Appendix

A Proof of Section 2

A.1 Proof of Theorem 3

Before presenting the proof, let us first prove a lemma.

Lemma 3. *Suppose Φ satisfies Assumptions 1 and 3. For any $y_1, y_2 \in \mathcal{Y}$, define $x^*(y_1)$ and $x^*(y_2)$ as $x^*(y_i) = \arg\min_{x\in\mathcal{X}} \Phi(x, y_i)$ for $i = 1, 2$. Similarly, for any $x_1, x_2 \in \mathcal{X}$, define $y^*(x_1)$ and $y^*(x_2)$ as $y^*(x_i) = \operatorname{argmax}_{y\in\mathcal{Y}} \Phi(x_i, y)$, $i = 1, 2$. Therefore, it holds that*

$$\|x^*(y_1) - x^*(y_2)\| \le \frac{L_{xy}}{\mu_x} |||y_1 - y_2|||, \qquad |||y^*(x_1) - y^*(x_2)||| \le \frac{L_{xy}}{\mu_y} \|x_1 - x_2\|.$$

The proof of this lemma is presented in Appendix E.1. Now we present the proof of the Theorem 3.

Proof. For the ease of notation, let us denote the Lipschitz constant

$$\ell_x(\xi) = \sup_{y\in\mathcal{Y}} \ell_x(\xi, y) \quad \text{and} \quad \ell_y(\xi) = \sup_{x\in\mathcal{X}} \ell_y(\xi, x).$$

For $\forall 1 \le i \le n$, denote

$$\begin{cases} x^*(\hat{y}_{(i)}) = \arg\min_{x\in\mathcal{X}} \Phi(x, \hat{y}_{(i)}), \\ y^*(\hat{x}_{(i)}) = \operatorname{argmax}_{y\in\mathcal{Y}} \Phi(\hat{x}_{(i)}, y). \end{cases}$$

Similarly we can define $x^*(\hat{y})$ and $y^*(\hat{x})$. Therefore,

$$\begin{aligned} &\Phi_{\xi_i}(\hat{x}_{(i)}, y^*(\hat{x}_{(i)})) - \Phi_{\xi_i}(x^*(\hat{y}_{(i)}), \hat{y}_{(i)}) \\ \overset{(a)}{\le}\; & \Phi_{\xi_i}(\hat{x}_{(i)}, y^*(\hat{x})) - \Phi_{\xi_i}(x^*(\hat{y}), \hat{y}_{(i)}) + \ell_x(\xi_i)\left\|x^*(\hat{y}_{(i)}) - x^*(\hat{y})\right\| + \ell_y(\xi_i)|||y^*(\hat{x}_{(i)}) - y^*(\hat{x})||| \\ \overset{(b)}{\le}\; & \Phi_{\xi_i}(\hat{x}_{(i)}, y^*(\hat{x})) - \Phi_{\xi_i}(x^*(\hat{y}), \hat{y}_{(i)}) + \frac{L_{xy}\ell_x(\xi_i)}{\mu_x}|||\hat{y}_{(i)} - \hat{y}||| + \frac{L_{xy}\ell_y(\xi_i)}{\mu_y}\left\|\hat{x}_{(i)} - \hat{x}\right\| \\ \overset{(c)}{\le}\; & \Phi_{\xi_i}(\hat{x}, y^*(\hat{x})) - \Phi_{\xi_i}(x^*(\hat{y}), \hat{y}) + \underbrace{\left(\frac{L_{xy}\ell_x(\xi_i)}{\mu_x} + \ell_y(\xi_i)\right)|||\hat{y}_{(i)} - \hat{y}||| + \left(\frac{L_{xy}\ell_y(\xi_i)}{\mu_y} + \ell_x(\xi_i)\right)\left\|\hat{x}_{(i)} - \hat{x}\right\|}_{T(i)}. \end{aligned} \tag{33}$$

The steps (a) and (c) are due to the function Lipschitz property in Assumption 2, and step (b) is due to Lemma 3. Consequently,

$$\begin{aligned} &\Delta^s(\hat{x}, \hat{y}) \\ =\;& \mathbf{E}\left[\max_{y\in\mathcal{Y}} \Phi(\hat{x}, y) - \min_{x\in\mathcal{X}} \Phi(x, \hat{y})\right] \\ \overset{(a)}{=}\;& \frac{1}{n}\sum_{i=1}^n \mathbf{E}\left[\Phi(\hat{x}_{(i)}, y^*(\hat{x}_{(i)})) - \Phi(x^*(\hat{y}_{(i)}), \hat{y}_{(i)})\right] \\ \overset{(b)}{=}\;& \frac{1}{n}\sum_{i=1}^n \mathbf{E}\left[\Phi_{\xi_i}(\hat{x}_{(i)}, y^*(\hat{x}_{(i)})) - \Phi_{\xi_i}(x^*(\hat{y}_{(i)}), \hat{y}_{(i)})\right] \\ \overset{(c)}{\le}\;& \mathbf{E}\left[\frac{1}{n}\sum_{i=1}^n \left(\Phi_{\xi_i}(\hat{x}, y^*(\hat{x})) - \Phi_{\xi_i}(x^*(\hat{y}), \hat{y})\right)\right] + \frac{1}{n}\sum_{i=1}^n \mathbf{E}[T(i)] \\ \overset{(d)}{\le}\;& \frac{1}{n}\sum_{i=1}^n \mathbf{E}[T(i)]. \end{aligned} \tag{34}$$

The step (a) is because $(\hat{x}, \hat{y})$ and $(\hat{x}_{(i)}, \hat{y}_{(i)})$ are identically distributed. And the step (b) is because the independence between ξ_i and $\Gamma(i)$, which indicates that

$$\mathbf{E}\left[\Phi_{\xi_i}(\hat{x}_{(i)}, y^*(\hat{x}_{(i)}))\right] \;=\; \mathbf{E}\left[\mathbf{E}\left[\Phi_{\xi_i}(\hat{x}_{(i)}, y^*(\hat{x}_{(i)}))\middle|\xi_i\right]\right] = \mathbf{E}\left[\Phi(\hat{x}_{(i)}, y^*(\hat{x}_{(i)}))\right].$$

The independence here is a crucial point and need to be carefully handled. The step (c) is due to (33). And the step (d) is because $(\hat{x}, \hat{y})$ solves the ESP problem (2), which implies $\hat{\Phi}_n(\hat{x}, y) - \hat{\Phi}_n(x, \hat{y}) \le 0$ for $\forall x \in \mathcal{X}, y \in \mathcal{Y}$. Consequently

$$\mathbf{E}\Big[\frac{1}{n}\sum_{i=1}^{n}\left(\Phi_{\xi_i}\left(\hat{x}, y^*(\hat{x})\right) - \Phi_{\xi_i}\left(x^*(\hat{y}), \hat{y}\right)\right)\Big] = \mathbf{E}\left[\hat{\Phi}_n\left(\hat{x}, y^*(\hat{x})\right) - \hat{\Phi}_n\left(x^*(\hat{y}), \hat{y}\right)\right] \le 0.$$

Therefore, the last step to bound $\Delta^s(\hat{x}, \hat{y})$ remains as follows,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[T(i)]\\
=\;&\mathbf{E}\left[\left(\frac{L_{xy}\ell_x(\xi_i)}{\mu_x}+\ell_y(\xi_i)\right)\|\hat{y}_{(i)}-\hat{y}\|+\left(\frac{L_{xy}\ell_y(\xi_i)}{\mu_y}+\ell_x(\xi_i)\right)\|\hat{x}_{(i)}-\hat{x}\|\right]\\
\overset{(a)}{\le}\;&\mathbf{E}\left[\sqrt{\left(\frac{L_{xy}\ell_y(\xi_i)}{\sqrt{\mu_x}\mu_y}+\frac{\ell_x(\xi_i)}{\sqrt{\mu_x}}\right)^2+\left(\frac{L_{xy}\ell_x(\xi_i)}{\mu_x\sqrt{\mu_y}}+\frac{\ell_y(\xi_i)}{\sqrt{\mu_y}}\right)^2}\times\sqrt{\mu_x\|\hat{x}-\hat{x}_{(i)}\|^2+\mu_y\|\hat{y}-\hat{y}_{(i)}\|^2}\right]\\
\overset{(b)}{\le}\;&\sqrt{\mathbf{E}\left[\left(\frac{L_{xy}\ell_y(\xi_i)}{\sqrt{\mu_x}\mu_y}+\frac{\ell_x(\xi_i)}{\sqrt{\mu_x}}\right)^2+\left(\frac{L_{xy}\ell_x(\xi_i)}{\mu_x\sqrt{\mu_y}}+\frac{\ell_y(\xi_i)}{\sqrt{\mu_y}}\right)^2\right]}\times\sqrt{\mathbf{E}\left[\mu_x\|\hat{x}-\hat{x}_{(i)}\|^2+\mu_y\|\hat{y}-\hat{y}_{(i)}\|^2\right]}\\
\overset{(c)}{\le}\;&\sqrt{\frac{2L_{xy}^2(\ell_y^s)^2}{\mu_x\mu_y^2}+\frac{2(\ell_x^s)^2}{\mu_x}+\frac{2L_{xy}^2(\ell_x^s)^2}{\mu_x^2\mu_y}+\frac{2(\ell_y^s)^2}{\mu_y}}\cdot\frac{2}{n}\cdot\sqrt{\frac{(\ell_x^s)^2}{\mu_x}+\frac{(\ell_y^s)^2}{\mu_y}}\\
\le\;&\frac{2\sqrt{2}}{n}\cdot\sqrt{\frac{L_{xy}^2}{\mu_x\mu_y}+1}\cdot\left(\frac{(\ell_x^s)^2}{\mu_x}+\frac{(\ell_y^s)^2}{\mu_y}\right).
\end{aligned}$$

The step (a) here is due to the Cauchy-Schwartz inequality, for any two vectors a and b, $a^\top b \le \|a\|_2 \cdot \|b\|_2$. The step (b) is the expectation version of Cauchy-Shwartz inequality, for any two random variables a and b, $\mathbf{E}[ab] \le \sqrt{\mathbf{E}[a^2]} \cdot \sqrt{\mathbf{E}[b^2]}$. And the step (c) is due to the fact that $(a+b)^2 \le 2(a^2+b^2)$ and the stability argument of Lemma 1.

Finally, substituting this bound into the inequality (34) proves the theorem. □

A.2 Proof of Theorem 4

To prove the Theorem 4, let us first present some definition and lemmas. We define the primal function $f(x)$ and dual function $g(y)$ as well as their empirical version $f_\Gamma(x)$ and $g_\Gamma(y)$:

$$\begin{cases} f(x) = \max_y \Phi(x, y), \\ \hat{f}_n(x) = \max_y \hat{\Phi}_n(x, y), \end{cases} \quad \text{and} \quad \begin{cases} g(y) = \min_x \Phi(x, y), \\ \hat{g}_n(y) = \min_x \hat{\Phi}_n(x, y). \end{cases} \tag{35}$$

For the ease of notation, we also denote

$$x_n^*(y) = \operatorname*{argmin}_x \hat{\Phi}_n(x, y) \quad \text{and} \quad y_n^*(x) = \operatorname*{argmax}_y \hat{\Phi}_n(x, y). \tag{36}$$

As a result the following property holds true.

Proposition 1. *Under Assumption 1 and 3, the primal function $f(x)$ and $\hat{f}_n$ are μ_x-strongly convex; $\nabla f(x)$ is L_f-Lipschitz continuous, with $L_f := L_x + L_{xy}^2/\mu_y$. Similarly, $g(y)$ and $\hat{g}_n$ are μ_y-storngly concave; $\nabla g(y)$ is L_g-Lipschitz continuous, with $L_g := L_y + L_{xy}^2/\mu_x$.*

This proposition is a well known results, see e.g. Sanjabi et al. (2018).

Lemma 4. *The squared distance from the emprical solution to the polulation solution is bounded as*

$$\begin{cases} \|\hat{x} - x^*\|_2^2 \le \frac{4}{\mu_x^2}\|\nabla_x \hat{\Phi}_n(x^*, y_n^*(x^*))\|_2^2 \\ \|\hat{y} - y^*\|_2^2 \le \frac{4}{\mu_y^2}\|\nabla_y \hat{\Phi}_n(x_n^*(y^*), y^*)\|_2^2. \end{cases} \tag{37}$$

and

$$\mathbf{E}\left[\|\nabla_x \hat{\Phi}_n(x^*, y_n^*(x^*))\|_2^2\right] \le \frac{1}{n}\left(\frac{8L_{xy}^2}{\mu_y^2}\mathbf{E}\left[\|\nabla_y \Phi_\xi(x^*, y^*)\|_2^2\right] + 2\mathbf{E}\left[\|\nabla_x \Phi_\xi(x^*, y^*)\|_2^2\right]\right) \tag{38}$$

and

$$\mathbf{E}\big[\|\nabla_y\hat{\Phi}_n(x_n^*(y^*),y^*)\|_2^2\big] \le \frac{1}{n}\left(\frac{8L_{xy}^2}{\mu_x^2}\mathbf{E}\big[\|\nabla_x\Phi_\xi(x^*,y^*)\|_2^2\big] + 2\mathbf{E}\big[\|\nabla_y\Phi_\xi(x^*,y^*)\|_2^2\big]\right). \quad (39)$$

We provide the proof in Appendix E.2. As a result of Proposition 1 and Lemma 4, the proof of Theorem 4 will follow the following argument. By the Lipschitz continuity of $\nabla f(x)$ and $\nabla g(y)$,

$$\begin{aligned}
&\max_{y\in\mathcal{Y}}\Phi(\hat{x},y) - \min_{x\in\mathcal{X}}\Phi(x,\hat{y})\\
=\;& f(\hat{x}) - g(\hat{y})\\
\le\;& f(x^*) + \frac{L_f}{2}\|\hat{x}-x^*\|_2^2 - \Big(g(y^*) - \frac{L_g}{2}\|\hat{y}-y^*\|_2^2\Big)\\
=\;& \frac{L_f}{2}\|\hat{x}-x^*\|_2^2 + \frac{L_g}{2}\|\hat{y}-y^*\|_2^2.
\end{aligned}$$

Taking expectation on both sides and substituting in the values of L_f and L_g in Proposition 1 proves the Theorem.

B Proof of Section 3

B.1 Assumptions on fast mixing time and uniform ergodicity

Assumption 5 (Uniformly bounded ergodicity)**.** *The Markov decision process is ergodic under any stationary policy π, and there exists $\tau > 1$ such that*

$$\frac{1}{\sqrt{\tau}|\mathcal{S}|}\cdot\mathbf{1} \le \sum_{a\in\mathcal{A}} y_a^\pi \le \frac{\sqrt{\tau}}{|\mathcal{S}|}\cdot\mathbf{1},$$

where y^π is the stationary state-action distribution under the policy π.

Assumption 6 (Fast mixing time)**.** *There exists a constant t_{mix} such that for any stationary policy π,*

$$t_{mix} \ge \min_t \Big\{t : \|P_\pi^t(s,\cdot) - \sum_{a\in\mathcal{A}} y_a^\pi\|_{TV} \le 1/4, \forall s\in\mathcal{S}\Big\},$$

where $\|\cdot\|_{TV}$ is the total variation norm, $P_\pi(s,s') = \sum_{a\in\mathcal{A}}\pi(a|s)P_a(s,s')$ is the transition probability matrix under policy π and $P_\pi^t(s,s')$ is the t-step transition probability from s to s'.

B.2 Proof of inequality (29)

To compute the upperbound of $\Delta^w(\bar{x},\bar{y})$, we will first need the following proposition on the Lipschitz constants ℓ_x^w and ℓ_y^w, whose proof is delegated to Appendix E.3.

Proposition 2. *For any ξ there exist constants $\ell_x(\xi,y)$ and $\ell_y(\xi,x)$ s.t. $\Phi_\xi(\cdot,y)$ is $\ell_x(\xi,y)$-Lipschitz under L_2-norm, and $\Phi_\xi(x,\cdot)$ is $\ell_y(\xi,x)$-Lipschitz under L_1-norm. Moreover,*

$$\begin{cases}(\ell_x^w)^2 := \sup_{y\in\mathcal{Y}}\mathbf{E}_\xi[\ell_x^2(\xi,y)] = \mathcal{O}\left(\tau^3/|\mathcal{S}|\right),\\ (\ell_y^w)^2 := \sup_{x\in\mathcal{X}}\mathbf{E}_\xi[\ell_y^2(\xi,x)] = \mathcal{O}\big(t_{mix}^2\big).\end{cases}$$

For the rest of the proof, it suffices to specify the following details for Lemma 2. For the Φ_ξ's, $\mu_x = \mu_y = 0$. The norm $\|\cdot\|$ is the L_2-norm $\|\cdot\|_2$ and the norm $|||\cdot|||$ os tje L_1-norm $\|\cdot\|_1$. We set the regularizer to be

$$\Psi(x,y) = \frac{\alpha_x}{2}\|x\|_2^2 - \alpha_y\sum_{sa} y_{sa}\log y_{sa}.$$

$\Psi(\cdot,y)$ is ν_x-strongly convex in x under the norm $\|\cdot\|_2$ with $\nu_x = \alpha_x$. $\Psi(x,\cdot)$ is ν_y-strongly concave in y under the norm $\|\cdot\|_1$ with $\nu_y = \alpha_y$. Furthermore, for any $x\in\mathcal{X}$ and $y\in\mathcal{Y}$, we know

$$\begin{aligned}
R &= \max_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}}|\Psi(x,y)|\\
&\le \max_{x\in\mathcal{X}}\frac{\alpha_x}{2}\|x\|_2^2 + \max_{y\in\mathcal{Y}}\Big|\alpha_y\sum_{sa}y_{sa}\log y_{sa}\Big|\\
&= 2\alpha_x|\mathcal{S}|t_{mix}^2 + \alpha_y\log(|\mathcal{S}||\mathcal{A}|).
\end{aligned}$$

From Proposition 2, we know $(\ell_x^w)^2 = \mathcal{O}(\tau^3/|\mathcal{S}|)$ and $(\ell_y^w)^2 = \mathcal{O}(t_{mix}^2)$. We get

$$\max_{y\in\mathcal{Y}} \mathbf{E}[\Phi(\hat{x}, y)] - \min_{x\in\mathcal{X}} \mathbf{E}[\Phi(x, \hat{y})] \le \mathcal{O}\left(\frac{\tau^3}{n|\mathcal{S}|\alpha_x} + \frac{t_{mix}^2}{n\alpha_y} + \alpha_x|\mathcal{S}|t_{mix}^2 + \alpha_y \log(|\mathcal{S}||\mathcal{A}|)\right).$$

To minimize the RHS, we set

$$\alpha_x = \frac{\tau^{3/2}}{\sqrt{n}|\mathcal{S}|t_{mix}} \quad \text{and} \quad \alpha_y = \frac{t_{mix}}{\sqrt{n\log(|\mathcal{S}||\mathcal{A}|)}}.$$

This imediately yields

$$\Delta^w(\hat{x}, \hat{y}) \quad = \quad \max_{y\in\mathcal{Y}} \mathbf{E}[\Phi(\hat{x}, y)] - \min_{x\in\mathcal{X}} \mathbf{E}[\Phi(x, \hat{y})] \le \mathcal{O}\left(\frac{t_{mix}}{\sqrt{n}}\left(\tau^{1.5} + \sqrt{\log(|\mathcal{S}||\mathcal{A}|)}\right)\right).$$

B.3 Proof of Inequality (30)

To prove this result, let us first introduce the primal and dual linear programming formulations of the aMDP problem, which are

$$\text{(Primal-LP)} \quad \min_{\hat{v}\in\mathbf{R}, x\in\mathbf{R}^{|\mathcal{S}|}} \hat{v} \text{ s.t. } \hat{v}\cdot\mathbf{1} + (I-P_a)x - r_a \ge 0, \ \forall a\in\mathcal{A}.$$

and

$$\begin{aligned} \text{(Dual-LP)} \qquad & \min_{y\in\mathbf{R}^{|\mathcal{S}|\times|\mathcal{A}|}} \langle y, r\rangle \\ & \text{s.t.} \quad y \ge 0, \quad \|y\|_1 = 1, \ \sum_{a\in\mathcal{A}} (I - P_a^\top) y_a = 0. \end{aligned}$$

Then our saddle point problem (3) is the min-max formulation of this primal-dual LP pair. Let $(\hat{v}^*, x^*)$ be the optimal solution to the (Primal-LP) (40) and let y^* be the optimal soltuion to the (Dual-LP), then (x^*, y^*) forms the saddle point of our problem (3). The following set of conditions are satisfied

$$\begin{cases} \hat{v}^* + (I-P_a)x^* - r_a \ge 0, \ \forall a \in \mathcal{A}, & \text{(Primal feasibility)} \\ y^* \ge 0, \ \|y^*\|_1 = 1, \sum_{a\in\mathcal{A}}(I-P_a^\top)y_a^* = 0, & \text{(Dual feasibility)} \\ \langle y^*, r\rangle = \hat{v}^*, & \text{(Complementarity slackness)} \end{cases} \tag{40}$$

With these preliminary results, let us now provide the proof of this lemma.

Proof. Note that $\Phi(x, y) := \langle y, r\rangle + \sum_{a\in\mathcal{A}} y_a^\top (P_a - I)x$, by direct computation, we have

$$\begin{aligned} & \max_{y\in\mathcal{Y}} \mathbf{E}\left[\Phi(\bar{x}, y)\right] - \min_{x\in\mathcal{X}} \mathbf{E}\left[\Phi(x, \bar{y})\right] \\ \ge\ & \mathbf{E}\left[\Phi(\bar{x}, y^*) - \Phi(x^*, \bar{y})\right] \\ =\ & \mathbf{E}\left[\sum_{a\in\mathcal{A}} (y_a^*)^\top((P_a - I)\bar{x} + r_a) - \sum_{a\in\mathcal{A}} (\bar{y}_a)^\top((P_a - I)x^* + r_a)\right] \\ \overset{(a)}{=}\ & \langle y^*, r\rangle - \mathbf{E}\left[\sum_{a\in\mathcal{A}} (\bar{y}_a)^\top((P_a - I)x^* + r_a)\right] \\ \overset{(b)}{=}\ & \hat{v}^* - \mathbf{E}\left[\sum_{a\in\mathcal{A}} (\bar{y}_a)^\top((P_a - I)x^* + r_a)\right]. \end{aligned}$$

In the step (a), we apply the feasibility of y^*: $\sum_{a\in\mathcal{A}}(y_a^*)^\top(P_a - I) = 0$. In the step (b), we applied the fact that $\hat{v}^* = \langle y^*, r\rangle$. Hence we complete the proof. ◻

C Proof of Section 4

C.1 Proof of Inequality (32)

Proof. For the ease of notation. We denote the Lipschitz constant of the gradient of f as L. We denote the upper bound of the diameters of $\mathcal{X}$ and $\mathcal{Y}$ as c_1 . Since A_ξ and b_ξ are uniformly upper bounded. We denote the upper bound of $\|A_\xi\|_2$ and $\|b\|_2$ as c_2. Then, let us specify the constants used in Theorem 3 as follows. Let

$$\begin{cases} \Phi_\xi(x, y) := r(x) + y^\top(A_\xi x - b_\xi) - f^*(y), \\ \Phi(x, y) := r(x) + y^\top(Ax - b) - f^*(y). \end{cases}$$

First $\mu_x = \mu$ and $\mu_y = 1/L$. For the gradient Lipschitz constant, we know $L_{xy} = \|A\|_2 \le c_2$ since $\nabla_{xy}\Phi(x, y) = A$. For the function Lipschitz constant, we know for any ξ,

$$\ell_x(\xi, y) = \ell_r + \|(A_\xi)^\top y\|_2 \le \ell_r + c_1c_2.$$

By Fenchel duality, we know when $f(Ax + b) = f^*(y)$, then $\|\nabla f^*(y)\| = \|A_\Gamma x + b\| \le (c_1 + 1)c_2$. Consequently, the Lipschitz constant of f^* in $\mathcal{Y}$ is $\ell_{f^*} = O(c_1c_2)$. Then

$$\ell_y(\xi, x) = \ell_{f^*} + \|A_\xi x - b_\xi\|_2 \le O(c_1c_2).$$

Therefore, we know that $\ell_x^s = O(\ell_r + c_1c_2)$ and $\ell_y^s = O(c_1c_2)$. Substituting all the above constants into Theorem 3 yields

$$\begin{aligned} & \mathbf{E}\left[\max_{y\in\mathcal{Y}} \Phi(\bar{x}, y) - \min_{x\in\mathcal{X}} \Phi(x, \bar{y})\right] \\ \le\ & O\left(\sqrt{\frac{L_{xy}^2}{\mu_x\mu_y}}\left(\frac{(\ell_x^s)^2}{n\mu_x} + \frac{(\ell_y^s)^2}{n\mu_y}\right)\right) \\ =\ & O\left(c_2\sqrt{L/\mu}\cdot\left(\frac{c_2^2c_1^2 + \ell_r^2}{n\mu} + \frac{c_1^2c_2^2}{n/L}\right)\right) \\ =\ & O\left(\frac{1}{n\mu^{1.5}}\right). \end{aligned} \tag{41}$$

Next, let $A = \mathbf{E}_\xi[A_\xi]$ and $b = \mathbf{E}_\xi[b_\xi]$, then

$$\begin{aligned} \max_{y\in\mathcal{Y}} \Phi(\bar{x}, y) &= \max_{y\in\mathcal{Y}} r(\bar{x}) + y^\top(A\bar{x} - b) - f^*(y) \\ &= r(\bar{x}) + f(A\bar{x} - b) \\ &= F(\bar{x}), \end{aligned} \tag{42}$$

and

$$-\min_{x\in\mathcal{X}} \Phi(x, \bar{y}) \ge -\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}} \Phi(x, y) = \min_{x\in\mathcal{X}} F(x). \tag{43}$$

Combining (41), (42) and (43) proves the result. □

C.2 Proof of Theorem 6

To prove this theorem, we would like to first present the following result.

Lemma 5. *(Hybrid generalization bound) Under the settings of Theorem 1, if Assumption 3 is satisfied in addition, we have*

$$\begin{aligned} & \mathbf{E}[\max_{y\in\mathcal{Y}} \Phi(\hat{x}, y)] - \min_{x\in\mathcal{X}} \mathbf{E}[\Phi(x, \hat{y})] \\ \le\ & O\left(\sqrt{\frac{L_{xy}^2(\ell_y^s)^2}{\mu_x\mu_y^2} + \frac{(\ell_x^s)^2}{\mu_x} + \frac{(\ell_y^s)^2}{\mu_y}}\cdot\sqrt{\mathbf{E}\left[\mu_x\|\hat{x} - \hat{x}_{(i)}\|^2 + \mu_y\|\|\hat{y} - \hat{y}_{(i)}\|\|^2\right]}\right). \end{aligned} \tag{44}$$

We provide the proof of this lemma in Appendix E.4. Based on this theorem and constants figured out in Appendix C.1, we get

$$\sqrt{\frac{L_{xy}^2(\ell_y^s)^2}{\mu_x\mu_y^2}+\frac{(\ell_x^s)^2}{\mu_x}+\frac{(\ell_y^s)^2}{\mu_y}}=O\left(\sqrt{\frac{L^2}{\mu}+\frac{\ell_r^2}{\mu}+L}\right)=O\left(\mu^{-\frac{1}{2}}\sqrt{L^2+\ell_r^2+L\mu}\right)$$

For the other term, it is worth noting that the result of Lemma lemma:stability-reg indicates the Lipschitz constant of the regularizer does not play a role in the stability bound. On the other hand, this result can also be interpretated as that the Lipschitz constant of the *deterministic* component of the objective function does not affect the stability bound. Therefore, if we view $r(x)$ and $f^*(y)$ as the "regularizers", then Lemma 1 indicates that

$$\sqrt{\mathbf{E}\left[\mu_x\|\hat{x}-\hat{x}_{(i)}\|^2+\mu_y\|\hat{y}-\hat{y}_{(i)}\|^2\right]}\leq O\left(\frac{1}{n}\sqrt{\frac{1}{\mu}+\frac{1}{1/L}}\right)=O\left(\frac{1}{n\sqrt{\mu}}\sqrt{1+L\mu}\right).$$

Due to our assumption that $L\mu\leq O(1)$, the above result indicates that

$$\mathbf{E}\Big[\max_{y\in\mathcal{Y}}\Phi(\bar{x},y)\Big]-\min_{x\in\mathcal{X}}\mathbf{E}\left[\Phi(x,\bar{y})\right]\leq O\left(\frac{L+\ell_r}{n\mu}\right)=O\left(\frac{1}{\mu\cdot n}\right).$$

In this case, (42) is still true, while (43) is replaced with

$$-\min_{x\in\mathcal{X}}\mathbf{E}\left[\Phi(x,\bar{y})\right]\overset{(a)}{\geq}-\min_{x\in\mathcal{X}}\Phi\big(x,\mathbf{E}[\bar{y}]\big)\geq-\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}}\Phi(x,y)=\min_{x\in\mathcal{X}}F(x),\tag{45}$$

where step (a) is due to the Jensen's inequality. Combinging (45), (42) and (45) proves the theorem.

D Proof of Section 5

D.1 Proof of Theorem 7

Proof. The result of this theorem is a direct corollary of Lemma 2. First, we will need to figure the corresponding algorithmic constants. Define

$$\Psi(x,y)=\frac{1}{\sqrt{n\log N_1}}\sum_i x_i\log x_i-\frac{1}{\sqrt{n\log N_2}}\sum_j y_j\log y_j$$

as the regularizer. Then we know Ψ is ν_x-strongly convex under L_1-norm and ν_y-strongly concave under L_1-norm, with $\nu_x=\frac{1}{\sqrt{n\log N_1}}$ and $\nu_y=\frac{1}{\sqrt{n\log N_2}}$. Morever, the magnitude of Ψ is upper bounder by

$$\begin{aligned}
R &= \max_{x\in\Delta_{N_1}}\max_{y\in\Delta_{N_2}}|\Psi(x,y)|\\
&= \frac{1}{\sqrt{n\log N_1}}\log N_1+\frac{1}{\sqrt{n\log N_2}}\log N_2\\
&= \frac{1}{\sqrt{n}}\left(\sqrt{\log N_1}+\sqrt{\log N_2}\right).
\end{aligned}$$

Denote $\Phi_\xi(x,y)=x^\top A_\xi y$, then clearly, Φ_ξ is not an SC-SC function. Hence $\mu_x=\mu_y=0$. Let the Lipschitz constants $\ell_x(\xi,y)$ and $\ell_y(\xi,x)$ be measured under the L_1-norm, then

$$\begin{aligned}
\ell_x(\xi,y) &= \max_{x\in\Delta_{N_1}}\|\nabla_x\Phi_\xi(x,y)\|_\infty\\
&= \max_{x\in\Delta_{N_1}}\|A_\xi y\|_\infty\\
&= \max_{ij}|A_\xi(i,j)|\\
&= 1.
\end{aligned}$$

Consequently, we have $\ell_x^w = 1$. Similarly, we have $\ell_y(\xi, x) \leq 1$ and $\ell_y^w \leq 1$. Consequently, by Lemma 2, we have

$$\begin{aligned}\Delta^w(\hat{x}, \hat{y}) &\leq \frac{2(\ell_x^w)^2}{n(\mu_x + \nu_x)} + \frac{2(\ell_y^w)^2}{n(\mu_y + \nu_y)} + 2R \\ &= 8\frac{\sqrt{\log N_1} + \sqrt{\log N_2}}{\sqrt{n}} \\ &\leq 16\frac{\sqrt{\log(N_1 N_2)}}{\sqrt{n}}.\end{aligned}$$

The last inequality is due to $\sqrt{\log N_1} + \sqrt{\log N_2} \leq 2\sqrt{\log N_1 + \log N_2} = 2\sqrt{\log(N_1 N_2)}$. Due to the definition of $\Delta^w(\hat{x}, \hat{y})$, we know that for any $x \in \Delta_{N_1}$ and $y \in \Delta_{N_2}$,

$$\begin{aligned}&\underbrace{\mathbf{E}\left[\bar{x}^\top A y - \bar{x}^\top A \bar{y}\right]}_{\geq 0} - \underbrace{\mathbf{E}\left[x^\top A \bar{y} - \bar{x}^\top A \bar{y}\right]}_{\leq 0} \\ = \;& \mathbf{E}\left[\bar{x}^\top A y - x^\top A \bar{y}\right] \leq \Delta^w(\hat{x}, \hat{y}) \\ \leq \;& O\left(\frac{\sqrt{\log(N_1 N_2)}}{\sqrt{n}}\right).\end{aligned}$$

Consequently, $\mathbf{E}\left[\bar{x}^\top A y - \bar{x}^\top A \bar{y}\right] \leq O\left(\frac{\sqrt{\log(N_1 N_2)}}{\sqrt{n}}\right)$ for any y. Meaning that when the player 1 plays $\bar{x}$, in expectation, it does not gain much benefit for player 2 if he switches to any other fixed strategy y. Symmetrically, we have $\mathbf{E}\left[x^\top A \bar{y} - \bar{x}^\top A \bar{y}\right]$, meaning that when the player 2 plays the strategy $\bar{y}$, in expectation, it does not cause more lost for player 2 if player 1 switches to any other fixed strategy x. □

E Other supporting lemmas

E.1 Proof of Lemma 3

Proof. First, let us prove the result for $\|x^*(y_1) - x^*(y_2)\|$ and the rest of the results can be proved parallelly. By the optimality condition we have

$$\begin{cases}\langle \nabla_x \Phi(x^*(y_1), y_1), x^*(y_2) - x^*(y_1)\rangle \geq 0, \\ \langle \nabla_x \Phi(x^*(y_2), y_2), x^*(y_1) - x^*(y_2)\rangle \geq 0.\end{cases}$$

Summing this up gives

$$\langle \nabla_x \Phi(x^*(y_2), y_2) - \nabla_x \Phi(x^*(y_1), y_1), x^*(y_1) - x^*(y_2)\rangle \geq 0.$$

By the strong convexity of $\Phi(\cdot, y)$ and the L_{xy}-Lipschitz continuity of $\nabla_x \Phi(x, y)$ in terms of y,

$$\begin{aligned}0 &\leq \langle \nabla_x \Phi(x^*(y_2), y_2) - \nabla_x \Phi(x^*(y_1), y_1), x^*(y_1) - x^*(y_2)\rangle \\ &= \langle \nabla_x \Phi(x^*(y_2), y_2) - \nabla_x \Phi(x^*(y_1), y_2), x^*(y_1) - x^*(y_2)\rangle + \langle \nabla_x \Phi(x^*(y_1), y_2) - \nabla_x \Phi(x^*(y_1), y_1), x^*(y_1) - x^*(y_2)\rangle \\ &\leq -\mu_x \|x^*(y_1) - x^*(y_2)\|^2 + \|\nabla_x \Phi(x^*(y_1), y_2) - \nabla_x \Phi(x^*(y_1), y_1)\|_* \|x^*(y_1) - x^*(y_2)\| \\ &\leq -\mu_x \|x^*(y_1) - x^*(y_2)\|^2 + L_{xy} |||y_1 - y_2||| \cdot \|x^*(y_1) - x^*(y_2)\|.\end{aligned}$$

Consequently,

$$\|x^*(y_1) - x^*(y_2)\| \leq \frac{L_{xy}}{\mu_x} |||y_1 - y_2|||.$$

The other part of the result follows the same line of proof. □

E.2 Proof of Lemma 4

Proof. Because $\hat{x} = \operatorname{argmin}_x \hat{f}_n(x)$, and $\hat{f}_n(\cdot)$ is μ_x-strongly convex, we have

$$\begin{aligned}0 &\geq \hat{f}_n(\hat{x}) - \hat{f}_n(x^*) \\ &\geq \langle \nabla \hat{f}_n(x^*), \hat{x} - x^*\rangle + \frac{\mu_x}{2}\|\hat{x} - x^*\|_2^2 \\ &= \langle \nabla_x \hat{\Phi}_n(x^*, y_n^*(x^*)), \hat{x} - x^*\rangle + \frac{\mu_x}{2}\|\hat{x} - x^*\|_2^2,\end{aligned}$$

where the last row is due to the Danskin's theorem. By rearranging the terms, we get

$$\frac{\mu_x}{2}\|\hat{x}-x^*\|_2^2 \le -\langle \nabla_x \hat{\Phi}_n(x^*, y_n^*(x^*)), \hat{x}-x^*\rangle \le \|\nabla_x \hat{\Phi}_n(x^*, y_n^*(x^*))\|_2 \cdot \|\hat{x}-x^*\|_2.$$

Deviding both sides by $\|\hat{x}-x^*\|_2$ and then square both sides proves the first inequality of (37). The second inequality for (37) can be proved similarly.

Next, let us focus on the first inequality of (38).

$$\begin{aligned}
&\mathbf{E}\big[\|\nabla_x \hat{\Phi}_n(x^*, y_n^*(x^*))\|_2^2\big] \\
=\ &\mathbf{E}\big[\|\nabla_x \hat{\Phi}_n(x^*, y_n^*(x^*)) - \nabla_x \hat{\Phi}_n(x^*, y^*) + \nabla_x \hat{\Phi}_n(x^*, y^*)\|_2^2\big] \\
\le\ &2\mathbf{E}\big[\|\nabla_x \hat{\Phi}_n(x^*, y_n^*(x^*)) - \nabla_x \hat{\Phi}_n(x^*, y^*)\|_2^2\big] + 2\mathbf{E}\big[\|\nabla_x \hat{\Phi}_n(x^*, y^*)\|_2^2\big] \\
\le\ &2L_{xy}^2\mathbf{E}\big[\|y^* - y_n^*(x^*)\|_2^2\big] + 2\mathbf{E}\big[\|\nabla_x \hat{\Phi}_n(x^*, y^*)\|_2^2\big].
\end{aligned} \tag{46}$$

Note that $\mathbf{E}\big[\nabla_x \hat{\Phi}_n(x^*, y^*)\big] = \nabla_x \Phi(x^*, y^*) = 0$, we have

$$\begin{aligned}
&\mathbf{E}\big[\|\nabla_x \hat{\Phi}_n(x^*, y^*)\|_2^2\big] \\
=\ &\mathbf{E}\left[\left\|\frac{1}{n}\sum_{\xi\in\Gamma} \nabla_x \Phi_\xi(x^*, y^*) - \mathbf{E}\left[\nabla_x \Phi_\xi(x^*, y^*)\right]\right\|_2^2\right] \\
=\ &\frac{1}{n}\mathbf{E}_\xi\left[\|\nabla_x \Phi_\xi(x^*, y^*)\|_2^2\right].
\end{aligned} \tag{47}$$

It is important that L_2-norm is used here so that the above variance equation chain holds. If another norm is used, (47) may not be true. For example if L_1-norm is used, an extra multiplicative factor of dimension will come into the bound. For the other term, note that

$$y_n^*(x^*) = \operatorname*{argmax}_y \hat{\Phi}_n(x^*, y) \qquad \text{and} \qquad y^* = \operatorname*{argmax}_y \Phi(x^*, y).$$

As a result, we have

$$\begin{aligned}
0 &\le \hat{\Phi}_n(x^*, y_n^*(x^*)) - \hat{\Phi}_n(x^*, y^*) \\
&\le \langle \nabla_y \hat{\Phi}_n(x^*, y^*), y_n^*(x^*) - y^*\rangle - \frac{\mu_y}{2}\|y_n^*(x^*) - y^*\|_2^2.
\end{aligned}$$

With slight rearranging and apply Cauchy-Schwartz inequality, we have

$$\|y_n^*(x^*) - y^*\|_2 \le \frac{2}{\mu_y}\|\nabla_y \hat{\Phi}_n(x^*, y^*)\|_2.$$

Taking expectation on both sides and we get

$$\begin{aligned}
\mathbf{E}\big[\|y_n^*(x^*) - y^*\|_2^2\big] &\le \frac{4}{\mu_y^2}\mathbf{E}\big[\|\nabla_y \hat{\Phi}_n(x^*, y^*)\|_2^2\big] \\
&= \frac{4}{\mu_y^2}\mathbf{E}\big[\|\nabla_y \hat{\Phi}_n(x^*, y^*) - \nabla_y \Phi(x^*, y^*)\|_2^2\big] \\
&\le \frac{4}{n\mu_y^2}\mathbf{E}_\xi\left[\|\nabla_y \Phi_\xi(x^*, y^*)\|_2^2\right].
\end{aligned} \tag{48}$$

The argument here is parallel to that of (47). Combining (46), (47), and (48), we have

$$\begin{aligned}
&\mathbf{E}\big[\|\nabla_x \hat{\Phi}_n(x^*, y_n^*(x^*))\|_2^2\big] \\
&\le \frac{1}{n}\left(\frac{8L_{xy}^2}{\mu_y^2}\mathbf{E}_\xi\left[\|\nabla_y \Phi_\xi(x^*, y^*)\|_2^2\right] + 2\mathbf{E}_\xi\left[\|\nabla_x \Phi_\xi(x^*, y^*)\|_2^2\right]\right).
\end{aligned}$$

The second inequality can be proved through a completely parallel way. □

E.3 Proof of Proposition 2

Proof. First, note that the Lipschitz continuity in x variable is measured under L_2-norm, and the Lipschitz continuity in y variable is measured under the L_1-norm. Because the dual norms of $\|\cdot\|_2$ and $\|\cdot\|_1$ are $\|\cdot\|_2$ and $\|\cdot\|_\infty$ repectively, we have

$$\begin{cases} \ell_y^2(\xi, x) = \sup_{y\in\mathcal{Y}} \|\nabla_y \Phi_\xi(x,y)\|_\infty^2, \\ \ell_x^2(\xi, y) = \sup_{x\in\mathcal{X}} \|\nabla_y \Phi_\xi(x,y)\|_2^2. \end{cases}$$

By direct calculation, we know

$$\nabla_{y_a}\Phi_\xi(x,y) = \hat{r}_a - x + u_a \quad \text{with} \quad u_{a,s} = \sum_{s'\in\mathcal{S}} \delta_\xi(s,a,s')x_{s'},$$

where $\hat{r}_a = [\hat{r}_{1,a}, \hat{r}_{2a}, ..., \hat{r}_{|\mathcal{S}|a}]^\top$. Consequently,

$$\begin{aligned} \|\nabla_y\Phi_\xi(x,y)\|_\infty &= \max_{a\in\mathcal{A}} \|\nabla_{y_a}\Phi_\xi(x,y)\|_\infty \\ &\le \max_{a\in\mathcal{A}} \|\hat{r}_a\|_\infty + \|x\|_\infty + \|u_a\|_\infty \\ &\le 1 + 4t_{mix}. \end{aligned}$$

As a result we have $(\ell_y^w)^2 = O(t_{mix}^2)$. For ℓ_x, we first compute the gradient as follows

$$\nabla_x\Phi_\xi(x,y) = -\sum_{a\in\mathcal{A}} y_a + w \quad \text{with} \quad w_{s'} = \sum_{s,a} y_{sa}\delta_\xi(s,a,s').$$

Consequently, for any fixed $y \in \mathcal{Y}$,

$$\begin{aligned} \mathbf{E}_\xi[\ell_x^2(\xi,y)] &= \|\nabla_x\Phi_\xi(x,y)\|_2^2 \\ &\le 2\Big\|\sum_{a\in\mathcal{A}} y_a\Big\|_2^2 + 2\mathbf{E}_\xi[\|w\|_2^2] \\ &= 2\Big\|\sum_{a\in\mathcal{A}} y_a\Big\|_2^2 + 2\|\mathbf{E}[w]\|_2^2 + 2\mathbf{E}_\xi[\|w - \mathbf{E}_\xi[w]\|_2^2]. \end{aligned} \tag{49}$$

Note that $\sum_{a\in\mathcal{A}} y_a \le \frac{\sqrt{\tau}}{|\mathcal{S}|}\cdot\mathbf{1}$, we know $\|\sum_{a\in\mathcal{A}} y_a\|_2^2 \le \frac{\tau}{|\mathcal{S}|}$. By directly calculating the expectation, we know

$$\mathbf{E}_\xi[w_{s'}] = \sum_{s,a} y_{sa}P_a(s,s').$$

Because for a particular s', $\delta_\xi(s_1,a_1,s')$ is independent from $\delta_\xi(s_2,a_2,s')$, we can compute the variance term as

$$\begin{aligned} & \mathbf{E}_\xi\left[\|w - \mathbf{E}_\xi[w]\|_2^2\right] \\ =\ & \sum_{s'} \mathbf{E}_\xi\left[\Big(\sum_{sa} y_{sa}(\delta_\xi(s,a,s') - P_a(s,s'))\Big)^2\right] \\ =\ & \sum_{s'}\sum_{sa} y_{sa}^2 \mathbf{E}_\xi\left[(\delta_\xi(s,a,s') - P_a(s,s'))^2\right] \\ =\ & \sum_{s'}\sum_{sa} y_{sa}^2(1 - P_a(s,s'))P_a(s,s') \\ \le\ & \sum_{s'}\sum_{sa} y_{sa}^2 P_a(s,s') \\ \le\ & \sum_{sa} y_{sa}^2 \\ \overset{(a)}{\le}\ & \Big\|\sum_a y_a\Big\|_2^2 \\ \le\ & \tau/|\mathcal{S}|. \end{aligned}$$

Where the step (a) is because $y \ge 0$. Now we bound the last term $\|\mathbf{E}_\xi[w]\|_2^2$. For the ease of discussion, let us define $\bar{y} = 0.5y + 0.5\frac{\mathbf{1}}{|\mathcal{S}||\mathcal{A}|}$, $\bar{\lambda} = \sum_{a\in\mathcal{A}} \bar{y}_a$, and $\bar{\pi}(a|s) = \bar{y}_{sa}/\bar{\lambda}_s$. Similarly, we define $\hat{y} = \frac{\mathbf{1}}{|\mathcal{S}||\mathcal{A}|}$, $\hat{\lambda} = \frac{\mathbf{1}}{|\mathcal{S}|}$, and $\hat{\pi}(a|s) = \frac{1}{|\mathcal{A}|}$.

Therefore, because both $\bar{\pi}$ and $\hat{\pi}$ are strictly positive, the corresponding Markov chains of the state transitionsis are ergodic. Hence,

$$
\begin{aligned}
&\|\mathbf{E}_\xi[w]\|_2^2\\
&= \sum_{s'}\left(\sum_{s,a}\left(y_{sa} + \frac{1}{|\mathcal{S}||\mathcal{A}|} - \frac{1}{|\mathcal{S}||\mathcal{A}|}\right)P_a(s,s')\right)^2\\
&\le 2\sum_{s'}\left(\sum_{s,a}\left(y_{sa} + \frac{1}{|\mathcal{S}||\mathcal{A}|}\right)P_a(s,s')\right)^2\\
&\quad +2\sum_{s'}\left(\sum_{s,a}\left(\frac{1}{|\mathcal{S}||\mathcal{A}|}\right)P_a(s,s')\right)^2\\
&\overset{(a)}{=} 8\sum_{s'}\left(\sum_{s,a}\bar{y}_{sa}P_a(s,s')\right)^2 + 2\sum_{s'}\left(\sum_{s,a}\hat{y}_{sa}P_a(s,s')\right)^2\\
&\overset{(b)}{=} 8\sum_{s'}\Big(\sum_{s}\bar{\lambda}_s\sum_a\bar{\pi}(a|s)P_a(s,s')\Big)^2\\
&\quad +2\sum_{s'}\Big(\sum_{s}\hat{\lambda}_s\sum_a\hat{\pi}(a|s)P_a(s,s')\Big)^2\\
&\overset{(c)}{=} 8\sum_{s'}\Big(\sum_{s}\bar{\lambda}_sP_{\bar{\pi}}(s,s')\Big)^2 + 2\sum_{s'}\Big(\sum_{s}\hat{\lambda}_sP_{\hat{\pi}}(s,s')\Big)^2\\
&= 8\|P_{\bar{\pi}}^\top\bar{\lambda}\|_2^2 + 2\|P_{\hat{\pi}}^\top\hat{\lambda}\|_2^2\\
&\overset{(d)}{\le} \frac{10\tau^3}{|\mathcal{S}|}.
\end{aligned}
$$

The step (a) and (b) follows directly from the definition of $\bar{y}$, $\bar{\lambda}$, $\bar{\pi}$ and $\hat{y}$, $\hat{\lambda}$, $\hat{\pi}$. The step (c) is we define $P_{\bar{\pi}}$ to be the state transition probability matrix under the policy $\bar{\pi}$, and $P_{\bar{\pi}}(s,s') := \sum_a \bar{\pi}(a|s)P_a(s,s')$; Similar argument is made for $P_{\hat{\pi}}$. Finally, the step (d) is due to the following argument. Let $\lambda^{\bar{\pi}}$ be the stationary state distribution under the policy $\bar{\pi}$, then by ergodicity property (Assumption 5) we have

$$0 \le P_{\bar{\pi}}^\top\mathbf{1} \le P_{\bar{\pi}}^\top(\sqrt{\tau}|\mathcal{S}|\lambda^{\bar{\pi}}) = \sqrt{\tau}|\mathcal{S}|\lambda^{\bar{\pi}} \le \sqrt{\tau}|\mathcal{S}|\cdot\frac{\sqrt{\tau}}{|\mathcal{S}|}\cdot\mathbf{1} = \tau\mathbf{1}.$$

As a result, $0 \le P_{\bar{\pi}}^\top\bar{\lambda} \le P_{\bar{\pi}}^\top\frac{\sqrt{\tau}}{|\mathcal{S}|}\mathbf{1} \le \frac{\tau^{3/2}}{|\mathcal{S}|}\mathbf{1}$ and consequently $\|P_{\bar{\pi}}^\top\bar{\lambda}\|_2^2 \le \tau^3/|\mathcal{S}|$. Similarly, $\|P_{\hat{\pi}}^\top\hat{\lambda}\|_2^2 \le \tau^3/|\mathcal{S}|$. Substituting the following bounds

$$
\begin{cases}
\|\sum_a y_a\|_2^2 \le O(\tau/|\mathcal{S}|)\\
\mathbf{E}_\xi\left[\|w-\mathbf{E}_\xi[w]\|_2^2\right] \le O(\tau/|\mathcal{S}|)\\
\left\|\mathbf{E}_\xi[w]\right\|_2^2 \le O(\tau^3/|\mathcal{S}|)
\end{cases}
$$

into (49) proves that $\mathbf{E}_\xi[\ell_x^2(\xi,y)] \le O(\tau^3/|\mathcal{S}|)$. Consequently, $(\ell_x^w)^2 = \sup_{y\in\mathcal{Y}}\mathbf{E}_\xi[\ell_x^2(\xi,y)] \le O(\tau^3/|\mathcal{S}|)$. This completes the proof of this proposition. □

E.4 Proof of Lemma 5

Proof. We still denote $\ell_x(\xi) := \sup_{y\in\mathcal{Y}}\ell_x(\xi,y)$ and $\ell_y(\xi) := \sup_{x\in\mathcal{X}}\ell_y(\xi,x)$. Then, by the Lipschitz continuity assumption, we have

$$
\begin{aligned}
&\Phi_{\xi_i}\left(\hat{x}_{(i)}, y^*\left(\hat{x}_{(i)}\right)\right) - \Phi_{\xi_i}\left(x,\hat{y}_{(i)}\right) \qquad (50)\\
&\le \Phi_{\xi_i}\left(\hat{x}_{(i)}, y^*\left(\hat{x}\right)\right) - \Phi_{\xi_i}\left(x,\hat{y}_{(i)}\right) + \ell_y(\xi_i)|||y^*\left(\hat{x}_{(i)}\right) - y^*\left(\hat{x}\right)|||\\
&\le \Phi_{\xi_i}\left(\hat{x}_{(i)}, y^*\left(\hat{x}\right)\right) - \Phi_{\xi_i}\left(x,\hat{y}_{(i)}\right) + \frac{L_{xy}\ell_y(\xi_i)}{\mu_y}\left\|\hat{x}_{(i)} - \hat{x}\right\|\\
&\le \Phi_{\xi_i}\left(\hat{x}, y^*\left(\hat{x}\right)\right) - \Phi_{\xi_i}\left(x,\hat{y}\right) + \underbrace{\ell_y(\xi_i)|||\hat{y}_{(i)} - \hat{y}||| + \left(\frac{L_{xy}\ell_y(\xi_i)}{\mu_y} + \ell_x(\xi_i)\right)\left\|\hat{x}_{(i)} - \hat{x}\right\|}_{T(i)}.
\end{aligned}
$$

Similar to the proof of Theorem 1, we have

$$
\begin{aligned}
&\mathbf{E}\big[\max_{y\in\mathcal{Y}}\Phi(\hat{x},y)\big]-\min_{x\in\mathcal{X}}\mathbf{E}\left[\Phi(x,\hat{y})\right] \\
&=\frac{1}{n}\sum_{i=1}^{n}\mathbf{E}\left[\Phi\big(\hat{x}_{(i)},y^*\big(\hat{x}_{(i)}\big)\big)\right]-\min_{x\in\mathcal{X}}\frac{1}{n}\sum_{i=1}^{n}\mathbf{E}\left[\Phi\big(x,\hat{y}_{(i)}\big)\right] \\
&=\max_{x\in\mathcal{X}}\frac{1}{n}\sum_{i=1}^{n}\mathbf{E}\left[\Phi_{\xi_i}\left(\hat{x}_{(i)},y^*\big(\hat{x}_{(i)}\big)\right)-\Phi_{\xi_i}\left(x,\hat{y}_{(i)}\right)\right] \\
&\le\max_{x\in\mathcal{X}}\mathbf{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\Phi_{\xi_i}\left(\hat{x},y^*\big(\hat{x}\big)\right)-\Phi_{\xi_i}\left(x,\hat{y}\right)\right)\right]+\frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[T(i)] \\
&\le\frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[T(i)].
\end{aligned}
\tag{51}
$$

Then similar to the proof of Theorem 1 and Theorem 3, we prove the result by providing the follwoing bound

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[T(i)] \\
&=\mathbf{E}\left[\left(\frac{L_{xy}\ell_y(\xi_i)}{\mu_y}+\ell_x(\xi_i)\right)\left\|\hat{x}_{(i)}-\hat{x}\right\|+\ell_y(\xi_i)|||\hat{y}_{(i)}-\hat{y}|||\right] \\
&\le\mathbf{E}\left[\sqrt{\left(\frac{L_{xy}\ell_y(\xi_i)}{\sqrt{\mu_x}\mu_y}+\frac{\ell_x(\xi_i)}{\sqrt{\mu_x}}\right)^2+\left(\frac{\ell_y(\xi_i)}{\sqrt{\mu_y}}\right)^2}\times\sqrt{\mu_x\|\hat{x}-\hat{x}_{(i)}\|^2+\mu_y|||\hat{y}-\hat{y}_{(i)}|||^2}\right] \\
&\le\sqrt{\mathbf{E}\left[\left(\frac{L_{xy}\ell_y(\xi_i)}{\sqrt{\mu_x}\mu_y}+\frac{\ell_x(\xi_i)}{\sqrt{\mu_x}}\right)^2+\left(\frac{\ell_y(\xi_i)}{\sqrt{\mu_y}}\right)^2\right]}\times\sqrt{\mathbf{E}\left[\mu_x\|\hat{x}-\hat{x}_{(i)}\|^2+\mu_y|||\hat{y}-\hat{y}_{(i)}|||^2\right]} \\
&\le O\left(\frac{1}{n}\sqrt{\frac{L_{xy}^2(\ell_y^s)^2}{\mu_x\mu_y^2}+\frac{(\ell_x^s)^2}{\mu_x}+\frac{(\ell_y^s)^2}{\mu_y}}\times\sqrt{\mathbf{E}\left[\mu_x\|\hat{x}-\hat{x}_{(i)}\|^2+\mu_y|||\hat{y}-\hat{y}_{(i)}|||^2\right]}\right)
\end{aligned}
$$

□