OT and duality

Legendre-Fenchel transform: $F : \mathbb{R}^n \to \mathbb{R}$ $\quad F^*(f) \triangleq \sup_{\alpha} \langle \alpha, f \rangle - F(\alpha)$, $f \in \mathbb{R}^n$

F^* convex, $(F^{**}) = F \Leftrightarrow F$ convex

NB : on $\mathcal{M}(X)$, $\forall f \in C(X)$, $F^*(f) \triangleq \sup_{\alpha \in \mathcal{M}(X)} \underbrace{\int f \, d\alpha}_{\langle f, \alpha \rangle} - F(\alpha)$

Dual of OT : $\min_{P \geq 0} \{ \langle P, C \rangle : P1 = a, \; P^T 1 = b \} = W_c(a,b)$

$= \min_{P \geq 0} \max_{(f,g) \in \mathbb{R}^n \times \mathbb{R}^n} \langle P, C \rangle + \langle f, a - P1 \rangle + \langle g, b - P^T 1 \rangle$

Prog lin $\to$ dualité forte si contraintes $\neq \emptyset$ (ici $P = ab^T$)

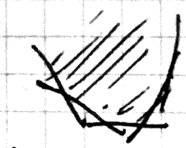
$= \max_{(f,g)} \langle f, a \rangle + \langle g, b \rangle + \boxed{\inf_{P \geq 0} \langle P, C - \underbrace{f 1^T - 1 g^T}_{-f \oplus g} \rangle} = \begin{cases} 0 \text{ si } C - f \oplus g \geq 0 \\ +\infty \text{ sinon} \end{cases}$

$= \max_{(f,g)} \{ \langle f, a \rangle + \langle g, b \rangle : f \oplus g \leq C \}$.

Sur $\mathcal{M}^+(X)$: $(f,g) \in C(X)^2$, $f(x) + g(y) \leq C(x,y)$

Rel° d'extrémalité : $\mathrm{supp}(P) \subset \{ (i,j) \; ; \; f_i + g_j = C_{ij} \}$.

Convexity : W_c is a max of linear form $\to$ Convex in (a,b)



c-transform : $W_c(\alpha, \beta) = \sup_{f,g} \{ D(f,g) \triangleq \langle f, \alpha \rangle + \langle g, \beta \rangle + \iota_{\mathcal{D}_c}(f,g) \}$ $\quad \mathcal{D}_c = \{(f,g) ; f \oplus g \leq c\}$.

"$f^c \triangleq \mathrm{argmax}_g \, D(f,g)$" $\to$ extended out of $\mathrm{supp}(\alpha)$

$g(y) \leq c(x,y) - f(x)$

$\Downarrow$

bc α is ≥ 0

$g(y) \leq \inf_x c(x,y) - f(x)$

$g(y) = \inf_x c(x,y) - f(x)$ $\hookrightarrow$ on the support of α

$\begin{cases} f^c(y) = \inf_x c(x,y) - f(x) \\ g^c(x) = \inf_y c(x,y) - g(y) \end{cases}$

$D(f,g) \leq D(f, f^c) \leq D(f^{cc}, f^c) = D(f^{cc}, f^{ccc})$

stop !

Prop° : $(f^c)^{cc} = f^c$ ie if g is a c-concave f°, $g^{cc} = g$

Rmq : if $c(x,y) = -\langle x, y \rangle$, then $f^c(y) = -\sup_x \langle x, y \rangle + f(x) = -(-f)^*$ (ie legendre for concave)

Special W1 : cf Book $\begin{cases} \to \text{Dual } f \text{: Lipsch} \\ \to \text{Div constraint} \\ \to O(n^2) \text{ network simplex} \end{cases}$

Dual norms : $\|\alpha\|_B \triangleq \sup_{f \in B} \langle f, \alpha \rangle$

ex: $B = \{f : \|f\|_\infty \leq 1\} \to$ TV norm

$B = \{f ; \|\nabla f\|_\infty \leq 1\} \to W_1$

$B = \{f : \|\nabla f\|_{L^2} \leq 1\} \to H^{-1}$

Prop : $\|.\|_B$ metrizes weak* conv $\Leftrightarrow$ $\overline{\text{Span}(B)} = \mathcal{C}(\mathcal{X})$ $\begin{cases} \to \text{Not ok for TV} \\ \to \text{ok for } W^1 \\ \to \text{ok for } H^{-1} \text{ in } d^\circ \leq \end{cases}$

Proof :

Semi-Discrete OT : $W_c(\alpha,\beta) = \sup_f \int f d\alpha + \int f^c d\beta = \mathbb{E}_{\alpha \otimes \beta}[f \oplus f^c]$

suppose $\alpha = \sum_{i=1}^n a_i \delta_{x_i} \rightsquigarrow f \in \mathbb{R}^n$, $f^c(y) = \min_{1 \leq i \leq n} c(x_i, y) - f_i \in \mathcal{C}(\mathcal{X})$

$W_c(\alpha,\beta) = \sup_{f \in \mathbb{R}^n} \sum_i a_i f_i + \underbrace{\int f^c(y) d\beta}_{\mathbb{E}_\beta(f^c)}$.

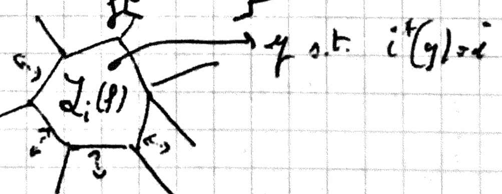
Stochastic GD : $y \sim \beta$, $f \leftarrow f - \tau_k [a + \nabla_f f^c(y)]$

Laguerre Cells
$i^*(y) = \operatorname{argmax}_i c(x_i, y) - f_i$

$\mathcal{L}_i(f) = \{y : c(x_i, y) - f_i \leq c(x_j, y) - f_j\}$ $\forall j$

$\mathcal{X} = \bigcup_i \mathcal{L}_i(f)$: for $c(x,y) = |x-y|^2 \to$ Polyhedral, $f = 0 \to$ Voronoi

Fast Algo



$\nabla_f [f^c(y)] = -\delta_{i^*(y)} \to O(n)$ to compute

Full GD : $\nabla_f \mathbb{E}_\beta(f^c) = \left(-\int_{\mathcal{L}_i(f)} d\beta\right)_{i=1}^n \rightsquigarrow$ need to compute Laguerre Diagr

③

Entropic OT: $\inf_{P\geq 0} \{ \langle P, c\rangle + \varepsilon KL(P|a\otimes b) : P\mathbb{1} = a, P^T\mathbb{1} = b\} \triangleq \overline{W}_c^\varepsilon(a,b)$

$$= \sup_{(f,g)} \langle f,a\rangle + \langle g,b\rangle + \inf_{P\geq 0} \langle P, c - f\oplus g\rangle + \varepsilon KL(P|a\otimes b)$$

$$\cdots \quad - \varepsilon \sup_{P\geq 0} \langle P, \frac{f\oplus g - c}{\varepsilon}\rangle - KL(P|a\otimes b)$$

$$\cdots \quad - \varepsilon\, KL^*\left(\frac{f\oplus g - c}{\varepsilon} \,\Big|\, a\otimes b\right)$$

Prop: $D_\varphi^*(u|a) = \sum_i \varphi^*(u_i)\, a_i$

$$\varphi(x) = x(\log(x) - 1) \rightsquigarrow \varphi^*(s) = e^s - 1$$

Cond°: $W_c^\varepsilon(a,b) = \sup_{(f,g)} \langle f,a\rangle + \langle g,b\rangle - \varepsilon \underbrace{\langle \exp\left(\frac{f\oplus g - c}{\varepsilon}\right) - 1, a\otimes b\rangle}$

"Soft $f\oplus g \leq c$" $\xrightarrow{\varepsilon\to 0}$ $f\oplus g \leq c$ hard

Soft c-transform: f fixed, min/g:

$$f^{c,\varepsilon}(y) \triangleq \mathrm{SoftMin}_{\alpha,\varepsilon}\left[c(\cdot,y) - f(\cdot)\right] \qquad \mathrm{SoftMin}_{\alpha,\varepsilon}(u) = -\varepsilon \log \int_X \exp\left(\frac{-u(x)}{\varepsilon}\right) d\alpha(x)$$

$\downarrow \varepsilon \to 0$

min on sup of α

Stabilizat°: LSE-trick $\qquad \mathrm{SoftMin}(u - c) = \mathrm{SoftMin}(u) - c$, $c \approx \min(u)$

Sinkhorn: $\to g \leftarrow f^{c,\varepsilon},\ f \leftarrow g^{c,\varepsilon}$ $\Leftrightarrow$ Sinkh on $(u,v) = (e^{f/\varepsilon}, e^{g/\varepsilon})$!!

Pbm: $W_c^\varepsilon(a,a) \neq 0$!

Debiased Sinkh. divergence: $\overline{W}_c^\varepsilon(a,b) = W_c^\varepsilon(a,b) - W_c^\varepsilon(a,a)/2 - W_c^\varepsilon(b,b)/2$

Thm: $W_c^\varepsilon \xrightarrow{\varepsilon\to 0} W_c$, $W_c^\varepsilon(\alpha,\beta) \xrightarrow{\varepsilon\to\infty} \frac{1}{2}\|\alpha - \beta\|^2_{-c}$

Max Mean Discr. (MMD): $\|\xi\|_k^2 = \iint k(x,y)\, d\xi(x)\, d\xi(y)$

if $k(x,y)$ positive kernel, ie, $(k(x_i, y_j))_{i,j} \succ 0$ on $\mathbb{1}^\perp$ then $\|.\|_k \geq 0$ and it is a norm

condit.

NB: $\|\alpha - \beta\|_k^2 = \iint k(x,x')\, d\alpha(x)\, d\alpha(x') + \iint k\, d\beta\, d\beta - 2\iint k\, d\alpha\, d\beta$

$\left[\alpha = \sum a_i \delta_{x_i},\ \beta = \sum b_j \delta_{y_j}\right] = \sum_{i,i'} a_i a_{i'} k(x_i, x_{i'}) + \sum_{j,j'} b_j b_{j'} k(y_j, y_{j'}) - 2\sum_{i,j} a_i b_j k(x_i, y_j) \Rightarrow O(n^2)$

Prop: $\|.\|_k$ metrize weak conv $\Leftrightarrow$ Span$(k(x,\cdot))$ dense in $\mathcal{C}(X)$

Ex: $\exp\left(-\frac{\|x-y\|^p}{\varepsilon}\right)$ $0<p\leq 2$; $-\|x-y\|^p$ $0<p<2$

!!

Proof of: $W_c^\varepsilon(\alpha,\beta) \xrightarrow{\varepsilon\to+\infty} \int c(x,y)\, d\alpha\, d\beta$

At optimum on f: $f_i = -\varepsilon \log \sum_j \exp\left(\frac{-C_{ij}+g_j}{\varepsilon}\right) b_j$

$$\Rightarrow 1 = \sum_j \exp\left(\frac{-C_{ij}+f_i+g_j}{\varepsilon}\right) b_j$$

$$\xRightarrow{\sum a_i \times} 1 = \left\langle \exp\left(\frac{f\oplus g - C}{\varepsilon}\right), a\otimes b\right\rangle$$

$$\Rightarrow \left\langle \exp\left(\frac{-C+f\oplus g}{\varepsilon}\right) - 1, a\otimes b\right\rangle$$

Hence $W_c^\varepsilon(\alpha,\beta) = \langle \underbrace{f^*_\varepsilon}_{\text{optimal}}, a\rangle + \langle \underbrace{g^*_\varepsilon}_{\text{optimal}}, b\rangle$

Primal limit: $\min_{\substack{P\geq 0\\ P1=a,\ P^T1=b}} \langle C,P\rangle + \varepsilon KL(P|ab^T) \xrightarrow{\varepsilon\to\infty} \min_{P\geq 0} KL(P|ab^T)$ $\boxed{P1=a,\ P^T1=b} \to$ useless

DL of ($\varepsilon\to+\infty$): $f_i = \mathrm{SoftMin}_{\varepsilon,j}(C_{ij} - g_j) \longrightarrow f = Cb - \langle b, g\rangle + O(\frac{1}{\varepsilon})$

$g_j = \dots \longrightarrow g = Ca - \langle a, f\rangle + O(1/\varepsilon)$

WLOG: $\langle g,b\rangle = 0$ so that $\begin{cases} f_\varepsilon \to Cb \\ g_\varepsilon \to Ca - \langle a, Cb\rangle \end{cases}$

In fine $W_c^\varepsilon(a,b) = \langle f_\varepsilon,a\rangle + \langle g_\varepsilon,b\rangle \to \langle Cb,a\rangle + \langle Ca,b\rangle - \langle a,Cb\rangle$
$= \langle Cb,a\rangle$ ■

Proof of positivity: if $k_\varepsilon \triangleq e^{-C/\varepsilon}$ is a positive kernel, $\overline{W}_c^\varepsilon(a,b)\geq 0$

• One has: $\overline{W}_c^\varepsilon(\alpha,\alpha) = \max_f 2\langle f,\alpha\rangle - \varepsilon\left\langle \exp\left(\frac{f\oplus f - C}{\varepsilon}\right) - 1, \alpha\otimes\alpha\right\rangle$
$= 2\langle f_{\alpha\alpha},\alpha\rangle$

let $\begin{cases} (f_{\alpha\beta}, g_{\alpha\beta}) \text{ optimum for } W_c^\varepsilon(\alpha,\beta) \\ (f_{\alpha\alpha}, f_{\alpha\alpha}) \quad \dots \quad W_c^\varepsilon(\alpha,\alpha) \\ (g_{\beta\beta}, g_{\beta\beta}) \quad \dots \quad W_c^\varepsilon(\beta,\beta) \end{cases}$

Then by optimality plugging $(f_{\alpha\alpha}, g_{\beta\beta})$ in dual pb:

$$W_c^\varepsilon(\alpha,\beta) \geq \underbrace{\langle\alpha,f_{\alpha\alpha}\rangle + \langle\beta,g_{\beta\beta}\rangle}_{=\frac12 W_c^\varepsilon(\alpha,\alpha) + \frac12 W_c^\varepsilon(\beta,\beta)} + \varepsilon\left\langle\alpha\otimes\beta, e^{\frac{f_{\alpha\alpha}\oplus g_{\beta\beta}-C}{\varepsilon}} - 1\right\rangle$$

Denoting $F \triangleq e^{f_{\alpha\alpha}/\varepsilon}$, $G\triangleq e^{g_{\beta\beta}/\varepsilon}$, $\Rightarrow \overline{W}_c^\varepsilon(\alpha,\beta) \geq \varepsilon(1 - \langle\alpha F, \beta G\rangle_{k_\varepsilon})$ ⊛

Optim. cond reads: $F\odot[k_\varepsilon(\alpha 1)] = 1 \Rightarrow \|F\alpha\|^2_{k_\varepsilon} = \langle\alpha F, k_\varepsilon(\alpha F)\rangle = \langle\alpha, F\odot k_\varepsilon(\alpha F)\rangle = \langle\alpha,1\rangle = 1$

Same $\|G\beta\|^2_{k_\varepsilon} = 1$, hence $\overline{W}_c^\varepsilon(\alpha,\beta) \geq \frac{\varepsilon}{2}\|\alpha F - \beta G\|^2_{k_\varepsilon} \geq 0$