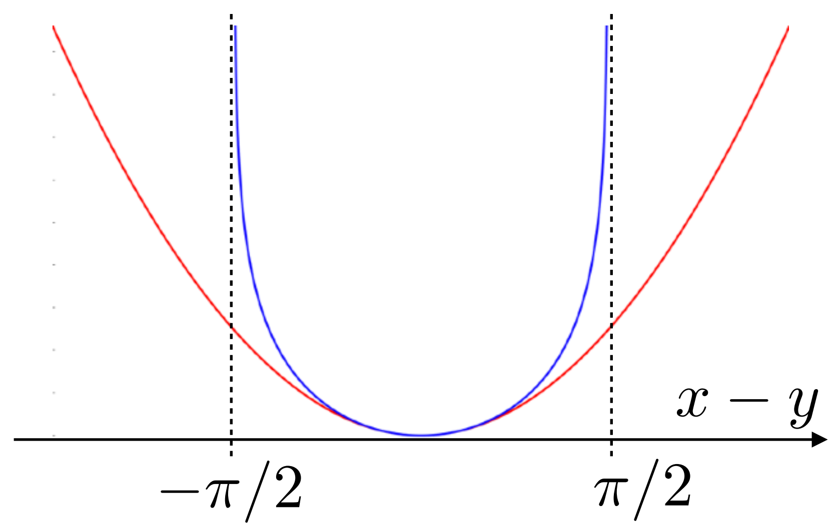
$$c(x,y) = d(x,y)^2$$ [Gaussian-Hellinger]

$$c(x,y) = -\log(\cos(\tfrac{\pi}{2} \wedge d(x,y)^2))$$ [Wasserstein-Fisher-Rao / Kantorovitch-Hellinger]

[Kondratyev, Monsaingeon, Vorotnikov 2015]

[Liereo, Mielke, Savaré 2015]

[Chizat, Schmitzer, Peyré, Vialard 2015]

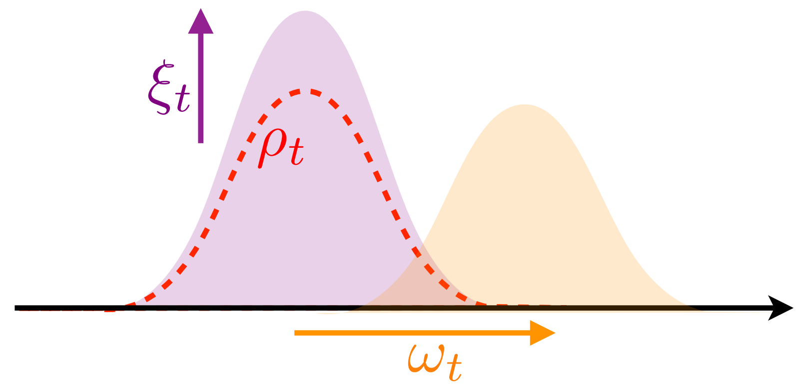


$|x-y|^2$ $\quad -\log(\cos(\frac{\pi}{2} \wedge |x-y|)^2)$

Dynamical Formulation

Wasserstein-Fisher-Rao /
Kantorovitch-Hellinger

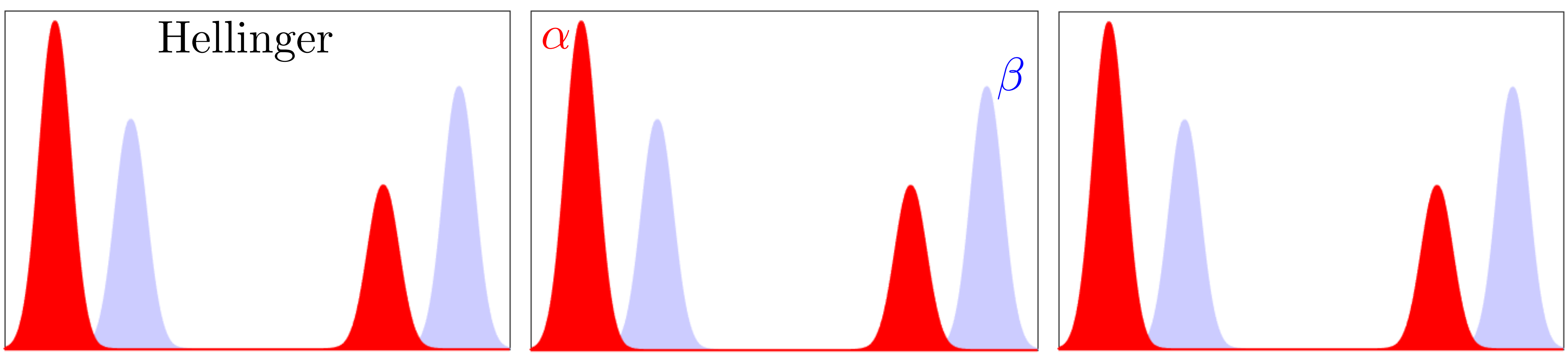
$$c(x,y) = -\log(\cos(\tfrac{1}{2\rho}\|x-y\| \wedge \tfrac{\pi}{2})^2)$$



Theorem:

$$\mathrm{UOT}(\alpha,\beta)^2 = \min_{\mu_0=\alpha,\mu_1=\beta} \left\{ \int_0^1 \int_{\mathbb{R}^d} \frac{\|\omega_t\|^2 + \rho\xi_t^2}{\mu_t} \mathrm{d}x\mathrm{d}t \;:\; \partial_t\mu_t + \mathrm{div}(\omega_t) = \xi_t \right\}$$

$$\int(\sqrt{\alpha}-\sqrt{\beta})^2 \xleftarrow{\;\rho\to 0\;} \text{Unbalanced} \xrightarrow{\;\rho\to+\infty\;} \text{Balanced}$$

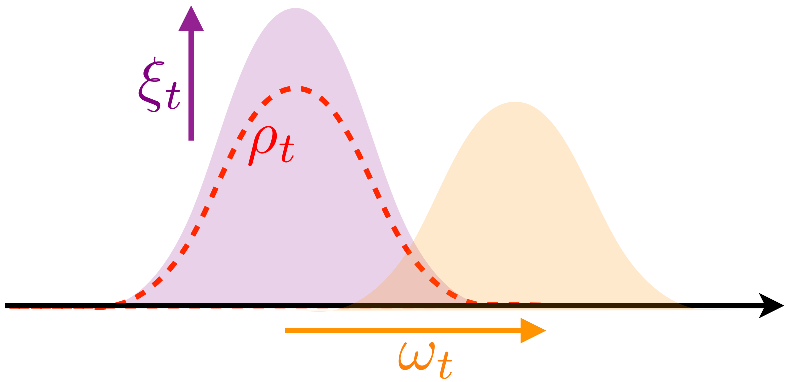


Dynamical Formulation

Wasserstein-Fisher-Rao /

Kantorovitch-Hellinger

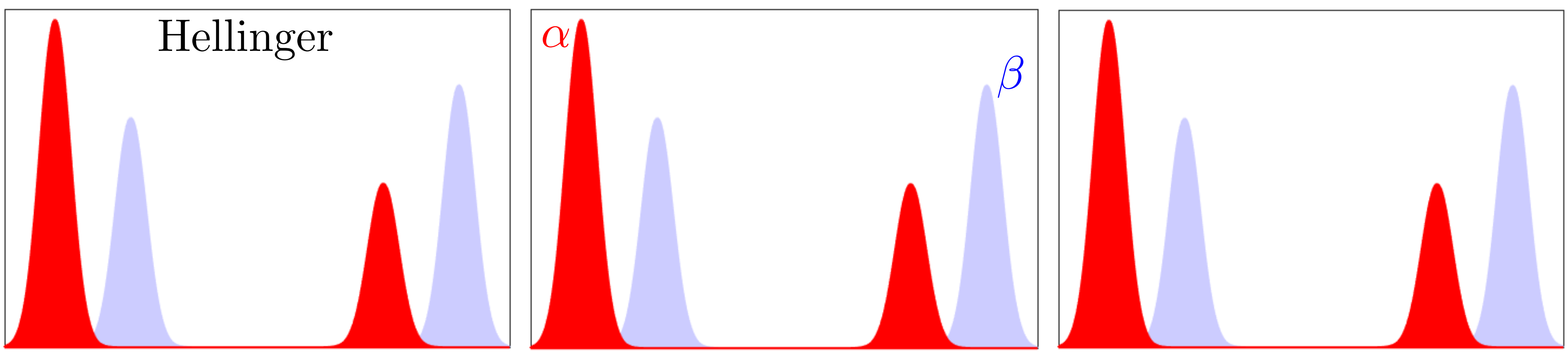
$$c(x,y) = -\log(\cos(\tfrac{1}{2\rho}\|x-y\| \wedge \tfrac{\pi}{2})^2)$$



Theorem:

$$\mathrm{UOT}(\alpha,\beta)^2 = \min_{\mu_0=\alpha,\mu_1=\beta} \left\{ \int_0^1 \int_{\mathbb{R}^d} \frac{\|\omega_t\|^2 + \rho\xi_t^2}{\mu_t} \mathrm{d}x\mathrm{d}t \;:\; \partial_t \mu_t + \mathrm{div}(\omega_t) = \xi_t \right\}$$

$$\int(\sqrt{\alpha} - \sqrt{\beta})^2 \xleftarrow{\rho \to 0} \text{Unbalanced} \xrightarrow{\rho \to +\infty} \text{Balanced}$$

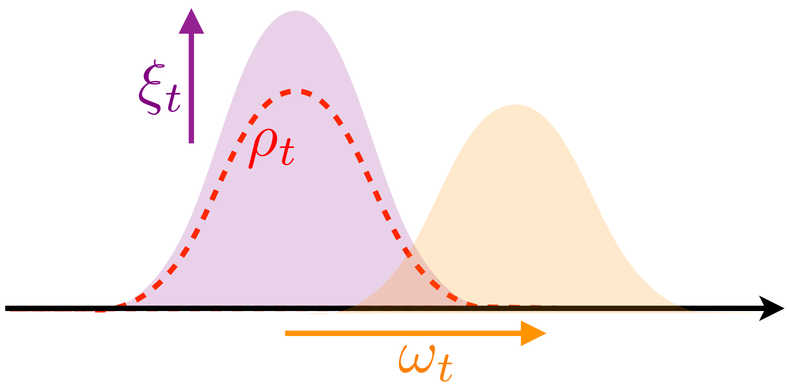


Dynamical Formulation

Wasserstein-Fisher-Rao /

Kantorovitch-Hellinger

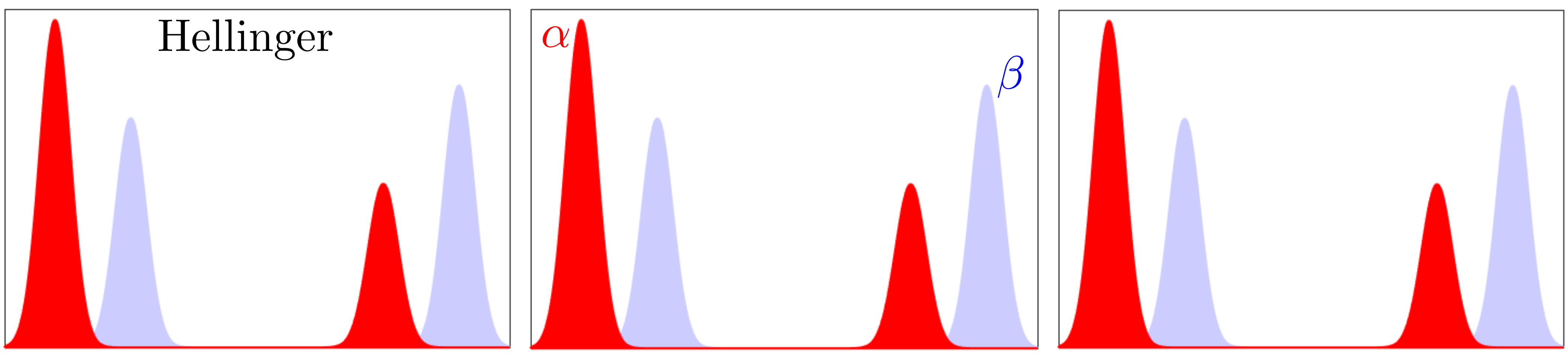
$$c(x,y) = -\log(\cos(\tfrac{1}{2\rho}\|x-y\| \wedge \tfrac{\pi}{2})^2)$$



Theorem:

$$\mathrm{UOT}(\alpha,\beta)^2 = \min_{\mu_0=\alpha,\mu_1=\beta} \left\{ \int_0^1 \int_{\mathbb{R}^d} \frac{\|\omega_t\|^2 + \rho\xi_t^2}{\mu_t} \mathrm{d}x\mathrm{d}t \;:\; \partial_t\mu_t + \mathrm{div}(\omega_t) = \xi_t \right\}$$

$$\int(\sqrt{\alpha}-\sqrt{\beta})^2 \xleftarrow{\rho \to 0} \text{Unbalanced} \xrightarrow{\rho \to +\infty} \text{Balanced}$$

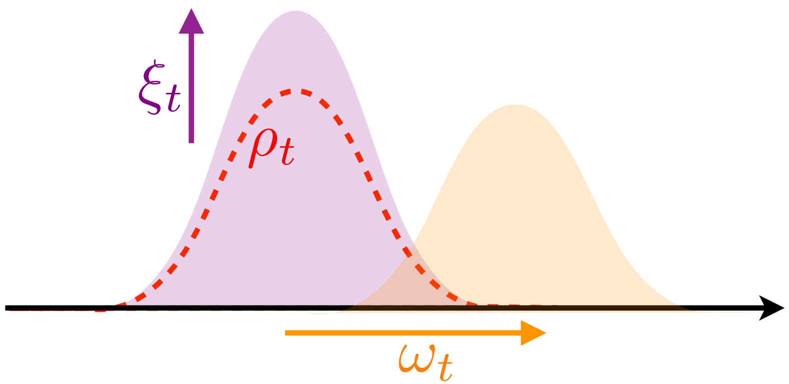


Dynamical Formulation

Wasserstein-Fisher-Rao /

Kantorovitch-Hellinger

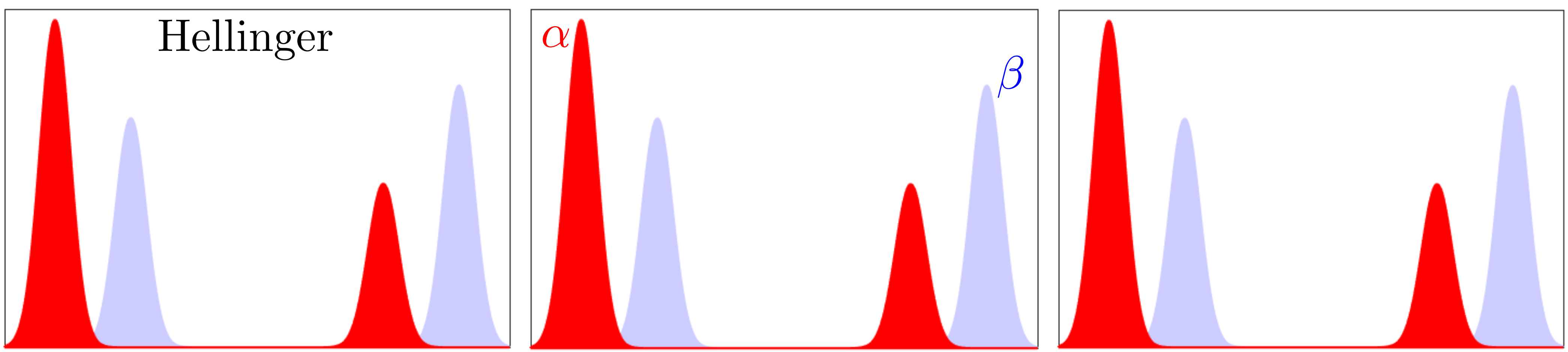
$$c(x,y) = -\log(\cos(\tfrac{1}{2\rho}\|x-y\| \wedge \tfrac{\pi}{2})^2)$$



Theorem:

$$\mathrm{UOT}(\alpha,\beta)^2 = \min_{\mu_0=\alpha,\mu_1=\beta} \left\{ \int_0^1 \int_{\mathbb{R}^d} \frac{\|\omega_t\|^2 + \rho\xi_t^2}{\mu_t} \mathrm{d}x\mathrm{d}t \;:\; \partial_t\mu_t + \mathrm{div}(\omega_t) = \xi_t \right\}$$

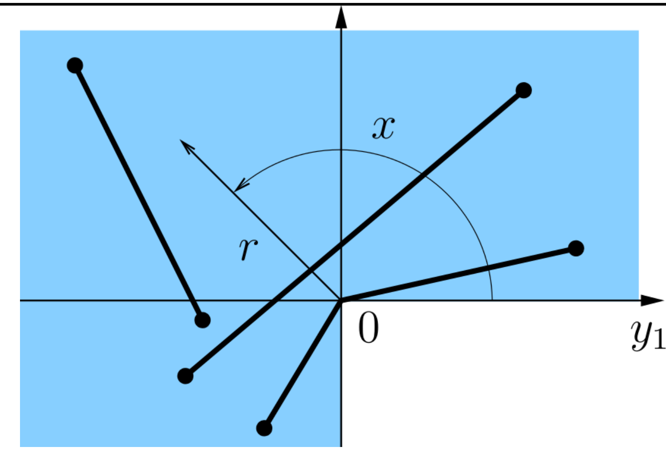
$\int(\sqrt{\alpha}-\sqrt{\beta})^2 \xleftarrow{\rho \to 0}$ Unbalanced $\xrightarrow{\rho \to +\infty}$ Balanced



Conic Formulation

Conification of $\mathcal{X}$: $(x,r) \in \mathfrak{C}[\mathcal{X}] \triangleq \mathcal{X} \times \mathbb{R}^+$.

Cone distance: $d_{\mathfrak{C}[\mathcal{X}]}((x,r),(y,s))^2$
$$\triangleq r^2 + s^2 - 2rs\cos(\pi \wedge d_{\mathcal{X}}(x,y)))$$



$$\sum_{i=1}^n r_i^2 \delta_{x_i} \in \mathcal{M}_+(\mathcal{X}) \xrightarrow{\text{Lifting}} \frac{1}{n}\sum_i \delta_{(r_i,x_i)} \in \mathcal{M}_+^1(\mathfrak{C}[\mathcal{X}])$$

$$\int_{\mathbb{R}^+} r^2 \mathrm{d}\gamma(\cdot,r) \in \mathcal{M}_+(\mathcal{X}) \xleftarrow{\text{Projection}} \gamma \in \mathcal{M}_+^1(\mathfrak{C}[\mathcal{X}])$$

Theorem: $\mathrm{UOT}(\alpha,\beta) = \inf_{\eta \in \mathcal{M}_+^1(\mathfrak{C}[\mathcal{X}]^2)} \left\{ \int_{\mathfrak{C}[\mathcal{X}]^2} d^2_{\mathfrak{C}[\mathcal{X}]} \mathrm{d}\eta : \begin{array}{l} \int r^2 \mathrm{d}\eta_1(\cdot,r) = \alpha \\ \int s^2 \mathrm{d}\eta_2(\cdot,s) = \beta \end{array} \right\}$