Supplementary Material for “Regularized Policies are Reward Robust”

1 Proofs of Main Results

We first introduce some notation that will be used exclusively for the Appendix. For any function $R : \mathscr{B}(\mathcal{X}) \to \mathbb{R}$, we define $R_+(\mu) = R(\mu) + \iota_{\mathscr{P}}(\mu)$ and $R_-(\mu) = R(\mu) - \iota_{\mathscr{P}}(\mu)$. Indeed, it should noted that if R is upper semi-continuous concave then R_- is upper semi-continuous concave and $-R_-$ is proper convex. The central benefit of rewriting R in this is way is due to

$$\sup_{\mu \in \mathcal{K}_{P,\gamma}} R(\mu) = \sup_{\mu \in \mathcal{K}_{P,\gamma}} R_-(\mu).$$

First we will show a technical result.

Lemma 1 *If $R : \mathscr{B}(\mathcal{X}) \to \mathbb{R}$ is upper semicontinuous and concave then $(-R_-)^\star$ is increasing.*

Proof Let $r, r' \in \mathcal{F}_b(\mathcal{X})$ such that $r \leq r'$ and let

$$\nu \in \operatorname*{arg\,sup}_{\mu \in \mathscr{P}(\mathcal{X})} \left(\int_{\mathcal{X}} r(x) d\mu(x) + R(\mu) \right),$$

noting that ν exists since the mapping $\mu \mapsto \int_{\mathcal{X}} r(x) d\mu(x) + R(\mu)$ is concave, upper semicontinuous and $\mathscr{P}(\mathcal{X})$ is compact. Next we have

$$\begin{aligned}
&(-R_-)^\star(r) - (-R_-)^\star(r') \\
&= \sup_{\mu \in \mathscr{P}(\mathcal{X})} \left(\int_{\mathcal{X}} r(x) d\mu(x) + R(\mu) \right) - \sup_{\mu \in \mathscr{P}(\mathcal{X})} \left(\int_{\mathcal{X}} r'(x) d\mu(x) + R(\mu) \right) \\
&\leq \int_{\mathcal{X}} r(x) d\nu(x) + R(\nu) - \int_{\mathcal{X}} r'(x) d\nu(x) - R(\nu) \\
&= \int_{\mathcal{X}} \left(r(x) - r'(x) \right) d\nu(x) \\
&\leq 0
\end{aligned}$$

■

We also recall some classical results regarding Fenchel duality between the spaces $\mathcal{F}_b(\mathcal{X})$ and $\mathscr{B}(\mathcal{X})$.

Definition 1 (Rockafellar (1968)) *For any proper convex function $F : \mathcal{F}_b(\mathcal{X}) \to (-\infty, \infty]$ and $\mu \in \mathscr{B}(\mathcal{X})$ we define*

$$F^\star(\mu) = \sup_{h \in \mathcal{F}_b} \left(\int_{\mathcal{X}} h d\mu - F(h) \right)$$

and for any $h \in \mathcal{F}_b(\Omega)$ we define

$$F^{\star\star}(h) = \sup_{\mu \in \mathscr{B}(\mathcal{X})} \left(\int_{\mathcal{X}} h d\mu - F^\star(\mu) \right).$$

Theorem 1 (Zalinescu (2002) Theorem 2.3.3) *If X is a Hausdorff locally convex space, and $F : X \to (-\infty, \infty]$ is a proper convex lower semi-continuous function then $F^{\star\star} = F$.*

1.1 Proof of Theorem 1

We have

$$\begin{aligned}
\sup_{\mu \in \mathcal{K}_{P,\gamma}} R(\mu) &= \sup_{\mu \in \mathcal{K}_{P,\gamma}} -(-R(\mu)) \\
&\overset{(1)}{=} \sup_{\mu \in \mathcal{K}_{P,\gamma}} -(-R(\mu))^{\star\star} \\
&\overset{(2)}{=} \sup_{\mu \in \mathcal{K}_{P,\gamma}} - \sup_{r' \in \mathcal{F}_b(\mathcal{X})} \left(\int_{\mathcal{X}} r'(x) d\mu(x) - (-R)^{\star}(r') \right) \\
&= \sup_{\mu \in \mathcal{K}_{P,\gamma}} \inf_{r' \in \mathcal{F}_b(\mathcal{X})} \left(\int_{\mathcal{X}} (-r'(x))\, d\mu(x) + (-R)^{\star}(r') \right) \\
&\overset{(3)}{=} \inf_{r' \in \mathcal{F}_b(\mathcal{X})} \sup_{\mu \in \mathcal{K}_{P,\gamma}} \left(\int_{\mathcal{X}} (-r'(x))\, d\mu(x) + (-R)^{\star}(r') \right) \\
&\overset{(4)}{=} \inf_{r' \in \mathcal{F}_b(\mathcal{X})} \left(\sup_{\mu \in \mathcal{K}_{P,\gamma}} \int_{\mathcal{X}} r'(x) d\mu(x) + (-R)^{\star}(-r') \right) \\
&\overset{(5)}{=} \inf_{r' \in \mathcal{F}_b(\mathcal{X})} \left(\mathrm{RL}_{P,\gamma}(r') + (-R)^{\star}(-r') \right)
\end{aligned}$$

where (1) holds since $-R$ is proper convex, (2) is the definition of the conjugate, (3) is an application of Ky Fan's minimax theorem (Fan, 1953, Theorem 2) noting that the set $\mathcal{K}_{P,\gamma}$ is compact, and that the mapping $r \mapsto \int_{\mathcal{X}} (-r'(x))\, d\mu(x) + (-R)^{\star}(r')$ is concave and the mapping $\mu \mapsto \int_{\mathcal{X}} (-r'(x))\, d\mu(x)$ is linear. (4) holds by negating r' since $-\mathcal{F}_b(\mathcal{X}) = \mathcal{F}_b(\mathcal{X})$ and (5) holds by definition.

1.2 Proof of Theorem 2

By definition, we have $\mathrm{RL}_{P,\gamma}(r^*) - \langle r^*, \mu^* \rangle \geq 0$. To show the other direction, it follows that

$$\begin{aligned}
\mathrm{RL}_{P,\gamma}(r^*) - \langle r^*, \mu^* \rangle &= (\mathrm{RL}_{P,\gamma}(r^*) + (-R)^{\star}(-r^*)) - (\langle r^*, \mu^* \rangle + (-R)^{\star}(-r^*)) \\
&\overset{(1)}{=} \inf_{r' \in \mathcal{F}_b(\mathcal{X})} (\mathrm{RL}_{P,\gamma}(r') + (-R)^{\star}(-r')) - (\langle r^*, \mu^* \rangle + (-R)^{\star}(-r^*)) \\
&\overset{(2)}{=} \sup_{\mu \in \mathcal{K}_{P,\gamma}} R(\mu) - (\langle r^*, \mu^* \rangle + (-R)^{\star}(-r^*)) \\
&\overset{(3)}{=} R(\mu^*) - (\langle r^*, \mu^* \rangle + (-R)^{\star}(-r^*)) \\
&= \langle -r^*, \mu^* \rangle - (-R)(\mu^*) - (-R)^{\star}(-r^*) \\
&\overset{(4)}{\leq} 0,
\end{aligned}$$

where (1) follows via optimality of r^*, (2) is due to the duality result, (3) follows via optimality of μ^* and (4) is an application of the Fenchel-Young inequality on the convex function $-R$. Finally, we have $\mathrm{RL}_{P,\gamma}(r^*) = \langle r^*, \mu^* \rangle$, which implies optimality of μ^* and concludes the proof.

1.3 Proof of Theorem 3

Using the classic linear programming duality result, we have

$$\mathrm{RL}_{P,\gamma}(r) = (1-\gamma) \inf_{V \in \mathcal{V}_{P,r,\gamma}} \int_{\mathcal{S}} V(s) d\mu_0(s), \tag{1}$$

where

$$\mathcal{V}_{P,r,\gamma} = \left\{ V \in \mathcal{F}_b(\mathcal{S}) : V(s) \geq r(s,a) + \gamma \int_{\mathcal{S}} V(s') dP(s' \mid s,a), \forall (s,a) \in \mathcal{X} \right\},$$

and define

$$r_V(s,a) := V(s) - \gamma \int_{\mathcal{S}} V(s')dP(s' \mid s,a). \tag{2}$$

It then holds that

$$\begin{aligned}
\sup_{\mu \in \mathcal{K}_{P,\gamma}} R(\mu) &\overset{(1)}{=} \inf_{r' \in \mathcal{F}_b(\mathcal{X})} \left(\mathrm{RL}_{P,\gamma}(r') + (-R)^\star(-r')\right) \\
&\overset{(2)}{=} \inf_{r' \in \mathcal{F}_b(\mathcal{X})} \left((1-\gamma) \inf_{V \in \mathcal{V}_{P,r',\gamma}} \int_{\mathcal{S}} V(s)d\mu_0(s) + (-R)^\star(-r') \right) \\
&= \inf_{r' \in \mathcal{F}_b(\mathcal{X})} \inf_{V \in \mathcal{F}_b(\mathcal{S})} \left((1-\gamma) \int_{\mathcal{S}} V(s)d\mu_0(s) + (-R)^\star(-r') + \iota_{\mathcal{V}_{P,r',\gamma}}(V) \right) \\
&= \inf_{V \in \mathcal{F}_b(\mathcal{S})} \inf_{r' \in \mathcal{F}_b(\mathcal{X})} \left((1-\gamma) \int_{\mathcal{S}} V(s)d\mu_0(s) + (-R)^\star(-r') + \iota_{\mathcal{V}_{P,r',\gamma}}(V) \right) \\
&= \inf_{V \in \mathcal{F}_b(\mathcal{S})} \inf_{r' \leq r_V} \left((1-\gamma) \int_{\mathcal{S}} V(s)d\mu_0(s) + (-R)^\star(-r') \right),
\end{aligned}$$

where (1) is due to Theorem 1, (2) is due to (1) and noting that $r \leq r_V$ implies $V(s) \geq r(s,a) + \gamma \int_{\mathcal{S}} V(s')dP(s' \mid s,a)$ concludes the proof.

1.4 Proof of Lemma 1

First note that for any $\mu \in \mathcal{K}_{P,\gamma}$, we have

$$\begin{aligned}
&\int_{\mathcal{X}} r_V(s,a)d\mu(s,a) \\
&= \left(\int_{\mathcal{S}} V(s)d\mu(s,a) - \gamma \int_{\mathcal{X}} \int_{\mathcal{S}} V(s')dP(s' \mid s,a)d\mu(s,a) \right) \\
&= \left(\int_{\mathcal{S}} V(s)d\mu(s,a) - \int_{\mathcal{S}} V(s)d\mu(s,a) + (1-\gamma) \int_{\mathcal{S}} V(s)d\mu_0(s) \right) \\
&= (1-\gamma) \int_{\mathcal{S}} V(s)d\mu_0(s),
\end{aligned}$$

and so we can conclude for any $V \in \mathcal{F}_b(\mathcal{S})$, we have

$$\mathrm{RL}_{P,\gamma}(r_V) = (1-\gamma) \int_{\mathcal{S}} V(s)d\mu_0(s).$$

Next, we have

$$\begin{aligned}
\sup_{\mu \in \mathcal{K}_{P,\gamma}} R(\mu) &= \inf_{V \in \mathcal{F}_b(\mathcal{S})} \left((1-\gamma) \int_{\mathcal{S}} V(s)d\mu_0(s) + (-R)^\star(-r_V) \right) \\
&= \inf_{V \in \mathcal{F}_b(\mathcal{S})} \left(\mathrm{RL}_{P,\gamma}(r_V) + (-R)^\star(-r_V)\right) \\
&\geq \inf_{r' \in \mathcal{F}_b(\mathcal{X})} \left(\mathrm{RL}_{P,\gamma}(r') + (-R)^\star(-r')\right) \\
&= \sup_{\mu \in \mathcal{K}_{P,\gamma}} R(\mu),
\end{aligned}$$

and since the lower bound can achieve equality, it implies that the optimal r^* is of the form r_V.

1.5 Proof of Corollary 1

We have

$$\begin{aligned}
(-R)^\star(-r') &= \sup_{\mu \in \mathscr{B}(\mathcal{X})} \left(\int_{\mathcal{X}} -r'(x) d\mu(x) + R(\mu) \right) \\
&= \sup_{\mu \in \mathscr{B}(\mathcal{X})} \left(\int_{\mathcal{X}} -r'(x) d\mu(x) + \int_{\mathcal{X}} r(x) d\mu(x) - \varepsilon \Omega(\mu) \right) \\
&= \sup_{\mu \in \mathscr{B}(\mathcal{X})} \left(\int_{\mathcal{X}} r(x) - r'(x) d\mu(x) - \varepsilon \Omega(\mu) \right) \\
&= \varepsilon \sup_{\mu \in \mathscr{B}(\mathcal{X})} \left(\int_{\mathcal{X}} \frac{r(x) - r'(x)}{\varepsilon} d\mu(x) - \Omega(\mu) \right) \\
&= \varepsilon \Omega^\star \left(\frac{r - r'}{\varepsilon} \right),
\end{aligned}$$

which concludes the proof.

1.6 Proof of Theorem 4

First define the set

$$\mathcal{Q}_{P,r,\gamma} = \left\{ Q \in \mathcal{F}_b(\mathcal{X}) : Q(s,a) \geq r(s,a) + \gamma \int_{\mathcal{X}} \sup_{a' \in \mathcal{A}} Q(s', a') dP(s' \mid s, a) \right\},$$

and define

$$r_Q(s,a) = Q(s,a) - \gamma \int_{\mathcal{X}} \sup_{a' \in \mathcal{A}} Q(s', a') dP(s' \mid s, a)$$

Next we can write

$$\mathrm{RL}_{P,\gamma}(r) = \inf_{Q \in \mathcal{Q}_{P,r,\gamma}} \int_{\mathcal{S}} \sup_{a \in \mathcal{A}} Q(s,a) d\mu_0(s), \tag{A}$$

next we have

$$\begin{aligned}
\sup_{\mu \in \mathcal{K}_{P,\gamma}} R(\mu) &\overset{(1)}{=} \inf_{r' \in \mathcal{F}_b(\mathcal{X})} \left(\mathrm{RL}_{P,\gamma}(r') + (-R)^\star(-r') \right) \\
&\overset{(2)}{=} \inf_{r' \in \mathcal{F}_b(\mathcal{X})} \left(\inf_{Q \in \mathcal{Q}_{P,r',\gamma}} \int_{\mathcal{S}} \sup_{a \in \mathcal{A}} Q(s,a) d\mu_0(s) + (-R)^\star(-r') \right) \\
&= \inf_{r' \in \mathcal{F}_b(\mathcal{X})} \left(\inf_{Q \in \mathcal{F}_b(\mathcal{X})} \left(\int_{\mathcal{S}} \sup_{a \in \mathcal{A}} Q(s,a) d\mu_0(s) + \iota_{\mathcal{Q}_{P,r',\gamma}}(Q) \right) + (-R)^\star(-r') \right) \\
&= \inf_{r' \in \mathcal{F}_b(\mathcal{X})} \inf_{Q \in \mathcal{F}_b(\mathcal{X})} \left(\int_{\mathcal{S}} \sup_{a \in \mathcal{A}} Q(s,a) d\mu_0(s) + (-R)^\star(-r') + \iota_{\mathcal{Q}_{P,r',\gamma}}(Q) \right) \\
&= \inf_{Q \in \mathcal{F}_b(\mathcal{X})} \inf_{r' \in \mathcal{F}_b(\mathcal{X})} \left(\int_{\mathcal{S}} \sup_{a \in \mathcal{A}} Q(s,a) d\mu_0(s) + (-R)^\star(-r') + \iota_{\mathcal{Q}_{P,r',\gamma}}(Q) \right) \\
&= \inf_{Q \in \mathcal{F}_b(\mathcal{X})} \left(\int_{\mathcal{S}} \sup_{a \in \mathcal{A}} Q(s,a) d\mu_0(s) + \inf_{r' \in \mathcal{F}_b(\mathcal{X})} \left((-R)^\star(-r') + \iota_{\mathcal{Q}_{P,r',\gamma}}(Q) \right) \right) \\
&= \inf_{Q \in \mathcal{F}_b(\mathcal{X})} \left(\int_{\mathcal{S}} \sup_{a \in \mathcal{A}} Q(s,a) d\mu_0(s) + \inf_{r' \leq r_Q} (-R)^\star(-r') \right) \\
&\overset{(3)}{=} \inf_{Q \in \mathcal{F}_b(\mathcal{X})} \left(\int_{\mathcal{S}} \sup_{a \in \mathcal{A}} Q(s,a) d\mu_0(s) + (-R)^\star(-r_Q) \right),
\end{aligned}$$

where (1) is due to Theorem 1, (2) is due to (A), and (3) follows since $(-R)^\star$ is increasing by assumption. Next, noting that $(-R)^\star(-r_Q) = \varepsilon\Omega^\star\left(\frac{r-r_Q}{\varepsilon}\right)$, and that

$$\begin{aligned}
r - r_Q &= r(s,a) - \frac{Q(s,a)}{1-\gamma} + \gamma \int_{\mathcal{X}} \sup_{a' \in \mathcal{A}} Q(s',a') dP(s' \mid s,a) \\
&= \left(r(s,a) + \gamma \int_{\mathcal{X}} \sup_{a' \in \mathcal{A}} Q(s',a') dP(s' \mid s,a)\right) - Q(s,a) \\
&= \mathcal{T}Q - Q,
\end{aligned}$$

which is the difference between the Bellman operator. Putting this together yields

$$\begin{aligned}
&\sup_{\mu \in \mathcal{K}_{P,\gamma}} R(\mu) \\
&= \inf_{Q \in \mathcal{F}_b(\mathcal{X})} \left(\varepsilon\Omega^\star\left(\frac{r - r_Q}{\varepsilon}\right) + \int_{\mathcal{S}} \sup_{a \in \mathcal{A}} Q(s,a) d\mu_0(s)\right) \\
&= \inf_{Q \in \mathcal{F}_b(\mathcal{X})} \left(\varepsilon\Omega^\star\left(\frac{\mathcal{T}Q - Q}{\varepsilon}\right) + \int_{\mathcal{S}} \sup_{a \in \mathcal{A}} Q(s,a) d\mu_0(s)\right)
\end{aligned}$$

1.7 Proof of Lemma 2

We first set $n = |A|$. Let $\mathcal{F}_b(\mathcal{S}, \mathbb{R}^n)$ denote the set of measurable and bounded functions mapping from $\mathcal{S}$ into $\mathbb{R}^n$. For any $\pi \in \mathcal{F}_b(\mathcal{S}, \mathbb{R}^n)$, we use $\pi(a \mid s)$ to denote the index corresponding to $a \in \mathcal{A}$ for the function π evaluated at $s \in \mathcal{S}$. Next, we define the following set:

$$\mathcal{B}_\times := \left\{\mu(s,a) = \pi(a \mid s) \cdot \mu_S(s) \mid \mu_S \in \mathscr{P}(\mathcal{S}), \pi \in \mathcal{F}_b(\mathcal{S}, \mathbb{R}^n)\right\},$$

noting that $\mathcal{B}_\times \subseteq \mathscr{B}(\mathcal{X})$. We also have that $\mathscr{P}(\mathcal{X}) \subset \mathcal{B}_\times$ since this corresponds to having each $\pi(a \mid s)$ satisfy $\pi(a \mid s) \in [0,1]$ and $\sum_{a \in \mathcal{A}} \pi(a \mid s) = 1$. We then redefine

$$\Omega(\mu) = \begin{cases} \mathbb{E}_{\mu(s,a)}\left[\mathrm{KL}(\pi_\mu(\cdot \mid s), U)\right] & \text{if } \mu \in \mathcal{B}_\times \\ \infty & \text{if } \mu \notin \mathcal{B}_\times \end{cases}$$

We will first show that this choice of Ω is convex. First we need a Lemma that will make it easier.

Lemma 2 *The functional $F : \mathbb{R}^n \to \mathbb{R}$ defined as*

$$F(\mathbf{x}) = \sum_{i=1}^n x_i \cdot \log\left(\frac{x_i}{\sum_{j=1} x_j}\right)$$

is convex over its domain $\mathbb{R}^n_{>0}$.

Proof We derive the Hessian of F which can be verified to be:

$$HF(\mathbf{x}) = \operatorname{diag}\left(\frac{1}{x_1}, \frac{1}{x_2}, \ldots, \frac{1}{x_n}\right) - \frac{1}{\sum_{i=1}^n x_i} \cdot \mathbf{1}^\intercal \mathbf{1}.$$

Next, we have for any vector $z \in \mathbb{R}^n$ and $x \in \operatorname{dom} F$:

$$\begin{aligned}
z^\intercal HF(x) z &= z^\intercal \operatorname{diag}\left(\frac{1}{x_1}, \frac{1}{x_2}, \ldots, \frac{1}{x_n}\right) z - \frac{1}{\sum_{i=1}^n x_i}\left(\sum_{i=1}^n z_i\right)^2 \\
&= \sum_{i=1}^n \frac{z_i^2}{x_i} - \frac{1}{\sum_{i=1}^n x_i}\left(\sum_{i=1}^n z_i\right)^2 \\
&= \frac{1}{\sum_{i=1}^n x_i}\left(\left(\sum_{i=1}^n x_i\right) \cdot \left(\sum_{i=1}^n \frac{z_i^2}{x_i}\right) - \left(\sum_{i=1}^n z_i\right)^2\right) \\
&\geq 0,
\end{aligned}$$

where the last inequality follows by an application of Cauchy-Schwarz inequality noting that $x \in \mathrm{Dom}\, F = \mathbb{R}^n_{>0}$. Since the Hessian is positive semi-definite, it follows that F is convex. ■

First denote by $\mu_S(s) = \sum_{a \in \mathcal{A}} \mu(s,a)$ and note that $\pi_\mu(a \mid s) = \mu(s,a)/\mu_S(s)$. For any $\mu \in \mathrm{dom}\,\Omega$, we have

$$\begin{aligned}
\Omega(\mu) &= \mathbb{E}_{\mu(s,a)}\left[\mathrm{KL}(\pi_\mu, U)\right] \\
&= \mathbb{E}_{\mu(s,a)}\left[\sum_{a \in \mathcal{A}} \pi_\mu(a \mid s) \cdot \log\left(\pi_\mu(a \mid s)\right) + \log n\right] \\
&= \mathbb{E}_{\mu_S(s)}\left[\sum_{a \in \mathcal{A}} \pi_\mu(a \mid s) \cdot \log\left(\pi_\mu(a \mid s)\right)\right] + \log n \\
&= \int_{\mathcal{S}} \sum_{a \in \mathcal{A}} \mu_S(s)\pi_\mu(a \mid s) \cdot \log\left(\pi_\mu(a \mid s)\right) ds + \log n \\
&= \int_{\mathcal{S}} \sum_{a \in \mathcal{A}} \mu(s,a) \cdot \log\left(\frac{\mu(s,a)}{\sum_{a' \in \mathcal{A}} \mu(s,a')}\right) ds + \log n,
\end{aligned}$$

and convexity follows by the above Lemma. Before we proceed, we need to also show that $\mathcal{B}_\times$ is convex so that our redefining of Ω does not break convexity established above. Consider $\mu, \nu \in \mathcal{B}_\times$ and so there exists $\mu_S, \nu_S \in \mathscr{P}(\mathcal{S})$ and $\pi_\mu, \pi_\nu \in \mathcal{F}_b(\mathcal{S}, \mathbb{R}^n)$ with $\mu(s,a) = \pi_\mu(a \mid s) \cdot \mu_S(s)$ and $\nu(s,a) = \pi_\nu(a \mid s) \cdot \nu_S(s)$. For any $\lambda \in [0,1]$, we have (setting $P_{\mu,\nu}(s) = \frac{\mu_S(s)+\nu_S(s)}{2}$)

$$\begin{aligned}
\lambda \cdot \mu(s,a) + (1-\lambda)\nu(s,a) &= \lambda\pi_\mu(a \mid s) \cdot \mu_S(s) + (1-\lambda) \cdot \pi_\nu(a \mid s) \cdot \nu_S(s) \\
&= P_{\mu,\nu}(s) \cdot \left(\lambda\pi_\mu(a \mid s) \cdot \frac{\mu_S(s)}{P_{\mu,\nu}(s)} + (1-\lambda) \cdot \pi_\nu(a \mid s) \cdot \frac{\nu_S(s)}{P_{\mu,\nu}(s)}\right).
\end{aligned}$$

By construction, both μ_S and ν_S are absolutely continuous with respect to $P_{\mu,\nu}$ and thus the terms inside the bracket are bounded and well-defined. Moreover $P_{\mu,\nu} \in \mathscr{P}(\mathcal{S})$ and thus this element is in $\mathcal{B}_\times$, which concludes the convexity proof. We now proceed to derive the conjugate. For any $r' \in \mathcal{F}_b(\mathcal{X})$ we have

$$\begin{aligned}
\Omega^\star(r') &= \sup_{\mu \in \mathscr{B}(\mathcal{X})} \left(\int_{\mathcal{X}} r'(s,a) d\mu(s,a) - \Omega(\mu)\right) \\
&\overset{(1)}{=} \sup_{\mu \in \mathcal{B}_\times} \left(\int_{\mathcal{X}} r'(s,a) d\mu(s,a) - \Omega(\mu)\right) \\
&= \sup_{\mu \in \mathcal{B}_\times} \left(\int_{\mathcal{X}} r'(s,a) d\mu(s,a) - \mathbb{E}_{\mu(s,a)}\left[\mathrm{KL}(\pi_\mu(\cdot \mid s), U)\right]\right) \\
&= \sup_{\mu \in \mathcal{B}_\times} \left(\int_{\mathcal{X}} \left(\int_{\mathcal{A}} r'(s,a) d\pi_\mu(a \mid s) - \mathrm{KL}(\pi_\mu(\cdot \mid s), U)\right) d\mu(s,a)\right) \\
&= \sup_{\mu_S \in \mathscr{P}(\mathcal{S})} \sup_{\pi_\mu(\cdot \mid s) \in \mathcal{F}_b(\mathcal{S}, \mathbb{R}^n)} \left(\int_{\mathcal{X}} \left(\int_{\mathcal{A}} r'(s,a) d\pi_\mu(a \mid s) - \mathrm{KL}(\pi_\mu(\cdot \mid s), U)\right) d\mu_S(s)\right) \\
&\overset{(2)}{=} \sup_{\mu_S \in \mathscr{P}(\mathcal{S})} \int_{\mathcal{X}} \sup_{\pi_\mu \in \mathbb{R}^n} \left(\int_{\mathcal{A}} r'(s,a) d\pi_\mu(a) - \mathrm{KL}(\pi_\mu, U)\right) d\mu_S(s) \\
&\overset{(3)}{=} \sup_{\mu_S \in \mathscr{P}(\mathcal{S})} \int_{\mathcal{X}} \sup_{\pi_\mu \in \mathscr{P}(\mathcal{A})} \left(\int_{\mathcal{A}} r'(s,a) d\pi_\mu(a) - \mathrm{KL}(\pi_\mu, U)\right) d\mu_S(s) \\
&\overset{(4)}{=} \sup_{\mu_S \in \mathscr{P}(\mathcal{S})} \int_{\mathcal{X}} \exp\left(r'(s,a)\right) dU(a) - 1 \\
&\overset{(5)}{=} \sup_{s \in \mathcal{S}} \int_{\mathcal{X}} \exp\left(r'(s,a)\right) dU(a) - 1,
\end{aligned}$$

where (1) holds since $\mathrm{dom}\,\Omega \subseteq \mathcal{B}_\times$. (2) holds from (Rockafellar and Wets, 2009, Theorem 14.60, p. 677) using the fact that $\mathcal{F}_b(\mathcal{S}, \mathbb{R}^n)$ is trivially a decomposable space in definition (Rockafellar and Wets, 2009, Definition 14.59, p. 676). (3) holds since $\mathrm{dom}\left(\mathrm{KL}(\cdot, U)\right) \subseteq \mathscr{P}(\mathcal{A}) \subset \mathbb{R}^n$. (4) is due to (Feydy et al., 2019, Proposition 5) and (5) follows by noting that the optimal μ_S is concentrated around the supremum.

1.8 Imitation Learning

1.8.1 f-divergence

Note that for any $r \in \mathcal{F}_b(\mathcal{X})$ we have

$$\begin{aligned}
(-R)^\star(r) &= \sup_{\nu \in \mathscr{B}(\mathcal{X})} \left(\int_{\mathcal{X}} r(x) d\nu(x) + R(\nu) \right) \\
&= \sup_{\nu \in \mathscr{B}(\mathcal{X})} \left(\int_{\mathcal{X}} r(x) d\nu(x) - \mathrm{KL}(\nu, \mu_E \right) \\
&\overset{(1)}{=} \int_{\mathcal{X}} r(x) d\mu_E(x) - 1,
\end{aligned}$$

where (1) holds due to (Feydy et al., 2019, Proposition 5). We will now show that $(-R)^\star$ is increasing for any $R(\mu) = -D_f(\mu, \mu_E)$ where D_f is an f-divergence. First let

$$\nu \in \operatorname*{arg\,sup}_{\mu \in \mathscr{P}(\mathcal{X})} \left(\int_{\mathcal{X}} r(x) d\mu(x) + R(\mu) \right),$$

noting that ν exists since the mapping $\mu \mapsto \int_{\mathcal{X}} r(x) d\mu(x) + R(\mu)$ is concave, upper semicontinuous and $\mathscr{P}(\mathcal{X})$ is compact. For any $r' \geq r$

$$\begin{aligned}
&(-R_-)^\star(r) - (-R_-)^\star(r') \\
&= \sup_{\mu \in \mathscr{B}(\mathcal{X})} \left(\int_{\mathcal{X}} r(x) d\mu(x) + R(\mu) \right) - \sup_{\mu \in \mathscr{B}(\mathcal{X})} \left(\int_{\mathcal{X}} r'(x) d\mu(x) + R(\mu) \right) \\
&\overset{(1)}{=} \sup_{\mu \in \mathscr{P}(\mathcal{X})} \left(\int_{\mathcal{X}} r(x) d\mu(x) + R(\mu) \right) - \sup_{\mu \in \mathscr{P}(\mathcal{X})} \left(\int_{\mathcal{X}} r'(x) d\mu(x) + R(\mu) \right) \\
&\leq \int_{\mathcal{X}} r(x) d\nu(x) + R(\nu) - \int_{\mathcal{X}} r'(x) d\nu(x) - R(\nu) \\
&= \int_{\mathcal{X}} \left(r(x) - r'(x) \right) d\nu(x) \\
&\leq 0,
\end{aligned}$$

where (1) holds due to the fact that $\operatorname{dom}\left(D_f(\cdot, \mu_E) \right) \subseteq \mathscr{P}(\mathcal{X})$.

1.8.2 InfoGAIL

In this case, we exploit the fact that $-R(\mu)$ takes the form of an Integral Probability Metric between μ and μ_E. Let $\mathcal{H}_L$ the set of functions that are L-Lipschitz with respect to d. For any $r \in \mathcal{F}_b(\mathcal{X})$ we have

$$\begin{aligned}
(-R)^\star(r) &= \sup_{\nu \in \mathscr{B}(\mathcal{X})} \left(\int_{\mathcal{X}} r(x) d\nu(x) - \sup_{h : \mathrm{Lip}_d(h) \leq L} \left(\int_{\mathcal{X}} h(x) d\nu(x) - \int_{\mathcal{X}} h(x) d\mu_E(x) \right) \right) \\
&\overset{(1)}{=} \int_{\mathcal{X}} r(x) d\mu_E(x) + \iota_{\mathcal{H}_L}(r),
\end{aligned}$$

where (1) is due to (Husain, 2020, Lemma 5). Thus, it holds that

$$\begin{aligned}
\sup_{\mu \in \mathcal{K}_{P,\gamma}} R(\mu) &= \inf_{r' \in \mathcal{F}_b(\mathcal{X})} \left(\mathrm{RL}_{P,\gamma}(r') + \int_{\mathcal{X}} -r'(x) d\mu_E(x) + \iota_{\mathcal{H}_L}(-r') \right) \\
&\overset{(2)}{=} \inf_{r' \in \mathcal{F}_b(\mathcal{X})} \left(\mathrm{RL}_{P,\gamma}(r') - \int_{\mathcal{X}} r'(x) d\mu_E(x) + \iota_{\mathcal{H}_L}(r') \right) \\
&= \inf_{r' : \mathrm{Lip}_d \leq L} \left(\mathrm{RL}_{P,\gamma}(r') - \int_{\mathcal{X}} r'(x) d\mu_E(x) \right),
\end{aligned}$$

where (2) holds since $\text{Lip}_d(-r) = \text{Lip}_d(r)$. We now show that adding an entropy term to

$$R(\mu) = -\sup_{h:\text{Lip}_d(h)\leq L}\left(\int_{\mathcal{X}} h(x)d\mu(x) - \int_{\mathcal{X}} h(x)d\mu_E(x)\right) - \varepsilon\mathbb{E}_{\mu(s,a)}\left[\text{KL}(\pi_\mu(\cdot \mid s), U_A)\right] \tag{3}$$

will ensure that $(-R)^\star$ is increasing. Using standard results from (Penot, 2012) that the conjugate of the sum of two functions is the infimal convolution between their conjugates mean we will convolve both (3) and entropy conjugate from Lemma 2 of the main file.:

$$\begin{aligned}(-R)^\star(r') &= \inf_{r\in\mathcal{F}_b(\mathcal{X})}\left(\sup_{s\in\mathcal{S}}\int_{\mathcal{X}}\exp\left(r'(s,a)-r(s,a)\right)dU(a) + \int_{\mathcal{X}} r d\mu_E + \iota_{\mathcal{H}_L}(r)\right) &(4)\\ &= \inf_{r\in\mathcal{H}_L}\left(\sup_{s\in\mathcal{S}}\int_{\mathcal{X}}\exp\left(r'(s,a)-r(s,a)\right)dU(a) + \int_{\mathcal{X}} r d\mu_E\right). &(5)\end{aligned}$$

Let $r'' \leq r'$ pointwise and define

$$r^* \in \arg\inf_{r\in\mathcal{H}_L}\left(\sup_{s\in\mathcal{S}}\int_{\mathcal{X}}\exp\left(r'(s,a)-r(s,a)\right)dU(a) + \int_{\mathcal{X}} r d\mu_E\right), \tag{6}$$

noting that since exists due to Weierstrass Theorem since $\mathcal{H}_L$ is compact and the mapping inside is convex and lower semicontinuous. Next, we have

$$\begin{aligned}&(-R)^\star(r'') - (-R)^\star(r') &(7)\\ &= \inf_{r\in\mathcal{H}_L}\left(\sup_{s\in\mathcal{S}}\int_{\mathcal{X}}\exp\left(r''(s,a)-r(s,a)\right)dU(a) + \int_{\mathcal{X}} r d\mu_E\right) - \inf_{r\in\mathcal{H}_L}\left(\sup_{s\in\mathcal{S}}\int_{\mathcal{X}}\exp\left(r'(s,a)-r(s,a)\right)dU(a) + \int_{\mathcal{X}} r d\mu_E\right) &(8)\\ &\leq \sup_{s\in\mathcal{S}}\int_{\mathcal{X}}\exp\left(r''(s,a)-r^*(s,a)\right)dU(a) + \int_{\mathcal{X}} r^* d\mu_E - \sup_{s\in\mathcal{S}}\int_{\mathcal{X}}\exp\left(r'(s,a)-r^*(s,a)\right)dU(a) - \int_{\mathcal{X}} r^* d\mu_E &(9)\\ &= \sup_{s\in\mathcal{S}}\int_{\mathcal{X}}\exp\left(r''(s,a)-r^*(s,a)\right)dU(a) - \sup_{s\in\mathcal{S}}\int_{\mathcal{X}}\exp\left(r'(s,a)-r^*(s,a)\right)dU(a) &(10)\\ &\leq 0, &(11)\end{aligned}$$

where the last inequality follows from the fact that $r'' \leq r'$ and thus this proves that $(-R)^\star$ is increasing.

1.9 Entropic Exploration

For any $r \in \mathcal{F}_b(\mathcal{X})$

$$\begin{aligned}(-R)^\star(r) &= \sup_{\mu\in\mathscr{B}(\mathcal{X})}\left(\int_{\mathcal{X}} r(x)d\mu(x) - \text{KL}(\mu, U_{\mathcal{X}})\right)\\ &\overset{(1)}{=} \int_{\mathcal{X}}\exp\left(r(x)\right)dU_{\mathcal{X}}(x) - 1,\end{aligned}$$

where (1) follows from (Feydy et al., 2019, Proposition 5).

References

Fan, K. (1953). Minimax theorems. *Proceedings of the National Academy of Sciences of the United States of America*, 39(1):42.

Feydy, J., Séjourné, T., Vialard, F.-X., Amari, S.-i., Trouvé, A., and Peyré, G. (2019). Interpolating between optimal transport and mmd using sinkhorn divergences. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2681–2690.

Husain, H. (2020). Distributional robustness with ipms and links to regularization and gans. *Advances in Neural Information Processing Systems*, 33.

Penot, J.-P. (2012). *Calculus without derivatives*, volume 266. Springer Science & Business Media.

Rockafellar, R. (1968). Integrals which are convex functionals. *Pacific journal of mathematics*, 24(3):525–539.

Rockafellar, R. T. and Wets, R. J.-B. (2009). *Variational analysis*, volume 317. Springer Science & Business Media.

Zalinescu, C. (2002). *Convex analysis in general vector spaces*. World scientific.