Additional Proofs

Appendix EC.1: Proofs for Appendix A

Proof of Lemma 1. The monotonicity of $\mathcal{L}(\rho)$ can be seen from the definition. Moreover, since $\mathcal{K}_c(\widehat{\mathbb{P}},\widehat{\mathbb{P}})=0$,

$$\mathcal{L}(\rho)\geq\mathcal{L}(0)\geq\mathbb{E}_{X\sim\widehat{\mathbb{P}}}[f(X)]>-\infty.$$

Therefore for all $\rho\geq 0$, $\mathcal{L}(\rho)$ is bounded from below. To verify the concavity, fix $\rho_0,\rho_1\geq 0$. Pick any $t\in[0,1]$ and $\mathbb{P}_0,\mathbb{P}_1\in\mathcal{P}(\mathcal{X})$ satisfying $\mathcal{K}_c(\widehat{\mathbb{P}},\mathbb{P}_j)\leq\rho_j$ and $\mathbb{E}_{\mathbb{P}_j}[f]>-\infty$, $j=0,1$, and denote $\mathbb{P}_t=(1-t)\mathbb{P}_0+t\mathbb{P}_1$. For arbitrary $\epsilon>0$, we can find transport plans $\gamma_0\in\Gamma(\widehat{\mathbb{P}},\mathbb{P}_0)$, $\gamma_1\in\Gamma(\widehat{\mathbb{P}},\mathbb{P}_1)$ such that $\mathbb{E}_{\gamma_0}[c]\leq\mathcal{K}_c(\widehat{\mathbb{P}},\mathbb{P}_0)+\epsilon$, $\mathbb{E}_{\gamma_1}[c]\leq\mathcal{K}_c(\widehat{\mathbb{P}},\mathbb{P}_1)+\epsilon$. Define $\gamma_t=(1-t)\gamma_0+t\gamma_1$, then $\gamma_t\in\Gamma(\widehat{\mathbb{P}},\mathbb{P}_t)$ and

$$\begin{aligned}\mathcal{K}_c(\widehat{\mathbb{P}},\mathbb{P}_t)\leq\mathbb{E}_{\gamma_t}[c]=(1-t)\mathbb{E}_{\gamma_0}[c]+t\mathbb{E}_{\gamma_1}[c]&\leq(1-t)(\mathcal{K}_c(\widehat{\mathbb{P}},\mathbb{P}_0)+\epsilon)+t(\mathcal{K}_c(\widehat{\mathbb{P}},\mathbb{P}_1)+\epsilon)\\&\leq(1-t)\mathcal{K}_c(\widehat{\mathbb{P}},\mathbb{P}_0)+t\mathcal{K}_c(\widehat{\mathbb{P}},\mathbb{P}_1)+\epsilon\leq(1-t)\rho_0+t\rho_1+\epsilon.\end{aligned}$$

Since it is true for any ϵ, we know $\mathcal{K}_c(\widehat{\mathbb{P}},\mathbb{P}_t)\leq(1-t)\rho_0+t\rho_1$, hence $\mathbb{P}_t$ is a feasible solution to (P) with $\rho=(1-t)\rho_0+t\rho_1$ and

$$\mathcal{L}((1-t)\rho_0+t\rho_1)\geq\mathbb{E}_{X\sim\mathbb{P}_t}[f(X)]=(1-t)\mathbb{E}_{X\sim\mathbb{P}_0}[f(X)]+t\mathbb{E}_{X\sim\mathbb{P}_1}[f(X)].$$

Taking the supremum over $\mathbb{P}_0$ and $\mathbb{P}_1$, we have

$$\mathcal{L}((1-t)\rho_0+t\rho_1)\geq(1-t)\mathcal{L}(\rho_0)+t\mathcal{L}(\rho_1),$$

which completes the proof. □

Proof of Lemma 3. Since $f(\rho)=+\infty$ for $\rho<0$, we have

$$f^*(-\lambda)=\sup_{\rho\in\mathbb{R}}\{-\lambda\rho-f(\rho)\}=\sup_{\rho\geq 0}\{-\lambda\rho-f(\rho)\}.$$

For each fixed $\rho\geq 0$, $-\lambda\rho-f(\rho)$ is a monotonically decreasing, lower semi-continuous convex function of λ, so the supremum over ρ is also monotonically decreasing, lower semi-continuous, and convex. Suppose $f(\rho_0)<+\infty$ at some $\rho_0\geq 0$. For each $\lambda<0$,

$$f^*(-\lambda)=\sup_{\rho\geq 0}\{-\lambda\rho-f(\rho)\}\geq\sup_{\rho\geq\rho_0}\{-\lambda\rho-f(\rho)\}\geq\sup_{\rho\geq\rho_1}\{-\lambda\rho-f(\rho_0)\}=+\infty.$$

Pick $\rho_1>\rho_0$. For each $\lambda\geq 0$,

$$f^*(-\lambda)=\sup_{\rho\geq 0}\{-\lambda\rho-f(\rho)\}=\sup_{0\leq\rho\leq\rho_1}\{-\lambda\rho-f(\rho)\}\vee\sup_{\rho\geq\rho_1}\{-\lambda\rho-f(\rho)\}.$$

When $0 \le \rho \le \rho_1$, $-\lambda\rho - f(\rho) \le -f(\rho_1) < +\infty$. When $\rho \ge \rho_1$, by convexity we have $f(\rho) \ge f(\rho_1) + (\rho - \rho_1)\frac{f(\rho_1)-f(\rho_0)}{\rho_1-\rho_0}$, so if $\lambda \ge \frac{f(\rho_0)-f(\rho_1)}{\rho_1-\rho_0}$ we must have

$$-\lambda\rho - f(\rho) \le -\lambda\rho - f(\rho_1) - (\rho-\rho_1)\frac{f(\rho_1)-f(\rho_0)}{\rho_1-\rho_0} \le -f(\rho_1) - \rho_1\frac{f(\rho_0)-f(\rho_1)}{\rho_1-\rho_0} < -f(\rho_1) < +\infty.$$

Hence $f^*(\lambda) \le -f(\rho_1) < +\infty$, so $f^* \not\equiv +\infty$. □

Proof of Lemma 2. The lower bound follows by setting $x = \widehat{X}$: for any $\lambda \in [0,\infty)$, we have

$$(-\mathcal{L})^*(-\lambda) \ge \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}[f(\widehat{X})] - \lambda\mathcal{K}_c(\widehat{\mathbb{P}},\widehat{\mathbb{P}}) = \mathbb{E}_{\widehat{\mathbb{P}}}[f], \qquad \mathcal{G}^*(\lambda) \ge \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[f(\widehat{X}) - \lambda c(\widehat{X},\widehat{X})\right] = \mathbb{E}_{\widehat{\mathbb{P}}}[f].$$

Here we used $\mathcal{K}_c(\widehat{\mathbb{P}},\widehat{\mathbb{P}}) = 0$ because $c(x,x) = 0$ for every $x \in \mathcal{X}$.

If $\mathcal{L}(\rho) < +\infty$ for all $\rho > 0$, then $(-\mathcal{L})^*(-\cdot)$ is decreasing, convex, and lower semi-continuous because of Lemma 3. If $\mathcal{L}(\rho) = +\infty$ for all $\rho > 0$, then $(-\mathcal{L})^* \equiv +\infty$.

Since $f(x) - \lambda c(\widehat{x},x)$ is decreasing and affine in λ, $\Phi(\lambda;\widehat{x}) := \sup_{x\in\mathcal{X}}\{f(x) - \lambda c(\widehat{x},x)\}$ is also a decreasing, convex, and lower semi-continuous function of λ. Recall that when $\lambda = 0$ and $c(\widehat{x},x) = +\infty$, we use the convention $0\cdot\infty = \infty$. We now verify that

(a) $\mathcal{G}^*$ is monotonically decreasing:

$$\lambda_1 \le \lambda_2 \implies \Phi(\lambda_1;\widehat{x}) \ge \Phi(\lambda_2;\widehat{x}) \text{ for all } \widehat{x}\in\mathcal{X} \implies \mathbb{E}_{\widehat{\mathbb{P}}}[\Phi(\lambda_1;\widehat{X})] \ge \mathbb{E}_{\widehat{\mathbb{P}}}[\Phi(\lambda_2;\widehat{X})].$$

(b) $\mathcal{G}^*$ is convex:

$$\begin{aligned}\lambda_\theta = (1-\theta)\lambda_1 + \theta\lambda_2 &\implies \Phi(\lambda_\theta;\widehat{x}) \le (1-\theta)\Phi(\lambda_0;\widehat{x}) + \theta\Phi(\lambda_1;\widehat{x}) \text{ for all } \widehat{x}\in\mathcal{X}\\ &\implies \mathbb{E}_{\widehat{\mathbb{P}}}[\Phi(\lambda_\theta;\widehat{X})] \le (1-\theta)\mathbb{E}_{\widehat{\mathbb{P}}}[\Phi(\lambda_0;\widehat{X})] + \theta\mathbb{E}_{\widehat{\mathbb{P}}}[\Phi(\lambda_1;\widehat{X})].\end{aligned}$$

(c) $\mathcal{G}^*$ is lower semi-continuous: note that $\Phi(\lambda;\widehat{x}) \ge f(\widehat{x})$ and $\mathbb{E}_{\widehat{\mathbb{P}}}[f] > -\infty$. Taking $\lambda_n \to \lambda$ where $\lambda_n, \lambda \in [0,\infty)$, then by Fatou's lemma, we have

$$\liminf_{\lambda_n\to\lambda}\mathbb{E}_{\widehat{\mathbb{P}}}[\Phi(\lambda_n;\widehat{X})] \ge \mathbb{E}_{\widehat{\mathbb{P}}}\left[\liminf_{\lambda_n\to\lambda}\Phi(\lambda_n;\widehat{X})\right] \ge \mathbb{E}_{\widehat{\mathbb{P}}}[\Phi(\lambda)].$$

With this we complete the proof. □

Proof of Lemma 4. $f(x) - \lambda c(\widehat{x},x) \le f(x) - 0 = f(x) - \lambda c(x,x)$, so $f - \lambda c$ is diagonally dominant. If ϕ is diagonally dominant, we define $\lambda := 1$, $f(x) := \phi(x,x)$, and $c(\widehat{x},x) := f(x) - \phi(\widehat{x},x)$ when $f(x) > -\infty$, $c(\widehat{x},x) = 0$ when $f(x) = -\infty$. Then $c(\widehat{x},x) \ge 0$ and $c(x,x) = 0$.

Appendix EC.2: Proof of Proposition 1

Before proving Proposition 1, we make a simple observation.

LEMMA EC.1. *$A \subset \mathcal{X}\times\mathcal{X}$ is a diagonally dominant set if and only if its indicator function $\mathbf{1}_A$ is a diagonally dominant function. $\phi : \mathcal{X}\times\mathcal{X} \to \mathbb{R}\cup\{-\infty\}$ is a diagonally dominant function if and only if its superlevel set $\{\phi > \alpha\}$ is a diagonally dominant set for any $\alpha \in \mathbb{R}$. $E : \mathcal{X} \to \mathscr{F}\setminus\{\varnothing\}$ is a diagonally dominant set-valued function if and only if its graph is a diagonally dominant set.*

Proof of Proposition 1. We first prove the sufficiency. Let ϕ be a diagonally dominant $(\mathscr{F} \otimes \mathscr{F})$-measurable function. Define $\Phi(\widehat{x}) = \sup_{x\in\mathcal{X}} \phi(\widehat{x}, x)$. For any $\alpha \in \mathbb{R}$, the superlevel set of Φ can be regarded as

$$\{\widehat{x} : \Phi(\widehat{x}) > \alpha\} = \{\widehat{x} : \exists x, \phi(\widehat{x}, x) > \alpha\} = \mathrm{Proj}_{\widehat{x}}(\{(\widehat{x}, x) : \phi(\widehat{x}, x) > \alpha\}).$$

The superlevel set $\{\phi > \alpha\}$ is diagonally dominant. By assumption (Proj), $\mathrm{Proj}_{\widehat{x}}$ maps $(\mathscr{F} \otimes \mathscr{F})$-measurable diagonally dominant sets to $\mathscr{F}_{\widehat{\mathbb{P}}}$-measurable sets, thus the superlevel set of Φ is $\mathscr{F}_{\widehat{\mathbb{P}}}$-measurable. Therefore, Φ is $\widehat{\mathbb{P}}$-measurable, and it remains to show that

$$\mathbb{E}_{\widehat{\mathbb{P}}}[\Phi] = \sup_{\gamma\in\Gamma_{\widehat{\mathbb{P}}}} \mathbb{E}_\gamma[\phi].$$

Since for any $x \in \mathcal{X}$, $\phi(\widehat{x}, x) \le \Phi(\widehat{x})$, it is clear that for any $\gamma \in \Gamma_{\widehat{\mathbb{P}}}$,

$$\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\Phi(\widehat{X})\right] \ge \mathbb{E}_{(\widehat{X},X)\sim\gamma}\left[\phi(\widehat{X}, X)\right].$$

To see the other direction, we may assume $\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}[\Phi(\widehat{X})] > -\infty$, otherwise the conclusion holds trivially. Then $\{\Phi = -\infty\}$ is a $\widehat{\mathbb{P}}$-nullset. We fix $\epsilon, M > 0$, and below we construct a near optimal $\gamma \in \Gamma_{\widehat{\mathbb{P}}}$.

Define $S_n = \{\widehat{x} \in \mathcal{X} : n\epsilon < \Phi(\widehat{x}) \le (n+1)\epsilon\}$ for $n \in \mathbb{Z}$. S_n is $\mathscr{F}_{\widehat{\mathbb{P}}}$-measurable, so we can find $B_n \subset S_n$ which is $\mathscr{F}$-measurable and $\widehat{\mathbb{P}}(S_n \setminus B_n) = 0$. Define set-valued function $E_n : \mathcal{X} \to \mathscr{F} \setminus \{\varnothing\}$ by

$$E_n(\widehat{x}) = \begin{cases} \mathcal{X} \setminus B_n & \widehat{x} \notin B_n \\ \{x \in \mathcal{X} : \phi(\widehat{x}, x) > n\epsilon\} & \widehat{x} \in B_n \end{cases}.$$

$\mathrm{Graph}(E_n) = (\mathcal{X} \setminus B_n) \times (\mathcal{X} \setminus B_n) \cup ((B_n \times \mathcal{X}) \cap \{\phi > n\epsilon\})$ is $(\mathscr{F} \otimes \mathscr{F})$-measurable. We claim E_n is diagonally dominant. That is, $x \in E_n(\widehat{x})$ implies $x \in E_n(x)$. To see this, note that if $x \notin B_n$ then $x \in \mathcal{X} \setminus B_n = E_n(x)$; if $\widehat{x} \notin B_n$ and $x \in E_n(\widehat{x}) = \mathcal{X} \setminus B_n$ then $x \notin B_n$ so $x \in E_n(x)$. Now suppose $\widehat{x}, x \in B_n$ and $x \in E_n(\widehat{x})$, then $\phi(x, x) \ge \phi(\widehat{x}, x) > n\epsilon$, so again we have $x \in E_n(x)$. This finishes the proof of the claim. By assumption (Sel*) we can find a measurable transport plan $\gamma_n \in \Gamma_{\widehat{\mathbb{P}}}$ supported in $\mathrm{Graph}(E_n)$.

Define $S_\infty = \{\widehat{x} \in \mathcal{X} : \Phi(\widehat{x}) = \infty\}$. S_∞ is $\mathscr{F}_{\widehat{\mathbb{P}}}$-measurable, so we can find $B_\infty \subset S_\infty$ which is $\mathscr{F}$-measurable and $\widehat{\mathbb{P}}(S_\infty \setminus B_\infty) = 0$. For some $M > 0$ to be determined, define set-valued function $E_\infty : \mathcal{X} \to \mathscr{F} \setminus \{\varnothing\}$ by

$$E_\infty(\widehat{x}) = \begin{cases} \mathcal{X} \setminus B_\infty & \widehat{x} \notin B_\infty \\ \{x \in \mathcal{X} : \phi(\widehat{x}, x) > M\} & \widehat{x} \in B_\infty \end{cases}.$$

Same as before, E_∞ is diagonally dominant, so we can find a measurable transport plan $\gamma_\infty \in \Gamma_{\widehat{\mathbb{P}}}$ supported in $\mathrm{Graph}(E_\infty)$.

Now we define measure $\gamma \in \mathcal{P}(\mathcal{X} \times \mathcal{X}, \mathscr{F} \otimes \mathscr{F})$ by

$$\gamma(A) = \sum_{n\in\mathbb{Z}\cup\{+\infty\}} \gamma_n(A \cap (B_n \times \mathcal{X})).$$

Then for any $S \subset \mathcal{X}$,

$$\gamma(S \times \mathcal{X}) = \sum_{n \in \mathbb{Z} \cup \{+\infty\}} \gamma_n((S \cap B_n) \times \mathcal{X}) = \sum_{n \in \mathbb{Z} \cup \{+\infty\}} \widehat{\mathbb{P}}(S \cap B_n) = \widehat{\mathbb{P}}(S),$$

so $\gamma \in \Gamma_{\widehat{\mathbb{P}}}$. In the last equality, we used that $B_n \subset S_n$ are pairwise disjoint and $\widehat{\mathbb{P}}(\mathcal{X} \setminus \bigcup_{n \in \mathbb{Z} \cup \{+\infty\}} B_n) = 0$. Moreover, γ is supported in

$$\{(\widehat{x}, x) \in \mathcal{X} \times \mathcal{X} : \phi(\widehat{x}, x) > \Phi(\widehat{x}) - \epsilon \text{ if } \Phi(\widehat{x}) < \infty, \phi(\widehat{x}, x) > M \text{ if } \Phi(\widehat{x}) = \infty\}.$$

Therefore, $\mathbb{E}_\gamma[\phi] \geq \mathbb{E}_{\widehat{\mathbb{P}}}[(\Phi - \epsilon)\mathbf{1}\{\Phi < +\infty\}] + M\widehat{\mathbb{P}}(\Phi = +\infty)$. By making ϵ arbitrarily small and M arbitrarily large, we have $\sup_{\gamma \in \Gamma_{\widehat{\mathbb{P}}}} \mathbb{E}_\gamma[\phi] \geq \mathbb{E}_{\widehat{\mathbb{P}}}[\Phi]$. This proves that (Proj) and (Sel*) combined imply (IP) holds for all diagonally dominant functions.

Next, we prove the necessity. Suppose (IP) holds for all diagonally dominant functions. Given $A \in \mathcal{F} \otimes \mathcal{F}$ diagonally dominant, let ϕ be the indicator of the set A: $\phi(\widehat{x}, x) = 1$ if $(\widehat{x}, x) \in A$ and 0 otherwise. Then ϕ is $\mathcal{F} \otimes \mathcal{F}$-measurable, diagonally dominant, and by (IP) the function $\Phi(\widehat{x}) = \sup_{x \in \mathcal{X}} \phi(\widehat{x}, x)$ is $\widehat{\mathbb{P}}$-measurable. Observe that

$$\mathrm{Proj}_{\widehat{x}}(A) = \{\widehat{x} \in \mathcal{X} : \Phi(\widehat{x}) \geq 1\},$$

which is the upper level set of Φ, and thus belongs to $\mathcal{F}_{\widehat{\mathbb{P}}}$. Therefore (IP) implies (Proj). Lastly, given a diagonally dominant set function $E : \mathcal{X} \to \mathcal{F} \setminus \{\varnothing\}$, let

$$\phi(\widehat{x}, x) = \begin{cases} 0 & x \in E(\widehat{x}) \\ -\infty & x \notin E(\widehat{x}) \end{cases}.$$

That is, $\phi = -\infty \cdot \mathbf{1}_{\mathcal{X} \setminus \mathrm{Graph}(E)}$. To see it is diagonally dominant, note that

$$\phi(\widehat{x}, x) = 0 \implies x \in E(\widehat{x}) \implies x \in E(x) \implies \phi(x, x) = 0.$$

So $\phi(\widehat{x}, x) \leq \phi(x, x)$. Since $E(\widehat{x}) \neq \varnothing$, $\Phi(\widehat{x}) := \sup_{x \in \mathcal{X}} \phi(\widehat{x}, x) = 0$. By (IP),

$$0 = \mathbb{E}_{\widehat{X} \sim \widehat{\mathbb{P}}}\left[\Phi(\widehat{X}, x)\right] = \sup_{\gamma \in \Gamma_{\widehat{\mathbb{P}}}} \left\{\mathbb{E}_{(\widehat{X}, X) \sim \gamma}[\phi(\widehat{X}, X)]\right\}.$$

Note that

$$\mathbb{E}_\gamma[\phi] = \begin{cases} 0, & \mathrm{supp}\, \gamma \subset \mathrm{Graph}(E), \\ -\infty, & \text{otherwise.} \end{cases}$$

Hence, there exists some $\gamma \in \Gamma_{\widehat{\mathbb{P}}}$ supported in A. Therefore (IP) implies (Sel*). □

Proof of Proposition 2. Note that ϕ is continuous in $\widehat{x}$, so $\Phi(\widehat{x}) = \sup_{x \in \mathcal{X}} \phi(\widehat{x}, x)$ is lower semi-continuous, and thus Borel measurable. Therefore, for any $\gamma \in \Gamma_{\widehat{\mathbb{P}}}$, it holds that $\mathbb{E}_\gamma[\phi] \leq \mathbb{E}_{\widehat{\mathbb{P}}}[\Phi]$. It remains to construct an ϵ-optimizer γ.

First, we assume $\mathbb{E}_{\widehat{\mathbb{P}}}[\Phi] < +\infty$, and fix $\epsilon > 0$. Since $\widehat{\mathbb{P}}$ is tight, there exists a compact subset $\widehat{K} \subset\subset \mathcal{X}$ sufficiently large such that $\mathbb{E}_{\widehat{\mathbb{P}}}[|f|\mathbf{1}_{\widehat{K}^c}] < \epsilon$ and $\mathbb{E}_{\widehat{\mathbb{P}}}[|\Phi|\mathbf{1}_{\widehat{K}^c}] < \epsilon$. Moreover, fix some $\widehat{x}_0 \in \mathcal{X}$, we define

$$K_n = \widehat{K} \cup B_n(\widehat{x}_0) = \widehat{K} \cup \{x \in \mathcal{X} : d(\widehat{x}_0, x) \le n\},$$

and define

$$\phi_n(\widehat{x}, x) = \phi(\widehat{x}, x) - \infty\mathbf{1}\{x \notin K_n\}, \qquad \Phi_n(\widehat{x}) = \sup_{x \in K_n} \phi(\widehat{x}, x) = \sup_{x \in \mathcal{X}} \phi_n(\widehat{x}, x).$$

Then $\Phi_n \to \Phi$ pointwise as $n \to \infty$. Since $f \le \Phi_n \le \Phi$ in $\widehat{K}$, From dominant convergence theorem, we have

$$\mathbb{E}_{\widehat{\mathbb{P}}}[\Phi\mathbf{1}_{\widehat{K}}] - \mathbb{E}_{\widehat{\mathbb{P}}}[\Phi_n\mathbf{1}_{\widehat{K}}] < \epsilon$$

for n sufficiently large. We fix n from now on. Note that $\phi(\widehat{x}, x)$ is uniformly continuous in $\widehat{x}$ in $\widehat{K} \times K_n$: there exists $\delta > 0$ such that if $\widehat{x}_1, \widehat{x}_2 \in \widehat{K}$ and $d(\widehat{x}_1, \widehat{x}_2) < \delta$, we must have $|\phi(\widehat{x}_1, x) - \phi(\widehat{x}_2, x)| \le \epsilon$ for all $x \in K_n$, and consequently $|\Phi_n(\widehat{x}_1) - \Phi_n(\widehat{x}_2)| \le \epsilon$. Since $\widehat{K}$ is compact, there exists a δ-net $\widehat{\mathcal{X}} = \{\widehat{x}_i\}_{i=1}^n \subset \widehat{K}$. Define $U_i = \widehat{K} \cap B_\delta(\widehat{x}_i) \setminus \bigcup_{j<i} B_\delta(\widehat{x}_j)$, then $\{U_i\}_{i=1}^n$ forms a partition of $\widehat{K}$. For each $\widehat{x}_i$, we can find x_i such that

$$\phi(\widehat{x}_i, x_i) > \Phi(\widehat{x}_i) - \epsilon.$$

Now we construct a Borel-measurable selection mapping

$$T(\widehat{x}) = \begin{cases} x_i & \widehat{x} \in U_i \\ x & \widehat{x} \in \widehat{K}^c \end{cases}, \qquad \widehat{x} \in \mathcal{X}.$$

This induces a measure $\gamma = (\mathrm{Id} \otimes T)_\# \widehat{\mathbb{P}}$. Under this selection, we have

$$\mathbb{E}_\gamma[\phi] = \mathbb{E}_{\widehat{\mathbb{P}}}\left[\phi(\widehat{X}, T(\widehat{X}))\right] = \mathbb{E}_{\widehat{\mathbb{P}}}\left[\phi(\widehat{X}, \widehat{X})\mathbf{1}\{\widehat{X} \in \widehat{K}^c\}\right] + \sum_{i=1}^n \mathbb{E}_{\widehat{\mathbb{P}}}\left[\phi(\widehat{X}, x_i)\mathbf{1}\{\widehat{X} \in U_i\}\right].$$

The first term is

$$\mathbb{E}_{\widehat{\mathbb{P}}}\left[\phi(\widehat{X}, \widehat{X})\mathbf{1}\{\widehat{X} \in \widehat{K}^c\}\right] = \mathbb{E}_{\widehat{\mathbb{P}}}[f\mathbf{1}_{\widehat{K}^c}] \ge -\epsilon.$$

For the second term, note that $|\widehat{x} - \widehat{x}_i| < \delta$ for $\widehat{x} \in U_i$, so by uniform continuity we have

$$\phi(\widehat{x}, x_i) \ge \phi_n(\widehat{x}, x_i) \ge \phi_n(\widehat{x}_i, x_i) - \epsilon \ge \Phi_n(\widehat{x}_i) - 2\epsilon \ge \Phi_n(\widehat{x}) - 3\epsilon.$$

Hence we have

$$\mathbb{E}_\gamma[\phi] \ge -4\epsilon + \sum_{i=1}^n \mathbb{E}_{\widehat{\mathbb{P}}}\left[\Phi_n(\widehat{X})\mathbf{1}\{\widehat{X} \in U_i\}\right] = -4\epsilon + \mathbb{E}_{\widehat{\mathbb{P}}}[\Phi_n\mathbf{1}_{\widehat{K}}] > -5\epsilon + \mathbb{E}_{\widehat{\mathbb{P}}}[\Phi\mathbf{1}_{\widehat{K}}] > -6\epsilon + \mathbb{E}_{\widehat{\mathbb{P}}}[\Phi].$$

Since ϵ can be chosen arbitrarily small, we proved interchangeability for ϕ. The case $\mathbb{E}_{\widehat{\mathbb{P}}}[\Phi] = +\infty$ is similar and we omit the proof. □

REMARK EC.1. It can be seen from the proof that beyond the Wasserstein setting, we need c to be continuous in $\widehat{x}$ uniformly in $\widehat{K} \times K_n$ for any compact $\widehat{K}$ and some sequence of $K_n \supset \widehat{K}$ such that $\bigcup_n K_n = \mathcal{X}$. In particular, this would hold if c is continuous and $\mathcal{X}$ is a σ-compact metrizable topological space.

Appendix EC.3: Proofs for Section 4

In this section, we provide additional details of the proof in Example 4, Example 5, and Example 6.

Proof of Example 4. We start with $t = T$. We have

$$V_T(s) = \inf_{a\in\mathcal{A}(s)} g_T(s,a),$$

which is lower semi-analytic [4, Proposition 7.47], and in particular, $\widehat{\mathbb{P}}$-measurable [4, Corollary 7.42.1]. Since g_T is bounded from below by our assumption, V_T is also bounded from below. In a Borel space or Polish space, any measure is tight. Thus by Proposition 2 we have

$$\begin{aligned}
V_{T-1}(s) &= \inf_{a\in\mathcal{A}(s)} \left\{ g_{T-1}(s,a) + \sup_{\mathbb{P}\in\mathfrak{M}(s,a)} \mathbb{E}_{s'\sim\mathbb{P}(\cdot|s,a)}[V_T(s')] \right\} \\
&= \inf_{\substack{a\in\mathcal{A}(s)\\ \lambda\geq 0}} \left\{ g_{T-1}(s,a) + \lambda\rho(s,a) + \mathbb{E}_{\widehat{s}'\sim\widehat{\mathbb{P}}(\cdot|s,a)}\left[\sup_{s'\in\mathcal{S}} \left\{ V_T(s') - \lambda c(\widehat{s}', s') \right\} \right] \right\},
\end{aligned}$$

which is also bounded from below since g_{T-1} is bounded from below by assumption and $\rho > 0$.

Suppose we have shown $V_{t+1}(\cdot)$ is lower semi-analytic, V_t is bounded from below and obtain the reformulation for V_t. Now let us show that $V_t(\cdot)$ is lower semi-analytic and derive the expression for V_{t-1} and show it is bounded from below. By the continuity of c, $\widehat{s}' \mapsto c(\widehat{s}', s')$ is continuous for each s', so the function $\widehat{s}' \mapsto \sup_{s'\in\mathcal{S}}\left\{V_{t+1}(s') - \lambda c(\widehat{s}', s')\right\}$ is the supremum of a family of continuous function, which is lower semi-continuous, and furthermore is Borel measurable and thus lower semi-analytic. Hence the function $(s,a) \mapsto \mathbb{E}_{\widehat{s}'\sim\widehat{\mathbb{P}}(\cdot|s,a)}\left[\sup_{s'\in\mathcal{S}}\left\{V_{t+1}(s') - \lambda c(\widehat{s}', s')\right\}\right]$ is lower semi-analytic [4, Proposition 7.48]. Since g_{t-1} and ρ are lower semi-analytic due to our assumptions, V_t is also lower semi-analytic [4, Proposition 7.47]. Then using Proposition 2 again we have

$$\begin{aligned}
V_{t-1}(s) &= \inf_{a\in\mathcal{A}(s)} \left\{ g_{t-1}(s,a) + \sup_{\mathbb{P}\in\mathfrak{M}(s,a)} \mathbb{E}_{s'\sim\mathbb{P}(\cdot|s,a)}[V_t(s')] \right\} \\
&= \inf_{\substack{a\in\mathcal{A}(s)\\ \lambda\geq 0}} \left\{ g_{t-1}(s,a) + \lambda\rho(s,a) + \mathbb{E}_{\widehat{s}'\sim\widehat{\mathbb{P}}(\cdot|s,a)}\left[\sup_{s'\in\mathcal{S}} \left\{ V_t(s') - \lambda c(\widehat{s}', s') \right\} \right] \right\},
\end{aligned}$$

which is bounded from below since g_{t-1}, ρ, V_t are all bounded from below. Therefore the proof is completed.

□

Proof of Example 5. We start with $t = T$. In this case,

$$Q_T(u_{T-1}, x_T) = \inf_{u\in\mathcal{U}_T(u_{T-1},x_T)} f_T(u, x_T).$$

Since f_T is random lower semi-continuous and $\mathcal{U}_T$ is uniformly bounded, $Q_T(\cdot,\cdot)$ is random lower semi-continuous [21, Theorem 9.50], and particularly, $Q_T(u_{T-1},\cdot)$ is measurable. Since f_T is bounded from below by our assumption, Q_T is also bounded from below. Thus Assumption 1 holds and using Example 3 we have

$$Q_{T-1}(u_{T-2}, x_{T-1}) = \inf_{u\in\mathcal{U}_{T-1}(u_{T-2},x_{T-1})} \left\{ f_{T-1}(u, x_{T-1}) + \sup_{\mathbb{P}\in\mathfrak{M}_T} \mathbb{E}_{\mathbb{P}}\left[Q_T(u, x_T)\right] \right\}$$

$$= \inf_{\substack{u \in \mathcal{U}_{T-1}(u_{T-2}, x_{T-1}) \\ \lambda \geq 0}} \left\{ f_{T-1}(u, x_{T-1}) + \lambda \rho_T + \mathbb{E}_{\widehat{\mathbb{P}}_T} \left[\sup_{x \in \mathcal{X}_T} \left\{ Q_T(u,x) - \lambda c(\widehat{x}_T, x) \right\} \right] \right\},$$

which is also bounded from below since f_{T-1}, ρ_T, Q_T are all bounded from below.

Suppose we have shown $Q_{t+1}(\cdot,\cdot)$ is random lower semi-continuous and obtain the reformulation for Q_t that is bounded from below. Now let us show that $Q_t(\cdot,\cdot)$ is random lower semi-continuous and derive the expression for Q_{t-1} and show it is bounded from below. Since $Q_{t+1}(\cdot,\cdot)$ is random lower semi-continuous, $u \mapsto Q_{t+1}(u,x)$ is lower semi-continuous for every $x \in \mathcal{X}_{t+1}$. Therefore, for every $x, \widehat{x}_{t+1} \in \mathcal{X}_{t+1}$, the function $u \mapsto Q_{t+1}(u,x) - \lambda c(\widehat{x}_{t+1}, x)$ is lower semi-continuous, and thus their supremum $u \mapsto \sup_{x \in \mathcal{X}_{t+1}} \{Q_{t+1}(u,x) - \lambda c(\widehat{x}_{t+1}, x)\}$ is lower semi-continuous, and taking the expectation with respect to $\widehat{\mathbb{P}}_{t+1}$, the function $\mathbb{E}_{\widehat{\mathbb{P}}_{t+1}} \left[\sup_{x \in \mathcal{X}_{t+1}} \left\{Q_{t+1}(u,x) - \lambda c(\widehat{x}_{t+1}, x)\right\}\right]$ is lower semi-continuous in u. It follows from [21, Theorem 9.50] that $Q_t(\cdot,\cdot)$ is random lower semi-continuous. Then using Example 3 we have

$$\begin{aligned} Q_{t-1}(u_{t-2}, x_{t-1}) &= \inf_{u \in \mathcal{U}_{t-1}(u_{t-2}, x_{t-1})} \left\{ f_{t-1}(u, x_{t-1}) + \sup_{\mathbb{P} \in \mathfrak{M}_t} \mathbb{E}_{\mathbb{P}} \left[Q_t(u, x_t) \right] \right\} \\ &= \inf_{\substack{u \in \mathcal{U}_{t-1}(u_{t-2}, x_{t-1}) \\ \lambda \geq 0}} \left\{ f_{t-1}(u, x_{t-1}) + \lambda \rho_t + \mathbb{E}_{\widehat{\mathbb{P}}_t} \left[\sup_{x \in \mathcal{X}_t} \left\{ Q_t(u,x) - \lambda c(\widehat{x}_t, x) \right\} \right] \right\}, \end{aligned}$$

which is again bounded from below, and thereby we complete the induction. □

Proof of Example 6. Define $f = \mathbf{1}_{\mathcal{S}^c}$, $c(\widehat{x}, x) = d(\widehat{x}, x)^p$. Then (3) is equivalent to

$$\mathcal{L}(\rho^p) = \sup_{\mathbb{P} \in \mathcal{P}(\mathcal{X})} \left\{ \mathbb{E}_{X \sim \mathbb{P}}[f(X)] : \mathcal{K}_c(\widehat{\mathbb{P}}, \mathbb{P}) \leq \rho^p \right\} \leq \beta.$$

For $p \in [1, \infty)$, we observe the following for each $\widehat{x} \in \mathcal{X}$:

$$\begin{aligned} \sup_{x \in \mathcal{X}} \{f(x) - \lambda c(\widehat{x}, x)\} &= \sup_{x \in \mathcal{S}^c} \{1 - \lambda d(\widehat{x}, x)^p\} \vee \sup_{x \in \mathcal{S}} \{0 - \lambda d(\widehat{x}, x)^p\} \\ &= (1 - \lambda d(\widehat{x}, \mathcal{S}^c)^p) \vee (-\lambda d(\widehat{x}, \mathcal{S})^p) \\ &= \begin{cases} 1, & \widehat{x} \in \mathcal{S}^c \\ (1 - \lambda d(\widehat{x}, \mathcal{S}^c)^p)_+, & \widehat{x} \in \mathcal{S} \end{cases} \\ &= (1 - \lambda d(\widehat{x}, \mathcal{S}^c)^p)_+. \end{aligned}$$

By Theorem 1, if $f - \lambda c$ satisfies (IP) for every $\lambda > 0$, then the dual problem can be calculated as the following:

$$(-\mathcal{L})^*(-\lambda) = \mathbb{E}_{\widehat{X} \sim \widehat{\mathbb{P}}} \left[\sup_{x \in \mathcal{X}} \left\{ f(x) - \lambda c(\widehat{X}, x) \right\} \right] = \mathbb{E}_{\widehat{X} \sim \widehat{\mathbb{P}}} \left[(1 - \lambda d(\widehat{X}, \mathcal{S}^c)^p)_+ \right].$$

Therefore, for every $\rho > 0$, we have

$$\mathcal{L}(\rho^p) = \min_{\lambda \geq 0} \{\lambda \rho^p + (-\mathcal{L})^*(-\lambda)\} = \min_{\lambda \geq 0} \left\{ \lambda \rho^p + \mathbb{E}_{\widehat{X} \sim \widehat{\mathbb{P}}} \left[(1 - \lambda d(\widehat{X}, \mathcal{S}^c)^p)_+ \right] \right\}.$$

For $\beta \in (0,1)$, the chance constraint can be written as

$$\begin{aligned}
\mathcal{L}(\rho^p) \le \beta &\iff \lambda\rho^p + \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[(1-\lambda d(\widehat{X},\mathcal{S}^c)^p)_+\right] \le \beta \text{ for some } \lambda \ge 0 \\
&\iff \lambda\rho^p + \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[(1-\lambda d(\widehat{X},\mathcal{S}^c)^p)_+\right] \le \beta \text{ for some } \lambda > 0 \\
&\iff \frac{\rho^p}{\beta} + \frac{1}{\beta}\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\left(\frac{1}{\lambda} - d(\widehat{X},\mathcal{S}^c)^p\right)_+\right] \le \frac{1}{\lambda} \text{ for some } \lambda > 0 \\
&\iff \alpha + \frac{1}{\beta}\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\left(-d(\widehat{X},\mathcal{S}^c)^p - \alpha\right)_+\right] \le -\frac{\rho^p}{\beta} \text{ for some } \alpha < 0 \\
&\iff \alpha + \frac{1}{\beta}\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\left(-d(\widehat{X},\mathcal{S}^c)^p - \alpha\right)_+\right] \le -\frac{\rho^p}{\beta} \text{ for some } \alpha \in \mathbb{R} \\
&\iff \mathbb{CV@R}_\beta^{\widehat{\mathbb{P}}}(-d(\widehat{X},\mathcal{S}^c)^p) = \min_{\alpha\in\mathbb{R}}\left\{\alpha + \frac{1}{\beta}\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\left(-d(\widehat{X},\mathcal{S}^c)^p - \alpha\right)_+\right]\right\} \le -\frac{\rho^p}{\beta}. \qquad \square
\end{aligned}$$

Appendix EC.4: Proofs for Section 5.1

Proof of Proposition 3. We compute $\overline{\mathcal{L}}^\circ(\rho)$ as follows.

$$\begin{aligned}
\overline{\mathcal{L}}^\circ(\rho) &= \sup_{\mathbb{P}\in\mathcal{P}(\mathcal{X})}\left\{\mathbb{E}_{X\sim\mathbb{P}}[f(X)] : \overline{\mathcal{K}}_c(\widehat{\mathbb{P}},\mathbb{P}) < \rho\right\} \\
&= \sup_{\mathbb{P}\in\mathcal{P}(\mathcal{X})}\left\{\mathbb{E}_{X\sim\mathbb{P}}[f(X)] : \inf_{\gamma\in\Gamma(\widehat{\mathbb{P}},\mathbb{P})} \gamma\text{-}\operatorname*{ess\,sup}_{\widehat{x},x\in\mathcal{X}} c(\widehat{x},x) < \rho\right\} \\
&= \sup_{\mathbb{P}\in\mathcal{P}(\mathcal{X})}\left\{\mathbb{E}_{X\sim\mathbb{P}}[f(X)] : \gamma\text{-}\operatorname*{ess\,sup}_{\widehat{x},x\in\mathcal{X}} c(\widehat{x},x) < \rho \text{ for some } \gamma\in\Gamma(\widehat{\mathbb{P}},\mathbb{P})\right\} \\
&= \sup_{\gamma\in\Gamma_{\widehat{\mathbb{P}}}}\left\{\mathbb{E}_{(\widehat{X},X)\sim\gamma}[f(X)] : \gamma\text{-}\operatorname*{ess\,sup}_{\widehat{x},x\in\mathcal{X}} c(\widehat{x},x) < \rho\right\} \\
&= \sup_{\gamma\in\Gamma_{\widehat{\mathbb{P}}}}\left\{\mathbb{E}_{(\widehat{X},X)\sim\gamma}\left[f(X) - \infty\mathbf{1}\{c(\widehat{X},X)\ge\rho\}\right]\right\} \\
&= \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_x\left\{f(x) - \infty\mathbf{1}\{c(\widehat{X},x)\ge\rho\}\right\}\right]. \\
&= \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_x\left\{f(x) : c(\widehat{X},x) < \rho\right\}\right].
\end{aligned}$$

In the second to the last line, we need (IP) on the function $\phi(\widehat{x},x) = f(x) - \infty\mathbf{1}\{c(\widehat{x},x)\ge\rho\}$.

Next, we compute the dual

$$\begin{aligned}
(-\overline{\mathcal{L}}^\circ)^*(-\lambda) = \sup_{\rho\ge0}\left\{\overline{\mathcal{L}}^\circ(\rho) - \lambda\rho\right\} &= \sup_{\rho\ge0}\sup_{\mathbb{P}\in\mathcal{P}(\mathcal{X})}\left\{\mathbb{E}_{X\sim\mathbb{P}}[f(X)] - \lambda\rho : \mathcal{W}_\infty(\widehat{\mathbb{P}},\mathbb{P}) < \rho\right\} \\
&= \sup_{\mathbb{P}\in\mathcal{P}(\mathcal{X})}\left\{\mathbb{E}_{X\sim\mathbb{P}}[f(X)] - \lambda\mathcal{W}_\infty(\widehat{\mathbb{P}},\mathbb{P})\right\} \\
&= \sup_{\rho\ge0}\sup_{\mathbb{P}\in\mathcal{P}(\mathcal{X})}\left\{\mathbb{E}_{X\sim\mathbb{P}}[f(X)] - \lambda\rho : \mathcal{W}_\infty(\widehat{\mathbb{P}},\mathbb{P}) \le \rho\right\} \\
&= \sup_{\rho\ge0}\left\{\overline{\mathcal{L}}(\rho) - \lambda\rho\right\} = (-\overline{\mathcal{L}})^*(-\lambda).
\end{aligned}$$

It follows that both $(-\overline{\mathcal{L}}^{\circ})^*(-\lambda)$ and $(-\overline{\mathcal{L}})^*(-\lambda)$ equal the soft-constrained robust loss. Thus

$$(-\overline{\mathcal{L}})^*(-\lambda) = (-\overline{\mathcal{L}}^{\circ})^*(-\lambda) = \sup_{\rho\geq 0}\left\{\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_x\left\{f(x) : c(\widehat{X},x) < \rho\right] - \lambda\rho\right\}\right\}.$$

This completes the proof of the proposition. We remark that $\overline{\mathcal{L}}^{\circ}$ is no longer necessarily concave, so $(-\overline{\mathcal{L}})^{**}(\rho)$ may differ from $-\overline{\mathcal{L}}^{\circ}(\rho)$. □

Proof of Theorem 2. Similarly to the above proof of Propositions 3, we have

$$\begin{aligned}
\overline{\mathcal{L}}(\rho) &= \sup_{\mathbb{P}\in\mathcal{P}(\mathcal{X})}\left\{\mathbb{E}_{X\sim\mathbb{P}}[f(X)] : \overline{\mathcal{K}}_c(\widehat{\mathbb{P}},\mathbb{P}) \leq \rho\right\} \\
&= \sup_{\mathbb{P}\in\mathcal{P}(\mathcal{X})}\left\{\mathbb{E}_{X\sim\mathbb{P}}[f(X)] : \inf_{\gamma\in\Gamma(\widehat{\mathbb{P}},\mathbb{P})} \gamma\text{-}\operatorname*{ess\,sup}_{\widehat{x},x\in\mathcal{X}} c(\widehat{x},x) \leq \rho\right\} \\
&\leq \sup_{\mathbb{P}\in\mathcal{P}(\mathcal{X})}\left\{\mathbb{E}_{X\sim\mathbb{P}}[f(X)] : \gamma\text{-}\operatorname*{ess\,sup}_{\widehat{x},x\in\mathcal{X}} c(\widehat{x},x) \leq \rho \text{ for some } \gamma\in\Gamma(\widehat{\mathbb{P}},\mathbb{P})\right\} \\
&= \sup_{\gamma\in\Gamma_{\widehat{\mathbb{P}}}}\left\{\mathbb{E}_{(\widehat{X},X)\sim\gamma}[f(X)] : \gamma\text{-}\operatorname*{ess\,sup}_{\widehat{x},x\in\mathcal{X}} c(\widehat{x},x) \leq \rho\right\} \\
&= \sup_{\gamma\in\Gamma_{\widehat{\mathbb{P}}}}\left\{\mathbb{E}_{(\widehat{X},X)\sim\gamma}\left[f(X) - \infty\mathbf{1}\{c(\widehat{X},X) > \rho\}\right]\right\} \\
&= \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_x\left\{f(x) - \infty\mathbf{1}\{c(\widehat{X},x) > \rho\} : c(\widehat{X},x) \leq \rho\right\}\right]. \\
&= \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_x\left\{f(x) : c(\widehat{X},x) \leq \rho\right\}\right].
\end{aligned}$$

Here we used (IP) as it holds for all measurable functions according to Example 2. Inequality becomes equality if

$$\inf_{\gamma\in\Gamma(\widehat{\mathbb{P}},\mathbb{P})} \gamma\text{-}\operatorname*{ess\,sup}_{\widehat{x},x\in\mathcal{X}} c(\widehat{x},x) \leq \rho \tag{EC.1}$$

can be achieved at some $\gamma\in\Gamma(\widehat{\mathbb{P}},\mathbb{P})$.

Now we use the additional information that $\mathcal{X}$ is Polish. If (EC.1) holds, we first find $\gamma_n\in\Gamma(\widehat{\mathbb{P}},\mathbb{P})$ with $\operatorname{supp}\gamma_n \subset \{c\leq\rho+\frac{1}{n}\}$. For any $\epsilon>0$, we can find compact sets $\widehat{K},K\subset\mathcal{X}$ with $\widehat{\mathbb{P}}[\widehat{K}]>1-\epsilon$ and $\mathbb{P}[K]>1-\epsilon$, because $\widehat{\mathbb{P}}$ and $\mathbb{P}$ are probability measures on a Polish space, which are tight. Then $\gamma_n[\widehat{K}\times K]>1-2\epsilon$ for each n. This shows $\{\gamma_n\}_n$ is a tight sequence. Since $\mathcal{X}$ is complete and separable, by Prokhorov theorem, there exists a weakly converging subsequence $\gamma_{n_k}\to\gamma\in\mathcal{P}(\mathcal{X}\times\mathcal{X})$. Marginals of γ_n also converge weakly to the marginals of γ, so $\gamma\in\Gamma(\widehat{\mathbb{P}},\mathbb{P})$. To show that γ is supported in $\{c\leq\rho\}$, define $g:\mathcal{X}\times\mathcal{X}\to\mathbb{R}$ by

$$g(\widehat{x},x) = 1\wedge(c(\widehat{x},x)-\rho)_+.$$

g is continuous and bounded on $\mathcal{X}\times\mathcal{X}$. Moreover, $\mathbb{E}_{\gamma_n}[g]\leq\frac{1}{n}$. By weak convergence, $\mathbb{E}_\gamma[g]=0$, so $c(\widehat{x},x)\leq\rho$ for γ-a.e. $(\widehat{x},x)\in\mathcal{X}\times\mathcal{X}$. That is, $\gamma\text{-}\operatorname{ess\,sup} c\leq\rho$.

Next, we compute the dual.

$$(-\overline{\mathcal{L}})^*(-\lambda) = \sup_{\rho\geq 0}\left\{\overline{\mathcal{L}}(\rho) - \lambda\rho\right\} = \sup_{\rho\geq 0}\sup_{\mathbb{P}\in\mathcal{P}(\mathcal{X})}\left\{\mathbb{E}_{X\sim\mathbb{P}}[f(X)] - \lambda\rho : \mathcal{W}_\infty(\widehat{\mathbb{P}},\mathbb{P})\leq\rho\right\}$$
$$= \sup_{\mathbb{P}\in\mathcal{P}(\mathcal{X})}\left\{\mathbb{E}_{X\sim\mathbb{P}}[f(X)] - \lambda\mathcal{W}_\infty(\widehat{\mathbb{P}},\mathbb{P})\right\}.$$

Thus

$$(-\overline{\mathcal{L}})^*(-\lambda) = \sup_{\rho\geq 0}\left\{\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_x\left\{f(x) : c(\widehat{X},x)\leq\rho\right]-\lambda\rho\right\}\right\}.$$

This completes the proof of the theorem. □

Proof of Example 7. By Theorem 2, we have

$$\overline{\mathcal{L}}(\rho) = \sup_{\mathbb{P}\in\mathcal{P}(\mathcal{X})}\{\mathbb{E}_{X\sim\mathbb{P}}[f(X)] : \mathcal{W}_\infty(\widehat{\mathbb{P}},\mathbb{P})\leq\rho\} = \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_x\left\{f(x) : d(\widehat{X},x)\leq\rho\right\}\right] = \widehat{\mathbb{P}}(d(\widehat{X},\mathcal{S}^c)\leq\rho).$$

In particular, $\overline{\mathcal{L}}(\rho) = 1$ if $\rho\geq d(\operatorname{supp}\widehat{\mathbb{P}},\mathcal{S}^c)$. We remark that now the corresponding soft robust problem is

$$(-\overline{\mathcal{L}})^*(-\lambda) = \sup_{\rho\geq 0}\left\{\widehat{\mathbb{P}}(d(\widehat{X},\mathcal{S}^c)\leq\rho) - \lambda\rho\right\}.$$ □

Appendix EC.5: Proofs for Section 5.2

Proof of Theorem 3. First, we consider the maximum transport cost $\overline{\mathcal{K}}_c$. Denote

$$J'(\mathbb{P}) := \inf_{\alpha\in A'}\mathbb{E}_{X\sim\mathbb{P}}[f_\alpha(X)].$$

By assumption (5), we know that

$$\overline{\mathcal{L}}_{\mathrm{J}}(\rho) = \sup_{\mathbb{P}\in\mathcal{P}(\mathcal{X})}\left\{J'(\mathbb{P}) : \overline{\mathcal{K}}_c(\widehat{\mathbb{P}},\mathbb{P})\leq\rho\right\}$$

Similar as Theorem 2, we have

$$\begin{aligned}
\overline{\mathcal{L}}_{\mathrm{J}}(\rho) &= \sup_{\mathbb{P}\in\mathcal{P}(\mathcal{X})}\left\{\inf_{\alpha\in A'}\mathbb{E}_{X\sim\mathbb{P}}[f_\alpha(X)] : \overline{\mathcal{K}}_c(\widehat{\mathbb{P}},\mathbb{P})\leq\rho\right\}\\
&= \sup_{\mathbb{P}\in\mathcal{P}(\mathcal{X})}\left\{\inf_{\alpha\in A'}\mathbb{E}_{X\sim\mathbb{P}}[f_\alpha(X)] : \inf_{\gamma\in\Gamma(\widehat{\mathbb{P}},\mathbb{P})}\gamma\text{-}\operatorname*{ess\,sup}_{\widehat{x},x\in\mathcal{X}} c(\widehat{x},x)\leq\rho\right\}\\
&= \sup_{\mathbb{P}\in\mathcal{P}(\mathcal{X})}\left\{\inf_{\alpha\in A'}\mathbb{E}_{X\sim\mathbb{P}}[f_\alpha(X)] : \gamma\text{-}\operatorname*{ess\,sup}_{\widehat{x},x\in\mathcal{X}} c(\widehat{x},x)\leq\rho \text{ for some } \gamma\in\Gamma(\widehat{\mathbb{P}},\mathbb{P})\right\}\\
&= \sup_{\gamma\in\Gamma_{\widehat{\mathbb{P}}}}\left\{\inf_{\alpha\in A'}\mathbb{E}_{X\sim\mathbb{P}}[f_\alpha(X)] : \gamma\text{-}\operatorname*{ess\,sup}_{\widehat{x},x\in\mathcal{X}} c(\widehat{x},x)\leq\rho\right\}\\
&= \sup_{\gamma\in\Gamma_{\widehat{\mathbb{P}}}}\left\{\inf_{\alpha\in A'}\mathbb{E}_{(\widehat{X},X)\sim\gamma}\left[f_\alpha(X) - \infty\mathbf{1}\{c(\widehat{X},X)>\rho\}\right]\right\}.
\end{aligned}$$

We claim that for each $\gamma\in\Gamma_{\widehat{\mathbb{P}}}$, $A'\ni\alpha\mapsto\mathbb{E}_\gamma[f_\alpha(X)]$ has the following properties:

(a) Convexity: given $\alpha_0, \alpha_1 \in \mathbb{R}$, $\theta \in (0,1)$, define $\alpha_\theta = (1-\theta)\alpha_0 + \theta\alpha_1$. Due to convexity of f_α in α we have $f_{\alpha_\theta}(x) \le (1-\theta) f_{\alpha_0}(x) + \theta f_{\alpha_1}(x)$ for every $x \in \mathcal{X}$. So

$$\mathbb{E}_\gamma[f_{\alpha_\theta}(X)] \le (1-\theta)\mathbb{E}_\gamma[f_{\alpha_0}(X)] + \theta\mathbb{E}_\gamma[f_{\alpha_1}(X)].$$

(b) Lower semi-continuity: let $\alpha_n \to \alpha$ as $n \to \infty$. Due to lower semi-continuity of f_α in α we have $f_\alpha(x) \le \liminf_{n\to\infty} f_{\alpha_n}(x)$ for every $x \in \mathcal{X}$. So

$$\mathbb{E}_\gamma[f_\alpha(X)] \le \mathbb{E}_\gamma\left[\liminf_{n\to\infty} f_{\alpha_n}(X)\right] \le \liminf_{n\to\infty} \mathbb{E}_\gamma[f_{\alpha_n}(X)].$$

The second inequality is due to Fatou's lemma thanks to $\{f_\alpha\}_{\alpha\in A'}$ being uniformly bounded from below.

Since A' is compact, by Sion's minimax theorem and interchangeability,

$$\begin{aligned}
\overline{\mathcal{L}}_{\mathrm{J}}(\rho) &= \inf_{\alpha\in A'} \left\{ \sup_{\gamma\in\Gamma_{\widehat{\mathbb{P}}}} \mathbb{E}_{(\widehat{X},X)\sim\gamma}\left[f_\alpha(X) - \infty\mathbf{1}\{c(\widehat{X},X) > \rho\}\right] \right\} \\
&= \inf_{\alpha\in A'} \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_x \left\{ f_\alpha(x) - \infty\mathbf{1}\{c(\widehat{X},x) > \rho\} : c(\widehat{X},x) \le \rho \right\}\right]. \\
&= \inf_{\alpha\in A'} \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_x \left\{ f_\alpha(x) : c(\widehat{X},x) \le \rho \right\}\right] =: \inf_{\alpha\in A'} \ell_\alpha.
\end{aligned}$$

To complete the proof, we need to enlarge A' to A again. Let α_n be a minimizing sequence:

$$\lim_{n\to\infty} \ell_{\alpha_n} = \inf_{\alpha\in A} \ell_\alpha.$$

Denote $A'_n = \mathrm{conv}(A' \cup \{\alpha_n\})$ to be the convex hull of A' and α_n. Clearly, if (5) holds for A', it should also hold for $A'_n \supset A'$. Since A'_n is still compact, the same argument on A'_n instead of A' gives $\overline{\mathcal{L}}_{\mathrm{J}}(\rho) = \inf_{\alpha\in A'_n} \ell_\alpha$. This holds for every n, so taking the limit yields

$$\overline{\mathcal{L}}_{\mathrm{J}}(\rho) = \inf_{\alpha\in A} \ell_\alpha = \inf_{\alpha\in A} \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_x \left\{ f_\alpha(x) : c(\widehat{X},x) \le \rho \right\}\right].$$

Next, we consider the Kantorovich cost $\mathcal{K}_c$. Similarly, we can restrict to A' by assumption (5):

$$\mathcal{L}_{\mathrm{J}}(\rho) = \sup_{\mathbb{P}\in\mathcal{P}(\mathcal{X})} \left\{ J'(\mathbb{P}) : \mathcal{K}_c(\widehat{\mathbb{P}},\mathbb{P}) \le \rho \right\}.$$

It is easy to see that J' is concave in $\mathbb{P}$. Similar as Lemma 1, it is a simple exercise to show $\mathcal{L}_{\mathrm{J}}(\cdot)$ is lower bounded by $\sup_{\alpha\in A'} \mathbb{E}_{\widehat{\mathbb{P}}}[f_\alpha]$, monotonically increasing, and concave on $[0,\infty)$. Now we take the dual of $-\mathcal{L}_{\mathrm{J}}$:

$$\begin{aligned}
(-\mathcal{L}_{\mathrm{J}})^*(-\lambda) := \sup_{\rho\ge 0}\{\mathcal{L}_{\mathrm{J}}(\rho) - \lambda\rho\} &= \sup_{\rho\ge 0, \mathbb{P}\in\mathcal{P}(\mathcal{X})} \left\{ J'(\mathbb{P}) - \lambda\rho : \mathcal{K}_c(\widehat{\mathbb{P}},\mathbb{P}) \le \rho \right\} \\
&= \sup_{\mathbb{P}\in\mathcal{P}(\mathcal{X})} \left\{ J'(\mathbb{P}) - \lambda\mathcal{K}_c(\widehat{\mathbb{P}},\mathbb{P}) \right\} \\
&= \sup_{\gamma\in\Gamma_{\widehat{\mathbb{P}}}} \inf_{\alpha\in A'} \left\{ \mathbb{E}_\gamma[f_\alpha(X) - \lambda c(\widehat{X},X)] \right\}.
\end{aligned}$$

We have shown that $\alpha \mapsto \mathbb{E}_\gamma[f_\alpha(X)]$ is lower semi-continuous and convex in the first half of the proof. By Sion's minimax theorem, we can exchange sup and inf, so

$$(-\mathcal{L}_{\mathrm{J}})^*(-\lambda) = \inf_{\alpha\in A'} \sup_{\gamma\in\Gamma_{\widehat{\mathbb{P}}}} \left\{\mathbb{E}_\gamma[f_\alpha(X) - \lambda c(\widehat{X}, X)]\right\} = \inf_{\alpha\in A'} \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}} \left[\sup_{x\in\mathcal{X}} \left\{f_\alpha(x) - \lambda c(\widehat{X}, x)\right\}\right].$$

Here we used (IP) on $\phi_{\lambda,\alpha} = f_\alpha - \lambda c$. By enlarging A' to A as in the maximal cost case, we have

$$(-\mathcal{L}_{\mathrm{J}})^*(-\lambda) = \inf_{\alpha\in A} \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}} \left[\sup_{x\in\mathcal{X}} \left\{f_\alpha(x) - \lambda c(\widehat{X}, x)\right\}\right]. \qquad \square$$

To apply Theorem 3 on Example 8, we need to verify that the infimum is achieved in a finite interval dependent only on the transport cost. The following lemma confirms this property for $\mathbb{CV@R}$ and $\mathbb{MAD}$.

LEMMA EC.2. *Suppose $\mathcal{X} = \mathbb{R}$. Let $\widehat{\mathbb{P}}, \mathbb{P} \in \mathcal{P}(\mathcal{X})$, $\beta \in (0,1)$. Let $\widehat{X} \sim \widehat{\mathbb{P}}$, $X \sim \mathbb{P}$. If $\mathbb{P}(X \geq \alpha) \geq \beta$ and $\mathbb{P}(X \leq \alpha) \geq 1-\beta$, then*

$$\alpha \in \left[-\mathbb{CV@R}^{\widehat{\mathbb{P}}}_{1-\beta}(-\widehat{X}) - \frac{1}{1-\beta}\mathcal{W}_1(\widehat{\mathbb{P}}, \mathbb{P}), \mathbb{CV@R}^{\widehat{\mathbb{P}}}_{\beta}(\widehat{X}) + \frac{1}{\beta}\mathcal{W}_1(\widehat{\mathbb{P}}, \mathbb{P})\right].$$

Proof. Given any $\gamma \in \Gamma(\widehat{\mathbb{P}}, \mathbb{P})$, we have

$$\begin{aligned}
\mathbb{E}_{(\widehat{X},X)\sim\gamma}\left[\|\widehat{X} - X\|\right] &\geq \mathbb{E}_{(\widehat{X},X)\sim\gamma}\left[\|X - \widehat{X}\|\mathbf{1}\{X \geq \alpha\}\right] \\
&\geq \mathbb{E}_{(\widehat{X},X)\sim\gamma}\left[(\alpha - \widehat{X})\mathbf{1}\{X \geq \alpha\}\right] \\
&= \alpha\mathbb{P}(X \geq \alpha) - \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\widehat{X}\mathbb{E}_{X\sim\gamma_{X|\widehat{X}}}\left[\mathbf{1}\{X \geq \alpha\} \mid \widehat{X}\right]\right] \\
&= \mathbb{P}(X \geq \alpha)\left(\alpha - \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\widehat{X}\cdot\frac{\mathbb{P}(X \geq \alpha \mid \widehat{X})}{\mathbb{P}(X \geq \alpha)}\right]\right)
\end{aligned}$$

Note that $\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\frac{\mathbb{P}(X \geq \alpha \mid \widehat{X})}{\mathbb{P}(X \geq \alpha)}\right] = 1$, and $\frac{\mathbb{P}(X \geq \alpha \mid \widehat{X})}{\mathbb{P}(X \geq \alpha)} \leq \frac{1}{\beta}$. Therefore

$$\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\widehat{X}\cdot\frac{\mathbb{P}(X \geq \alpha \mid \widehat{X})}{\mathbb{P}(X \geq \alpha)}\right] \leq \sup_{\widehat{\mathbb{Q}}\in\mathcal{P}(\mathcal{X})}\left\{\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{Q}}}[\widehat{X}] : \widehat{\mathbb{Q}} \ll \widehat{\mathbb{P}}, \frac{d\widehat{\mathbb{Q}}}{d\widehat{\mathbb{P}}} \leq \frac{1}{\beta}\right\} = \mathbb{CV@R}^{\widehat{\mathbb{P}}}_{\beta}(\widehat{X}).$$

Here we used the dual formulation for CVaR in [9]. Hence, we have shown that

$$\alpha - \mathbb{CV@R}^{\widehat{\mathbb{P}}}_{\beta}(\widehat{X}) \leq \frac{1}{\mathbb{P}(X \geq \alpha)}\mathbb{E}_{(\widehat{X},X)\sim\gamma}\left[\|\widehat{X} - X\|\right] \leq \frac{1}{\beta}\mathbb{E}_{(\widehat{X},X)\sim\gamma}\left[\|\widehat{X} - X\|\right].$$

The proof of the upper bound is completed by taking infimum over $\gamma \in \Gamma(\widehat{\mathbb{P}}, \mathbb{P})$. The proof of the lower bound is similar:

$$\begin{aligned}
\mathbb{E}_{(\widehat{X},X)\sim\gamma}\left[\|\widehat{X} - X\|\right] &\geq \mathbb{E}_{(\widehat{X},X)\sim\gamma}\left[\|\widehat{X} - X\|\mathbf{1}\{X \leq \alpha\}\right] \\
&\geq \mathbb{E}_{(\widehat{X},X)\sim\gamma}\left[(\widehat{X} - \alpha)\mathbf{1}\{X \leq \alpha\}\right] \\
&= \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\widehat{X}\mathbb{E}_{X\sim\gamma_{X|\widehat{X}}}\left[\mathbf{1}\{X \leq \alpha\} \mid \widehat{X}\right]\right] - \alpha\mathbb{P}(X \leq \alpha) \\
&= \mathbb{P}(X \leq \alpha)\left(\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\widehat{X}\cdot\frac{\mathbb{P}(X \leq \alpha \mid \widehat{X})}{\mathbb{P}(X \leq \alpha)}\right] - \alpha\right) \\
&\geq \mathbb{P}(X \leq \alpha)\left(-\mathbb{CV@R}^{\widehat{\mathbb{P}}}_{1-\beta}(-\widehat{X}) - \alpha\right).
\end{aligned} \qquad \square$$

Proof of Example 8 ($\mathbb{C}$V@R*).* Given $\widehat{Z} \sim \widehat{\mathbb{Q}}$ and $Z \sim \mathbb{Q}$, define $\widehat{X} = b^\top \widehat{Z}$ and $X = b^\top Z$, and let $\widehat{\mathbb{P}}$, $\mathbb{P}$ denote the law of $\widehat{X}$ and X respectively. We observe the following

(a) $\mathbb{C}\text{V@R}^{\mathbb{Q}}_\beta(b^\top Z) = \mathbb{C}\text{V@R}^{\mathbb{P}}_\beta(X)$ and $\mathbb{C}\text{V@R}^{\widehat{\mathbb{Q}}}_\beta(b^\top \widehat{Z}) = \mathbb{C}\text{V@R}^{\widehat{\mathbb{P}}}_\beta(\widehat{X})$.

(b) For any $\mathbb{Q} \in \mathcal{P}(\mathcal{Z})$, $\mathcal{W}_p(\widehat{\mathbb{P}}, \mathbb{P}) \le \|b\|_* \mathcal{W}_p(\widehat{\mathbb{Q}}, \mathbb{Q})$.

(c) For any $\mathbb{P}' \in \mathcal{P}(\mathbb{R})$, we can find $\mathbb{Q} \in \mathcal{P}(\mathcal{Z})$ such that $\mathbb{P} = \mathbb{P}'$, and $\mathcal{W}_p(\widehat{\mathbb{P}}, \mathbb{P}) \ge \|b\|_* \mathcal{W}_p(\widehat{\mathbb{Q}}, \mathbb{Q})$.

If all these claims are true, then

$$\sup_{\mathbb{Q}\in\mathcal{P}(\mathcal{Z})} \left\{ \mathbb{C}\text{V@R}^{\mathbb{Q}}_\beta(b^\top Z) : \mathcal{W}_p(\widehat{\mathbb{Q}}, \mathbb{Q}) \le \rho \right\} = \sup_{\mathbb{P}\in\mathcal{P}(\mathbb{R})} \left\{ \mathbb{C}\text{V@R}^{\mathbb{P}}_\beta(X) : \mathcal{W}_p(\widehat{\mathbb{P}}, \mathbb{P}) \le \|b\|_* \rho \right\}.$$

The first two claims are direct: $X = b^\top Z$, $\widehat{X} = b^\top \widehat{Z}$, and $|X - \widehat{X}| \le \|b\|_* \|Z - \widehat{Z}\|$. For the third claim, we prove it as follows. Let $b^* \in \mathcal{Z}$ be the unit dual of $b^\top$, i.e. $b^\top b^* = \|b^\top\|_*$ and $\|b^*\| = 1$. Given $\widehat{Z} \sim \widehat{\mathbb{P}}$, $\widehat{X} = b^\top X \sim \widehat{\mathbb{Q}}$, $X' \sim \mathbb{P}'$, define $Z = \widehat{Z} + (X' - \widehat{X}) b^* / \|b^\top\|_*$. Then $X = b^\top Z = X$, $\mathbb{P}' = \mathbb{P}$, and $\|\widehat{Z} - Z\| = \|b^\top\|_*^{-1} \|\widehat{X} - X\|$, thus $\mathcal{W}_p(\widehat{\mathbb{Q}}, \mathbb{Q}) \le \|b^\top\|_*^{-1} \mathcal{W}_p(\widehat{\mathbb{P}}, \mathbb{P})$. We have transformed the problem to the following form: with $p < \infty$, $c(\widehat{x}, x) = |\widehat{x} - x|^p$,

$$\sup_{\mathbb{Q}\in\mathcal{P}(\mathcal{Z})} \left\{ \mathbb{C}\text{V@R}^{\mathbb{Q}}_\beta(b^\top Z) : \mathcal{W}_p(\widehat{\mathbb{Q}}, \mathbb{Q}) \le \rho \right\} = \sup_{\mathbb{P}\in\mathcal{P}(\mathcal{X})} \left\{ J(\mathbb{P}) : \mathcal{K}_c(\widehat{\mathbb{P}}, \mathbb{P}) \le (\|b^\top\|_* \rho)^p \right\} = \mathcal{L}_{\mathrm{J}}((\|b^\top\|_* \rho)^p),$$

and with $p = \infty$, $c(\widehat{x}, x) = |\widehat{x} - x|^p$,

$$\sup_{\mathbb{Q}\in\mathcal{P}(\mathcal{Z})} \left\{ \mathbb{C}\text{V@R}^{\mathbb{Q}}_\beta(b^\top Z) : \mathcal{W}_p(\widehat{\mathbb{Q}}, \mathbb{Q}) \le \rho \right\} = \sup_{\mathbb{P}\in\mathcal{P}(\mathcal{X})} \left\{ J(\mathbb{P}) : \overline{\mathcal{K}}_c(\widehat{\mathbb{P}}, \mathbb{P}) \le \|b^\top\|_* \rho \right\} = \overline{\mathcal{L}}_{\mathrm{J}}(\|b^\top\|_* \rho).$$

For simplicity, assume $\|b^*\| = 1$ from now on.

To apply Theorem 3, we verify the following prerequisites:

- f_α satisfies Assumption 1: $f_\alpha \ge \alpha$ so $\mathbb{E}_{\widehat{\mathbb{P}}}[f_\alpha] \ge \alpha > -\infty$.
- f_α is lower semi-continuous and convex in α: this is obvious since $f_\alpha = \max\{\alpha, \frac{1}{\beta}x + (1 - \frac{1}{\beta})\alpha\}$ is the maximum of two affine functions.
- $\inf_{\alpha\in A', x\in\mathcal{X}} f_\alpha(x) > -\infty$ for compact $A' \subset \mathbb{R}$: $\inf_{\alpha\in A', x\in\mathcal{X}} f_\alpha(x) = \inf_{\alpha\in A'} \alpha = \min_{\alpha\in A'} \alpha > -\infty$, since A' is compact and bounded.
- $f_\alpha - \lambda c$ satisfies (IP): this is because $\mathcal{X}$ is a Euclidean space, which is complete and separable.
- (5) holds for $\mathbb{P}$ in Wasserstein ball: from [17] we know that for CVaR problem, the minimum of (4) is attained on a nonempty closed bounded interval $\alpha \in [\underline{\alpha}, \overline{\alpha}]$ (possibly a singleton). This interval contains α such that $\mathbb{P}(X \ge \alpha) \ge \beta$ and $\mathbb{P}(X \le \alpha) \ge 1 - \beta$. By Lemma EC.2, this interval is contained in

$$A' = \left[-\mathbb{C}\text{V@R}^{\widehat{\mathbb{P}}}_{1-\beta}(-\widehat{X}) - \frac{1}{1-\beta}\mathcal{W}_1(\widehat{\mathbb{P}}, \mathbb{P}), \mathbb{C}\text{V@R}^{\widehat{\mathbb{P}}}_\beta(\widehat{X}) + \frac{1}{\beta}\mathcal{W}_1(\widehat{\mathbb{P}}, \mathbb{P}) \right].$$

Since $\mathcal{W}_1(\widehat{\mathbb{P}}, \mathbb{P}) \le \mathcal{W}_p(\widehat{\mathbb{P}}, \mathbb{P})$ for $p \in [1, \infty]$, we have verified all the prerequisites of Theorem 3.

When $p = \infty$, by Theorem 3, we conclude

$$\overline{\mathcal{L}}_{\mathrm{J}}(\rho) = \sup_{\mathbb{P}\in\mathcal{P}} \left\{ \mathbb{C}\text{V@R}^{\mathbb{P}}_\beta(X) : \mathcal{W}_\infty(\widehat{\mathbb{P}}, \mathbb{P}) \le \rho \right\} = \inf_{\alpha\in A} \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}} \left[\sup_x \left\{ f_\alpha(x) : |\widehat{X} - x| \le \rho \right\} \right],$$

where

$$\sup_{x}\left\{f_\alpha(x):|\widehat{X}-x|\le\rho\right\}=\sup_{x}\left\{\alpha+\frac{1}{1-\beta}(x-\alpha)_+:|\widehat{X}-x|\le\rho\right\}=\alpha+\frac{1}{1-\beta}(\widehat{X}+\rho-\alpha)_+.$$

We thus conclude that

$$\overline{\mathcal{L}}_{\mathrm{J}}(\rho)=\inf_{\alpha\in A}\left\{\alpha+\frac{1}{1-\beta}\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}[(\widehat{X}+\rho-\alpha)_+]\right\}=\mathbb{C}\mathrm{V@R}_{\beta}^{\widehat{\mathbb{P}}}(\widehat{X})+\rho.$$

When $p=1$,

$$\sup_{x\in\mathcal{X}}\{f_\alpha(x)-\lambda c(\widehat{x},x)\}=\sup_{x\in\mathcal{X}}\left\{\alpha+\frac{1}{1-\beta}(x-\alpha)_+-\lambda|\widehat{x}-x|\right\}=\begin{cases}\alpha+\dfrac{1}{1-\beta}(\widehat{x}-\alpha)_+ & \lambda\ge\frac{1}{1-\beta}\\ \infty & \lambda<\frac{1}{1-\beta}\end{cases}.$$

Therefore for $\lambda\ge\frac{1}{1-\beta}$,

$$\inf_{\alpha\in A}\left\{\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_{x\in\mathcal{X}}\left\{f_\alpha(x)-\lambda c(\widehat{X},x)\right\}\right]\right\}=\inf_{\alpha\in A}\left\{\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\alpha+\frac{1}{1-\beta}(\widehat{X}-\alpha)_+\right]\right\}=\mathbb{C}\mathrm{V@R}_{\beta}^{\widehat{\mathbb{P}}}(\widehat{X}).$$

Thus

$$\mathcal{L}_{\mathrm{J}}(\rho)=\inf_{\alpha\in A,\lambda\ge0}\left\{\lambda\rho+\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_{x\in\mathcal{X}}\left\{f_\alpha(x)-\lambda c(\widehat{X},x)\right\}\right]\right\}=\mathbb{C}\mathrm{V@R}_{\beta}^{\widehat{\mathbb{P}}}(\widehat{X})+\frac{\rho}{1-\beta}.$$

When $p>1$,

$$\begin{aligned}\sup_{x\in\mathcal{X}}\{f_\alpha(x)-\lambda|\widehat{x}-x|^p\}&=\sup_{x\in\mathcal{X}}\left\{\alpha+\frac{1}{1-\beta}(x-\alpha)_+-\lambda|\widehat{x}-x|^p\right\}\\&=\sup_{x\in\mathcal{X}}\left\{\alpha+\frac{1}{1-\beta}(x-\alpha)-\lambda|\widehat{x}-x|^p\right\}\vee\sup_{x\in\mathcal{X}}\{\alpha-\lambda|\widehat{x}-x|^p\}\\&=\left(\alpha+\frac{1}{1-\beta}(\widehat{x}-\alpha)+\sup_{t\in\mathbb{R}}\left\{\frac{t}{1-\beta}-\lambda|t|^p\right\}\right)\vee\alpha\\&=\alpha+\left(\frac{1}{1-\beta}(\widehat{x}-\alpha)+C\lambda^{-\frac{1}{p-1}}\right)_+\\&=\alpha+\left(\frac{1}{1-\beta}\left(\widehat{x}-(\alpha-C(1-\beta)\lambda^{-\frac{1}{p-1}}\right)\right)_+,\end{aligned}$$

where $C=(p-1)(p(1-\beta))^{-\frac{p}{p-1}}$. Thus

$$\inf_{\alpha\in A}\left\{\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_{x\in\mathcal{X}}\left\{f_\alpha(x)-\lambda c(\widehat{X},x)\right\}\right]\right\}=\mathbb{C}\mathrm{V@R}_{\beta}^{\widehat{\mathbb{P}}}(\widehat{X})+C(1-\beta)\lambda^{-\frac{1}{p-1}}.$$

Therefore

$$\mathcal{L}_{\mathrm{J}}(\rho^p)=\mathbb{C}\mathrm{V@R}_{\beta}^{\widehat{\mathbb{P}}}(\widehat{X})+\min_{\lambda\ge0}\left\{\lambda\rho^p+C(1-\beta)\lambda^{-\frac{1}{p-1}}\right\}=\mathbb{C}\mathrm{V@R}_{\beta}^{\widehat{\mathbb{P}}}(\widehat{X})+\rho(1-\beta)^{-\frac{1}{p}}.$$

In conclusion, for $p\in[1,\infty]$, it holds that

$$\sup_{\mathbb{P}\in\mathcal{P}}\left\{\mathbb{C}\mathrm{V@R}_{\beta}^{\mathbb{P}}(X):\mathcal{W}_p(\widehat{\mathbb{P}},\mathbb{P})\le\rho\right\}=\mathbb{C}\mathrm{V@R}_{\beta}^{\widehat{\mathbb{P}}}(\widehat{X})+\rho(1-\beta)^{-\frac{1}{p}}.\qquad\square$$

Proof of Example 8 ($\mathbb{V}$ar). Same as Example 8 $\mathbb{C}$V@R, we can reduce to a one-dimensional problem and assume without loss of generality that $\|b^\top\|_* = 1$.

It is well-known that the optimal α is the expectation:

$$\mathbb{V}\mathrm{ar}^{\mathbb{P}}(X) = \min_{\alpha} \mathbb{E}[(X-\alpha)^2] = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

Given $\widehat{\mathbb{P}}, \mathbb{P}$ and a transport plan $\gamma \in \Gamma(\widehat{\mathbb{P}}, \mathbb{P})$, the transport cost

$$\mathbb{E}_\gamma[\|\widehat{X} - X\|] \geq |\mathbb{E}_\gamma[\widehat{X} - X]| \geq |\mathbb{E}_{\widehat{\mathbb{P}}}[\widehat{X}] - \mathbb{E}_{\mathbb{P}}[X]|.$$

Minimizing over all $\gamma \in \Gamma(\widehat{\mathbb{P}}, \mathbb{P})$ gives

$$\mathbb{E}_{X\sim\mathbb{P}}[X] \leq \left[\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}[\widehat{X}] - \mathcal{W}_1(\widehat{X}, X), \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}[\widehat{X}] + \mathcal{W}_1(\widehat{X}, X)\right].$$

Same as before, we can verify that $f_\alpha(x) = (x-\alpha)^2$ meets all the prerequisites of Theorem 3.

When $p = \infty$, by Theorem 3, we conclude

$$\overline{\mathcal{L}}_{\mathrm{J}}(\rho) = \sup_{\mathbb{P}\in\mathcal{P}} \left\{\mathbb{V}\mathrm{ar}^{\mathbb{P}}(X) : \mathcal{W}_\infty(\widehat{\mathbb{P}}, \mathbb{P}) \leq \rho\right\} = \inf_{\alpha\in A} \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_x \left\{f_\alpha(x) : |\widehat{X} - x| \leq \rho\right\}\right],$$

where

$$\sup_x \left\{f_\alpha(x) : |\widehat{X} - x| \leq \rho\right\} = \sup_x \left\{(x-\alpha)^2 : |\widehat{X} - x| \leq \rho\right\} = (|\widehat{X} - \alpha| + \rho)^2.$$

We thus conclude that

$$\overline{\mathcal{L}}_{\mathrm{J}}(\rho) = \inf_{\alpha\in A} \left\{\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[(|\widehat{X} - \alpha| + \rho)^2\right]\right\}.$$

When $1 \leq p < 2$, for any $\lambda \geq 0$, it holds that

$$\sup_{x\in\mathcal{X}} \{f_\alpha(x) - \lambda c(\widehat{x}, x)\} = \sup_{x\in\mathcal{X}} \left\{(x-\alpha)^2 - \lambda|\widehat{x} - x|^p\right\} = +\infty.$$

Thus for any $\rho > 0$ we must have

$$\mathcal{L}_{\mathrm{J}}(\rho^p) = +\infty.$$

When $p = 2$,

$$\begin{aligned}\sup_{x\in\mathcal{X}} \left\{f_\alpha(x) - \lambda|\widehat{x} - x|^2\right\} &= \sup_{x\in\mathcal{X}} \left\{(x-\alpha)^2 - \lambda(\widehat{x} - x)^2\right\} \\ &= \begin{cases} +\infty & 0 \leq \lambda < 1, \text{ or } \lambda = 1, \widehat{x} \neq \alpha, \text{ or} \\ 0 & \lambda = 1, \widehat{x} = \alpha \\ \frac{\lambda}{\lambda-1}(\widehat{x} - \alpha)^2 & \lambda > 1. \end{cases}\end{aligned}$$

Thus

$$\inf_{\alpha\in A}\left\{\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_{x\in\mathcal{X}}\left\{f_\alpha(x)-\lambda c(\widehat{X},x)\right\}\right]\right\}=\begin{cases}+\infty & 0\le\lambda<1,\text{ or }\lambda=1,\widehat{\mathbb{P}}\neq\boldsymbol{\delta}_{\widehat{x}}\text{ for any }\widehat{x}\in\mathcal{X}\\ 0 & \lambda=1,\widehat{\mathbb{P}}=\boldsymbol{\delta}_{\widehat{x}}\text{ for some }\widehat{x}\in\mathcal{X}\\ \frac{\lambda}{\lambda-1}\mathbb{V}\mathrm{ar}^{\widehat{\mathbb{P}}}(\widehat{X}) & \lambda>1.\end{cases}$$

We now conclude

$$\mathcal{L}_{\mathrm{J}}(\rho^2)=\inf_{\lambda>1}\left\{\lambda\rho^2+\frac{\lambda}{\lambda-1}\mathbb{V}\mathrm{ar}^{\widehat{\mathbb{P}}}(\widehat{X})\right\}=(\mathbb{V}\mathrm{ar}^{\widehat{\mathbb{P}}}(\widehat{X})^{\frac{1}{2}}+\rho)^2. \qquad \square$$

Proof of Example 8 ($\mathbb{MAD}$*).* Same as Example 8 $\mathbb{CV@R}$, we can reduce to a one-dimensional problem and assume without loss of generality that $\|b^\top\|_*=1$.

It is well-known that the optimal α is the median. By Lemma EC.2, the median of $\mathbb{P}$ inside the Wasserstein uncertainty set is attained in

$$\alpha\in\left[-\mathbb{CV@R}^{\widehat{\mathbb{P}}}_{\frac{1}{2}}(-\widehat{X})-2\mathcal{W}_1(\widehat{\mathbb{P}},\mathbb{P}),\mathbb{CV@R}^{\widehat{\mathbb{P}}}_{\frac{1}{2}}(\widehat{X})+2\mathcal{W}_1(\widehat{\mathbb{P}},\mathbb{P})\right].$$

Same as in Example 8 $\mathbb{CV@R}$, we can verify that $f_\alpha(x)=|x-\alpha|$ meets all the prerequisites of Theorem 3.

When $p=\infty$, by Theorem 3, we conclude

$$\overline{\mathcal{L}}_{\mathrm{J}}(\rho)=\sup_{\mathbb{P}\in\mathcal{P}}\left\{\mathbb{MAD}^{\mathbb{P}}(X):\mathcal{W}_\infty(\widehat{\mathbb{P}},\mathbb{P})\le\rho\right\}=\inf_{\alpha\in A}\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_{x}\left\{f_\alpha(x):|\widehat{X}-x|\le\rho\right\}\right],$$

where

$$\sup_{x}\left\{f_\alpha(x):|\widehat{X}-x|\le\rho\right\}=\sup_{x}\left\{|x-\alpha|:|\widehat{X}-x|\le\rho\right\}=|\widehat{X}-\alpha|+\rho.$$

We thus conclude that

$$\overline{\mathcal{L}}_{\mathrm{J}}(\rho)=\inf_{\alpha\in A}\left\{\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}[|\widehat{X}-\alpha|]+\rho\right\}=\mathbb{MAD}^{\widehat{\mathbb{P}}}(\widehat{X})+\rho.$$

When $p=1$,

$$\sup_{x\in\mathcal{X}}\{f_\alpha(x)-\lambda c(\widehat{x},x)\}=\sup_{x\in\mathcal{X}}\{|x-\alpha|-\lambda|\widehat{x}-x|\}=\begin{cases}|\widehat{x}-\alpha| & \lambda\ge1\\ \infty & \lambda<1\end{cases}.$$

Therefore for $\lambda\ge1$,

$$\inf_{\alpha\in A}\left\{\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_{x\in\mathcal{X}}\left\{f_\alpha(x)-\lambda c(\widehat{X},x)\right\}\right]\right\}=\inf_{\alpha\in A}\left\{\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[|\widehat{X}-\alpha|\right]\right\}=\mathbb{MAD}^{\widehat{\mathbb{P}}}(\widehat{X}).$$

Thus

$$\mathcal{L}_{\mathrm{J}}(\rho)=\inf_{\alpha\in A,\lambda\ge0}\left\{\lambda\rho+\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_{x\in\mathcal{X}}\left\{f_\alpha(x)-\lambda c(\widehat{X},x)\right\}\right]\right\}=\mathbb{MAD}^{\widehat{\mathbb{P}}}(\widehat{X})+\rho.$$

Robust loss for $p=1$ and $p=\infty$ are both $\mathbb{MAD}^{\widehat{\mathbb{P}}}(\widehat{X})+\rho$. As $\mathcal{W}_1(\widehat{\mathbb{P}},\mathbb{P})\le\mathcal{W}_p(\widehat{\mathbb{P}},\mathbb{P})\le\mathcal{W}_\infty(\widehat{\mathbb{P}},\mathbb{P})$, we have for any $1\le p\le\infty$:

$$\sup_{\mathbb{P}\in\mathcal{P}}\left\{\mathbb{CV@R}^{\mathbb{P}}_\beta(X):\mathcal{W}_p(\widehat{\mathbb{P}},\mathbb{P})\le\rho\right\}=\mathbb{MAD}^{\widehat{\mathbb{P}}}(\widehat{X})+\rho. \qquad \square$$

Proof of Example 8 ($\mathbb{E}$nt*).* Same as Example 8 $\mathbb{C}$V@R, we can reduce to a one-dimensional problem and assume without loss of generality that $\|b^\top\|_* = 1$.

$\alpha \mapsto \mathbb{E}[f_\alpha(X)] = \alpha + \frac{1}{\theta}\left(\mathbb{E}[e^{\theta(X-\alpha)}] - 1\right)$ is convex, and $\lim_{\alpha\to\pm\infty} \mathbb{E}[f_\alpha(X)] = +\infty$, so the minimizer α^* satisfies

$$0 = \frac{d}{d\alpha}\bigg|_{\alpha=\alpha^*} \mathbb{E}[f_\alpha(X)] = 1 - \mathbb{E}[e^{\theta(X-\alpha^*)}].$$

Therefore, $e^{\theta\alpha^*} = \mathbb{E}[e^{\theta X}]$. Suppose $\widehat{\alpha}^*$ is the minimizer to $\mathbb{E}[f_\alpha(\widehat{X})]$, and γ-$\operatorname{ess\,sup}_{\widehat{x},x} \|\widehat{x} - x\| \le \rho$, then

$$e^{\theta\alpha^*} = \mathbb{E}[e^{\theta X}] \le \mathbb{E}[e^{\theta(\widehat{X}+\rho)}] = \mathbb{E}[e^{\theta\widehat{X}}]e^{\theta\rho} = e^{\theta(\widehat{\alpha}^*+\rho)}.$$

Hence $\alpha^* \le \widehat{\alpha}^* + \rho$. Similarly, $\alpha^* \ge \widehat{\alpha}^* - \rho$. Therefore,

$$\alpha \in [\widehat{\alpha}^* - \mathcal{W}_\infty(\widehat{\mathbb{P}}, \mathbb{P}), \widehat{\alpha}^* + \mathcal{W}_\infty(\widehat{\mathbb{P}}, \mathbb{P})].$$

By Theorem 3, we conclude for $p = \infty$:

$$\begin{aligned}
\overline{\mathcal{L}}_{\mathrm{J}}(\rho) &= \inf_{\alpha\in\mathbb{R}} \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_{x\in\mathbb{R}}\left\{\alpha + \frac{1}{\theta}\left(e^{\theta(x-\alpha)} - 1\right) : |\widehat{X} - x| \le \rho\right\}\right] \\
&= \inf_{\alpha\in\mathbb{R}} \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\alpha + \frac{1}{\theta}\left(e^{\theta(\widehat{X}+\rho-\alpha)} - 1\right) - \alpha\right] \\
&= \frac{1}{\theta}\log(\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}[e^{\theta\widehat{X}}]) + \rho = \mathbb{E}\mathrm{nt}_\theta^{\widehat{\mathbb{P}}}(\widehat{X}) + \rho.
\end{aligned}$$

For $p < \infty$, we verify $\mathcal{L}_{\mathrm{J}}(\rho) = +\infty$ directly. We define $\mathbb{P}_\epsilon = (1-\epsilon)\widehat{\mathbb{P}} + \epsilon\widehat{\mathbb{P}}_{M_\epsilon}$, where $\widehat{\mathbb{P}}_M = (x \mapsto x + M)_\#\widehat{\mathbb{P}}$ is right-translation of $\widehat{\mathbb{P}}$ by M, and $M_\epsilon = \rho\epsilon^{-1/p}$. Then $\mathcal{W}_p(\widehat{\mathbb{P}}, \mathbb{P}) \le \rho$. However,

$$\mathbb{E}_{\mathbb{P}_\epsilon}[e^{\theta X}] = (1-\epsilon)\mathbb{E}[e^{\theta\widehat{X}}] + \epsilon\mathbb{E}[e^{\theta(\widehat{X}+M_\epsilon)}] = (1 - \epsilon + \epsilon e^{\theta M_\epsilon})\mathbb{E}[e^{\theta\widehat{X}}],$$

and accordingly

$$\mathbb{E}\mathrm{nt}_\theta^{\mathbb{P}_\epsilon}(X) = \frac{1}{\theta}\log(\mathbb{E}_{X\sim\mathbb{P}_\epsilon}[e^{\theta X}]) = \frac{1}{\theta}\log(\mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}[e^{\theta\widehat{X}}]) + \frac{1}{\theta}\log(1 + \epsilon(e^{\theta\epsilon^{-1/p}} - 1)),$$

which tends to infinity as $\epsilon \to 0$. □

Appendix EC.6: Proofs for Section 5.3

Proof of Proposition 4. For a fixed θ, first we apply Theorem 1 to $-\mathcal{L}_{\mathrm{G}}(\cdot, \theta)$ by taking the Fenchel conjugate

$$\begin{aligned}
(-\mathcal{L}_{\mathrm{G}}(\cdot,\theta))^*(-\lambda) &= \sup_{\mathbb{P},\widetilde{\mathbb{P}}\in\mathcal{P}(\mathcal{X})}\left\{\mathbb{E}_{X\sim\mathbb{P}}[f(X)] - \lambda\mathcal{K}_c(\widetilde{\mathbb{P}}, \mathbb{P}) : \mathcal{K}_{\tilde{c}}(\widehat{\mathbb{P}}, \widetilde{\mathbb{P}}) \le \theta\right\} \\
&= \sup_{\widetilde{\mathbb{P}}\in\mathcal{P}(\mathcal{X})}\left\{\mathbb{E}_{\widetilde{X}\sim\widetilde{\mathbb{P}}}\left[\sup_{x\in\mathcal{X}}\left\{f(x) - \lambda c(\widetilde{X}, x)\right\}\right] : \mathcal{K}_{\tilde{c}}(\widehat{\mathbb{P}}, \widetilde{\mathbb{P}}) \le \theta\right\}.
\end{aligned}$$

Denote $\tilde{f}(\tilde{x}) = \sup_{x\in\mathcal{X}} f(x) - \lambda c(\tilde{x}, x)$. Then we apply Theorem 1 to $\tilde{f}, \tilde{c}$ and $-(-\mathcal{L}_{\mathrm{G}}(\cdot,\theta))^*(-\lambda)$ by taking Fenchel conjugate $\theta \to -\mu$:

$$\begin{aligned}\mathcal{L}_{\mathrm{G}*}^{*}(-\lambda,-\mu) &= \sup_{\widetilde{\mathbb{P}}\in\mathcal{P}(\mathcal{X})} \left\{ \mathbb{E}_{\widetilde{X}\sim\mathbb{P}}\left[\sup_{x\in\mathcal{X}}\left\{ f(x) - \lambda c(\widetilde{X}, x)\right\}\right] - \mu\mathcal{K}_{\tilde{c}}(\widehat{\mathbb{P}}, \widetilde{\mathbb{P}})\right\} \\ &= \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_{x,\tilde{x}\in\mathcal{X}}\left\{ f(x) - \lambda c(\tilde{x},x) - \mu\tilde{c}(\widehat{X},\tilde{x})\right\}\right].\end{aligned}$$

Since $(-\mathcal{L}_{\mathrm{G}}(\cdot,\theta))^*(-\lambda)$ is concave in θ, we recover it by

$$\begin{aligned}(-\mathcal{L}_{\mathrm{G}}(\cdot,\theta))^*(-\lambda) &= \min_{\mu\geq 0}\left\{\mu\theta + (-\mathcal{L}_{\mathrm{G}})_*^*(-\lambda,-\mu)\right\} \\ &= \min_{\mu\geq 0}\left\{\mu\theta + \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_{x,\tilde{x}\in\mathcal{X}}\left\{ f(x) - \lambda c(\tilde{x},x) - \mu\tilde{c}(\widehat{X},\tilde{x})\right\}\right]\right\}.\end{aligned}$$

Since $\mathcal{L}_{\mathrm{G}}(\rho,\theta)$ is concave in ρ, we recover it by

$$\mathcal{L}_{\mathrm{G}}(\rho,\theta) = \min_{\lambda,\mu\geq 0}\left\{\lambda\rho + \mu\theta + \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_{x,\tilde{x}\in\mathcal{X}}\left\{ f(x) - \lambda c(\tilde{x},x) - \mu\tilde{c}(\widehat{X},\tilde{x})\right\}\right]\right\}.$$

In particular, if $c(x_1,x_2) = \tilde{c}(x_1,x_2) = d(x_1,x_2)$ are the same metric, then

$$\begin{aligned}\mathcal{L}_{\mathrm{G}*}^{*}(-\lambda,-\mu) &= \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_{x,\tilde{x}\in\mathcal{X}}\left\{ f(x) - \lambda d(\tilde{x},x) - \mu d(\widehat{X},\tilde{x})\right\}\right] \\ &= \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_{x\in\mathcal{X}}\left\{ f(x) - (\lambda\wedge\mu) d(\widehat{X},x)\right\}\right]\end{aligned}$$

by taking $\tilde{x} = x$ when $\lambda \geq \mu$ and $\tilde{x} = \widehat{X}$ when $\lambda \leq \mu$. Correspondingly,

$$\begin{aligned}(-\mathcal{L}_{\mathrm{G}})^*(-\lambda,\theta) &= \min_{\mu\geq 0}\left\{\mu\theta + \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_{x\in\mathcal{X}}\left\{ f(x) - (\lambda\wedge\mu) d(\widehat{X},x)\right\}\right]\right\} \\ &= \min_{\mu\in[0,\lambda]}\left\{\mu\theta + \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_{x\in\mathcal{X}}\left\{ f(x) - \mu d(\widehat{X},x)\right\}\right]\right\},\end{aligned}$$

and

$$\begin{aligned}\mathcal{L}_{\mathrm{G}}(\rho,\theta) &= \min_{0\leq\mu\leq\lambda}\left\{\lambda\rho + \mu\theta + \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_{x\in\mathcal{X}}\left\{ f(x) - \mu d(\widehat{X},x)\right\}\right]\right\} \\ &= \min_{\mu\geq 0}\left\{\mu(\rho+\theta) + \mathbb{E}_{\widehat{X}\sim\widehat{\mathbb{P}}}\left[\sup_{x\in\mathcal{X}}\left\{ f(x) - \mu d(\widehat{X},x)\right\}\right]\right\}. \qquad \square\end{aligned}$$