Supplementary material

A. Omitted Proofs

Proof of Theorem 1. We assume there exists an algorithm such that $|\sum_{t=1}^T x_t^\top A_t y_t - \min_{x\in X}\max_{y\in Y}\sum_{t=1}^T x^\top A_t y| \leq o(T)$, $\sum_{t=1}^T x_t^\top A_t y_t - \min_{x\in\Delta_X}\sum_{t=1}^T x^\top A_t y_t \leq o(T)$, and $\max_{y\in\Delta_Y}\sum_{t=1}^T x_t^\top A_T y - \sum_{t=1}^T x_t^\top A_t y_t \leq o(T)$ for all possible sequences of matrices $\{A_t\}_{t=1}^T$ with bounded entries between $[-1,1]$. We now construct two sequences of functions for which all the three guarantees hold and lead that to a contradiction. Let T be divisible by 2. In scenario 1: $A_t = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ for $1 \leq t \leq \frac{T}{2}$ and $A_t = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ for $\frac{T}{2} < t \leq T$. In scenario 2: $A_t = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ for $1 \leq t \leq \frac{T}{2}$ and $A_t = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}$ for $\frac{T}{2} < t \leq T$. It is easy to see that for both scenarios it holds that $\min_{x\in X}\max_{y\in Y}\sum_{t=1}^T x^\top A_t y = 0$. Since $d_1 = d_2 = 2$ and we can parametrize any $x \in \Delta_X$ as $x = [\alpha; 1-\alpha]$ and any $y \in \Delta_Y$ as $y = [\beta; 1-\beta]$ for some $0 \leq \alpha, \beta \leq 1$. By assumption we have that $\max_{y\in\Delta_Y}\sum_{t=1}^T x_t^\top A_t y - \min_{x\in X}\max_{y\in Y}\sum_{t=1}^T x^\top A_t y \leq o(T)$ for all sequences of matrices $\{A_t\}_{t=1}^T$. This implies for scenario 1 that $\max_{0\leq\beta\leq 1}\sum_{t=1}^{\frac{T}{2}} 4\alpha_t\beta - 2\beta + 1 - 2\alpha_t \leq o(T)$ which also implies that $\sum_{t=1}^{\frac{T}{2}} 2\alpha_t - 1 \leq o(T)$ and $\sum_{t=1}^{\frac{T}{2}} 1 - 2\alpha_t \leq o(T)$ since $\sum_{t=1}^{\frac{T}{2}} 4\alpha_t\beta - 2\beta + 1 - 2\alpha_t$ is a linear function of β and thus its maximum occurs at $\beta = 0$ or $\beta = 1$.

For scenario 2 $\max_{y\in\Delta_Y}\sum_{t=1}^T x_t^\top A_t y - \min_{x\in X}\max_{y\in Y}\sum_{t=1}^T x^\top A_t y \leq o(T)$ reduces to $\max_{0\leq\beta\leq 1}\sum_{t=1}^{\frac{T}{2}} 4\alpha_t\beta - 2\beta + 1 - 2\alpha_t + \frac{T}{2}(2\beta - 1) \leq o(T)$ which implies $\sum_{t=1}^{\frac{T}{2}} 2\alpha_t - 1 + \frac{T}{2} \leq o(T)$ and $\sum_{t=1}^{\frac{T}{2}} 1 - 2\alpha_t + \frac{T}{2} \leq o(T)$. Finally, notice that $\sum_{t=1}^{\frac{T}{2}} 2\alpha_t - 1 + \frac{T}{2} \leq o(T)$ implies $\frac{T}{2} \leq o(T) + \sum_{t=1}^{\frac{T}{2}} 1 - 2\alpha_t$ but from scenario 1 we have that $\sum_{t=1}^{\frac{T}{2}} 1 - 2\alpha_t \leq o(T)$ since $\frac{T}{2} \leq o(T)$ is a contradiction we get the result. □

Proof of Lemma 1. We omit the subscript t.

$$
\begin{aligned}
\|\nabla x^\top A y\|_2 &= \left\| \begin{bmatrix} \nabla_x x^\top A y \\ \nabla_y x^\top A y \end{bmatrix} \right\|_2 \\
&= \left\| \begin{bmatrix} A_{[1,:]}^\top y \\ \dots \\ A_{[d_1,:]}^\top y \\ A_{[:,1]}^\top x \\ \dots \\ A_{[:,d_2]}^\top x \end{bmatrix} \right\|_2 \\
&\leq \left\| \begin{bmatrix} A_{[1,:]}^\top y \\ \dots \\ A_{[d_1,:]}^\top y \end{bmatrix} \right\|_2 + \left\| \begin{bmatrix} A_{[:,1]}^\top x \\ \dots \\ A_{[:,d_2]}^\top x \end{bmatrix} \right\|_2 \\
&\leq \sqrt{\sum_{i=1}^{d_1} (A_{[i,:]}^\top y)^2} + \left\| \begin{bmatrix} A_{[:,1]}^\top x \\ \dots \\ A_{[:,d_2]}^\top x \end{bmatrix} \right\|_2 \\
&\leq \sqrt{d_1 (\|A_{[i,:]}\|_\infty \|y\|_1)^2} + \left\| \begin{bmatrix} A_{[:,1]}^\top x \\ \dots \\ A_{[:,d_2]}^\top x \end{bmatrix} \right\|_2 \quad \text{by Generalized Cauchy Schwartz} \\
&\leq \sqrt{c d_1} + \left\| \begin{bmatrix} A_{[:,1]}^\top x \\ \dots \\ A_{[:,d_2]}^\top x \end{bmatrix} \right\|_2 \\
&\leq \sqrt{c d_1} + \sqrt{c d_2}. \quad \text{(using the same reasoning)}
\end{aligned}
$$

The second part of the claim follows by bounding $\|\nabla x^\top A y\|_\infty$ using the same argument. □

Proof of Lemma 2.

$$\begin{aligned}
\min_{x\in X}\max_{y\in Y}\sum_{t=1}^{T}\mathcal{L}_t(x,y) &= \sum_{t=1}^{T}[\bar{\mathcal{L}}_t(x_{T+1},y_{T+1}) + \frac{1}{\eta}R_X(x_{T+1}) - \frac{1}{\eta}R_Y(y_{T+1})] \\
&\le \sum_{t=1}^{T}[\bar{\mathcal{L}}_t(\bar{x}_{T+1},y_{T+1}) + \frac{1}{\eta}R_X(\bar{x}_{T+1}) - \frac{1}{\eta}R_Y(y_{T+1})] && \text{by Equation (2)} \\
&\le \sum_{t=1}^{T}[\bar{\mathcal{L}}_t(\bar{x}_{T+1},\bar{y}_{T+1}) + \frac{1}{\eta}R_X(\bar{x}_{T+1}) - \frac{1}{\eta}R_Y(y_{T+1})] && \text{by Equation (2)} \\
&= \min_{x\in X}\max_{y\in Y}\sum_{t=1}^{T}[\bar{\mathcal{L}}_t(x,y) + \frac{T}{\eta}R_X(\bar{x}_{T+1}) - \frac{T}{\eta}R_Y(y_{T+1})] \\
&\le \min_{x\in X}\max_{y\in Y}\sum_{t=1}^{T}\bar{\mathcal{L}}_t(x,y) + \frac{T}{\eta}R_X(\bar{x}_{T+1}).
\end{aligned}$$

The other inequality can be obtained by a similar argument. ☐

Proof of Lemma 3. We first prove the second inequality. We proceed by induction. The base case $t = 1$ holds by definition of (x_2, y_2), indeed

$$\mathcal{L}_1(x_2,y_2) + G_{\mathcal{L}}\|y_1 - y_2\| \ge \mathcal{L}_1(x_2,y_2) := \min_{x\in X}\max_{y\in Y}\mathcal{L}_1(x,y).$$

We now assume the following claim holds for $T-1$:

$$\min_{x\in X}\max_{y\in Y}\sum_{t=1}^{T-1}\mathcal{L}_t(x,y) \ge \sum_{t=1}^{T-1}\mathcal{L}_t(x_{t+1},y_{t+1}) - G_{\mathcal{L}}\sum_{t=1}^{T-1}\|y_t - y_{t+1}\|, \tag{8}$$

and show it holds for T.

$$\begin{aligned}
&\min_{x\in X}\max_{y\in Y}\sum_{t=1}^{T}\mathcal{L}_t(x,y) \\
&= \sum_{t=1}^{T-1}\mathcal{L}_t(x_{T+1},y_{T+1}) + \mathcal{L}_T(x_{T+1},y_{T+1}) \\
&\ge \sum_{t=1}^{T-1}\mathcal{L}_t(x_{T+1},y_T) + \mathcal{L}_T(x_{T+1},y_T) && \text{by Equation (2)} \\
&\ge \sum_{t=1}^{T-1}\mathcal{L}_t(x_T,y_T) + \mathcal{L}_T(x_{T+1},y_T) && \text{by Equation (2)} \\
&\ge \sum_{t=1}^{T-1}\mathcal{L}_t(x_{t+1},y_{t+1}) - G_{\mathcal{L}}\sum_{t=1}^{T-1}\|y_t - y_{t+1}\| + \mathcal{L}_T(x_{T+1},y_T) && \text{by Equation (8)} \\
&= \sum_{t=1}^{T}\mathcal{L}_t(x_{t+1},y_{t+1}) - G_{\mathcal{L}}\sum_{t=1}^{T-1}\|y_t - y_{t+1}\| + \mathcal{L}_T(x_{T+1},y_T) - \mathcal{L}_T(x_{T+1},y_{T+1}) \\
&\ge \sum_{t=1}^{T}\mathcal{L}_t(x_{t+1},y_{t+1}) - G_{\mathcal{L}}\sum_{t=1}^{T-1}\|y_t - y_{t+1}\| - G_{\mathcal{L}}\|y_T - y_{T+1}\| \\
&= \sum_{t=1}^{T}\mathcal{L}_t(x_{t+1},y_{t+1}) - G_{\mathcal{L}}\sum_{t=1}^{T}\|y_t - y_{t+1}\|.
\end{aligned}$$

We now show by induction that

$$\min_{x\in X}\max_{y\in Y}\sum_{t=1}^{T}\mathcal{L}_t(x,y) \le \sum_{t=1}^{T}\mathcal{L}_t(x_{t+1},y_{t+1}) + G_{\mathcal{L}}\sum_{t=1}^{T}\|x_t - x_{t+1}\|.$$

Indeed, $t=1$ follows from the definition of (x_2,y_2). We now assume the claim holds for $T-1$ and prove it for T:

$$\begin{aligned}
&\min_{x\in X}\max_{y\in Y}\sum_{t=1}^{T}\mathcal{L}_t(x,y) \\
&= \sum_{t=1}^{T}\mathcal{L}_t(x_{T+1},y_{T+1}) \\
&\le \sum_{t=1}^{T-1}\mathcal{L}_t(x_T,y_{T+1}) + \mathcal{L}_T(x_T,y_{T+1}) && \text{by Equation (2)} \\
&\le \sum_{t=1}^{T-1}\mathcal{L}_t(x_T,y_T) + \mathcal{L}_T(x_T,y_{T+1}) && \text{by Equation (2)} \\
&\le \sum_{t=1}^{T-1}\mathcal{L}_t(x_{t+1},y_{t+1}) + G_{\mathcal{L}}\sum_{t=1}^{T-1}\|x_t - x_{t+1}\| \\
&\qquad + \mathcal{L}_T(x_T,y_{T+1}) + \mathcal{L}_T(x_{T+1},y_{T+1}) - \mathcal{L}_T(x_{T+1},y_{T+1}) && \text{by induction claim} \\
&\le \sum_{t=1}^{T}\mathcal{L}_t(x_{t+1},y_{t+1}) + G_{\mathcal{L}}\sum_{t=1}^{T}\|x_t - x_{t+1}\| && \text{since } \mathcal{L}_T \text{ is } G_{\mathcal{L}}\text{-Lipschitz.}
\end{aligned}$$

□

Proof of Lemma 3. Fix t, define $J(x,y) \triangleq \sum_{\tau=1}^{t-1}\mathcal{L}_\tau(x,y) + \mathcal{L}_t(x,y)$ and notice it is $\frac{t}{\eta}$-strongly convex strongly concave with respect to norm $\|\cdot\|$. Also notice that (x_{t+1},y_{t+1}) is the unique saddle point of $J(x,y)$.

By strong convexity of J and definition of x_{t+1} it holds that for any $x\in X$ and any $y\in Y$

$$J(x,y) \ge J(x_{t+1},y) + \nabla_x J(x_{t+1},y)^\top(x - x_{t+1}) + \frac{t}{2\eta}\|x - x_{t+1}\|^2.$$

Plugging in $y = y_{t+1}$ and recalling the KKT condition $\nabla_x J(x_{t+1},y_{t+1})^\top(x - x_{t+1}) \ge 0$, we have that for any $x\in X$

$$\frac{2\eta}{t}\big[J(x,y_{t+1}) - J(x_{t+1},y_{t+1})\big] \ge \|x - x_{t+1}\|^2. \tag{9}$$

Similarly, since J is $\frac{t}{\eta}$ strongly concave. That is, for any $y\in Y$

$$J(x_{t+1},y) \le J(x_{t+1},y_{t+1}) + \nabla_y J(x_{t+1},y_{t+1})^\top(y - y_{t+1}) - \frac{t}{2\eta}\|y - y_{t+1}\|^2.$$

Together with the KKT condition $\nabla_y J(x_{t+1},y_{t+1})^\top(y - y_{t+1}) \le 0$ we get that for any $y\in Y$

$$\frac{2\eta}{t}\big[J(x_{t+1},y_{t+1}) - J(x_{t+1},y)\big] \ge \|y - y_{t+1}\|^2. \tag{10}$$

Adding up Equations (9) and (10), plugging $x = x_t$ and $y = y_t$ we get

$$\begin{aligned}
&\frac{2\eta}{t}\big[J(x_t,y_{t+1}) - J(x_{t+1},y_t)\big] \ge \|x_t - x_{t+1}\|^2 + \|y - y_{t+1}\|^2. \\
\iff &\frac{2\eta}{t}\big[\sum_{\tau=1}^{t-1}\mathcal{L}_\tau(x_t,y_{t+1}) + \mathcal{L}_t(x_t,y_{t+1}) - [\sum_{\tau=1}^{t-1}\mathcal{L}_\tau(x_{t+1},y_t) + \mathcal{L}_t(x_{t+1},y_t)]\big]
\end{aligned}$$

$$
\begin{aligned}
&\geq \|x_t - x_{t+1}\|^2 + \|y - y_{t+1}\|^2 \\
\implies \frac{2\eta}{t}\Big[&\sum_{\tau=1}^{t-1} \mathcal{L}_\tau(x_t, y_t) + \mathcal{L}_t(x_t, y_{t+1}) - \Big[\sum_{\tau=1}^{t-1} \mathcal{L}_\tau(x_{t+1}, y_t) + \mathcal{L}_t(x_{t+1}, y_t)\Big]\Big] \\
&\geq \|x_t - x_{t+1}\|^2 + \|y - y_{t+1}\|^2,
\end{aligned}
$$

since $\sum_{\tau=1}^{t-1} \mathcal{L}_\tau(x_t, y_{t+1}) \leq \sum_{\tau=1}^{t-1} \mathcal{L}_\tau(x_t, y_t)$.

Additionally, since $\sum_{\tau=1}^{t-1} \mathcal{L}_\tau(x_t, y_t) \leq \sum_{\tau=1}^{t-1} \mathcal{L}_\tau(x_{t+1}, y_t)$, we have

$$
\begin{aligned}
&\frac{2\eta}{t}\Big[\sum_{\tau=1}^{t-1} \mathcal{L}_\tau(x_t, y_t) + \mathcal{L}_t(x_t, y_{t+1}) - \sum_{\tau=1}^{t-1} \mathcal{L}_\tau(x_t, y_t) - \mathcal{L}_t(x_{t+1}, y_t)\Big] \\
&\qquad \geq \|x_t - x_{t+1}\|^2 + \|y - y_{t+1}\|^2 \\
\iff &\frac{2\eta}{t}\big[\mathcal{L}_t(x_t, y_{t+1}) - \mathcal{L}_t(x_{t+1}, y_t)\big] \geq \|x_t - x_{t+1}\|^2 + \|y_t - y_{t+1}\|^2 \\
\iff &\frac{2\eta}{t}\big[\bar{\mathcal{L}}_t(x_t, y_{t+1}) + \frac{1}{\eta}R_X(x_t) - \frac{1}{\eta}R_Y(y_{t+1}) - \bar{\mathcal{L}}_t(x_{t+1}, y_t) - \frac{1}{\eta}R_X(x_{t+1}) + \frac{1}{\eta}R_Y(y_t)\big] \\
&\qquad \geq \|x_t - x_{t+1}\|^2 + \|y_t - y_{t+1}\|^2 \\
\implies &\frac{2\eta}{t}\big[G_{\bar{\mathcal{L}}}\|[x_t; y_{t+1}] - [x_{t+1}; y_t]\| + \frac{1}{\eta}R_X(x_t) - \frac{1}{\eta}R_X(x_{t+1}) + \frac{1}{\eta}R_Y(y_t) - \frac{1}{\eta}R_Y(y_{t+1})\big] \\
&\qquad \geq \|x_t - x_{t+1}\|^2 + \|y_t - y_{t+1}\|^2 \\
\implies &\frac{2\eta}{t}\big[G_{\bar{\mathcal{L}}}\|x_t - x_{t+1}\| + G_{\bar{\mathcal{L}}}\|y_t - y_{t+1}\| + \frac{1}{\eta}R_X(x_t) - \frac{1}{\eta}R_X(x_{t+1}) + \frac{1}{\eta}R_Y(y_t) - \frac{1}{\eta}R_Y(y_{t+1})\big] \\
&\qquad \geq \|x_t - x_{t+1}\|^2 + \|y_t - y_{t+1}\|^2 \\
\implies &\frac{2\eta}{t}\big[G_{\bar{\mathcal{L}}}\|x_t - x_{t+1}\| + G_{\bar{\mathcal{L}}}\|y_t - y_{t+1}\| + \frac{G_{R_X}}{\eta}\|x_t - x_{t+1}\| + \frac{G_{R_Y}}{\eta}\|y_t - y_{t+1}\|\big] \\
&\qquad \geq \|x_t - x_{t+1}\|^2 + \|y_t - y_{t+1}\|^2 \\
\implies &\frac{2\eta}{t}[G_{\bar{\mathcal{L}}} + \frac{1}{\eta}\max(G_{R_X}, G_{R_Y})]\big[\|x_t - x_{t+1}\| + \|y_t - y_{t+1}\|\big] \geq \|x_t - x_{t+1}\|^2 + \|y_t - y_{t+1}\|^2 \\
\iff &\frac{2\eta}{t}[G_{\bar{\mathcal{L}}} + \frac{1}{\eta}\max(G_{R_X}, G_{R_Y})] \geq \frac{\|x_t - x_{t+1}\|^2 + \|y_t - y_{t+1}\|^2}{\|x_t - x_{t+1}\| + \|y_t - y_{t+1}\|}.
\end{aligned}
$$

Finally, since x^2 is a convex function $\frac{a^2}{2} + \frac{b^2}{2} \geq \left(\frac{a+b}{2}\right)^2$, we have $a^2 + b^2 \geq \frac{(a+b)^2}{2}$. This, together with the last implication, yields the result

$$
\frac{4\eta}{t}[G_{\bar{\mathcal{L}}} + \frac{1}{\eta}\max(G_{R_X}, G_{R_Y})] \geq \|x_t - x_{t+1}\| + \|y_t - y_{t+1}\|.
$$

□

Proof of Theorem 2.

$$
\begin{aligned}
&\sum_{t=1}^{T} \bar{\mathcal{L}}_t(x_t, y_t) - \min_{x \in X}\max_{y \in Y}\sum_{t=1}^{T} \bar{\mathcal{L}}_t(x, y) \\
&\leq \sum_{t=1}^{T} \mathcal{L}_t(x_t, y_t) - \min_{x \in X}\max_{y \in Y}\sum_{t=1}^{T} \bar{\mathcal{L}}_t(x, y) + \sum_{t=1}^{T}\frac{1}{\eta}R_Y(y_t) \quad \text{by Equation 6} \\
&\leq \sum_{t=1}^{T} \mathcal{L}_t(x_t, y_t) - \min_{x \in X}\max_{y \in Y}\sum_{t=1}^{T} \mathcal{L}_t(x, y) + \sum_{t=1}^{T}\frac{1}{\eta}R_Y(y_t) + \frac{T}{\eta}R_X(x_{T+1}) \quad \text{by Lemma 2}
\end{aligned}
$$

$$
\begin{aligned}
&\leq \sum_{t=1}^{T} \mathcal{L}_t(x_t, y_t) - \sum_{t=1}^{T} \mathcal{L}_t(x_{t+1}, y_{t+1}) + \sum_{t=1}^{T} \frac{1}{\eta} R_Y(y_t) + \frac{T}{\eta} R_X(x_{T+1}) + G_{\mathcal{L}} \sum_{t=1}^{T} \|y_t - y_{t+1}\| \quad \text{by Lemma 3} \\
&\leq \sum_{t=1}^{T} G_{\mathcal{L}}(\|x_t - x_{t+1}\| + \|y_t - y_{t+1}\|) + \sum_{t=1}^{T} \frac{1}{\eta} R_Y(y_t) + \frac{T}{\eta} R_X(x_{T+1}) + G_{\mathcal{L}} \sum_{t=1}^{T} \|y_t - y_{t+1}\| \\
&\leq \sum_{t=1}^{T} G_{\mathcal{L}}(\|x_t - x_{t+1}\| + \|y_t - y_{t+1}\|) + \sum_{t=1}^{T} \frac{1}{\eta} R_Y(y_t) + \frac{T}{\eta} R_X(x_{T+1}) + G_{\mathcal{L}} \sum_{t=1}^{T} \|y_t - y_{t+1}\| \\
&\leq 2 \sum_{t=1}^{T} G_{\mathcal{L}}(\|x_t - x_{t+1}\| + \|y_t - y_{t+1}\|) + \sum_{t=1}^{T} \frac{1}{\eta} R_Y(y_t) + \frac{T}{\eta} R_X(x_{T+1}) \\
&\leq 2 \sum_{t=1}^{T} G_{\mathcal{L}}(\frac{4\eta}{t}[G_{\bar{\mathcal{L}}} + \frac{1}{\eta} \max(G_{R_X}, G_{R_Y})]) + \sum_{t=1}^{T} \frac{1}{\eta} R_Y(y_t) + \frac{T}{\eta} R_X(x_{T+1}) \\
&\leq 8 G_{\mathcal{L}} \eta [G_{\bar{\mathcal{L}}} + \frac{1}{\eta} \max(G_{R_X}, G_{R_Y})](1 + \int_1^T \frac{1}{t} dt) + \sum_{t=1}^{T} \frac{1}{\eta} R_Y(y_t) + \frac{T}{\eta} R_X(x_{T+1}) \\
&\leq 8 G_{\mathcal{L}} \eta [G_{\bar{\mathcal{L}}} + \frac{1}{\eta} \max(G_{R_X}, G_{R_Y})](1 + \ln(T)) + \frac{T}{\eta} \max_{y \in Y} R_Y(y) + \frac{T}{\eta} \max_{x \in X} R_X(x) \\
&\leq 8\eta [G_{\bar{\mathcal{L}}} + \frac{1}{\eta} \max(G_{R_X}, G_{R_Y})]^2 (1 + \ln(T)) + \frac{T}{\eta} \max_{y \in Y} R_Y(y) + \frac{T}{\eta} \max_{x \in X} R_X(x).
\end{aligned}
$$

Notice that $\min_{x\in X} \max_{y\in Y} \sum_{t=1}^T \bar{\mathcal{L}}_t(x, y) - \sum_{t=1}^T \bar{\mathcal{L}}_t(x_t, y_t)$ can be upper bounded by the same quantity using the same argument. This concludes the proof. □

Proof of Lemma 5. We need to find $G_R > 0$ such that $\|\nabla R(x)\|_\infty \leq G_R$ for all $x \in \Delta_\theta$. Notice that $[\nabla R(x)]_i = 1 + \ln(x_i)$ for $i = 1, ...d$. Moreover, since for every $i = 1, ..., d$ we have $\theta \leq x_i \leq 1$ the following sequence of inequalities hold: $\ln(\theta) \leq 1 + \ln(\theta) \leq 1 + \ln(x_i) \leq 1$. It follows that $G_R = \max\{|\ln(\theta)|, 1\}$. □

Proof of Lemma 6. Choose $z^* = [1; 0; 0; ...; 0; 0]$, it is easy to see that $z_p^* = [1 - \theta(d-1); \theta; \theta; ...; \theta, \theta]$ and $\|z^* - z_p^*\|_1 = 2\theta(d-1)$. □

Proof of Lemma 7. Let (x^*, y^*) be any saddle point pair for $\sum_{t=1}^T \bar{\mathcal{L}}_t(x, y)$ with $x^* \in \Delta, y^* \in \Delta$. Let (x_θ^*, y_θ^*) be any saddle point pair for $\sum_{t=1}^T \bar{\mathcal{L}}_t(x, y)$ with $x_\theta^* \in \Delta, y_\theta^* \in \Delta$. Let x_p^*, y_p^* be the projection of x^*, y^* onto the respective simplexes using the $\|\cdot\|_\infty$ norm. We first show the second inequality. Notice that

$$
\begin{aligned}
\sum_{t=1}^{T} \bar{\mathcal{L}}_t(x^*, y^*) &\leq \sum_{t=1}^{T} \bar{\mathcal{L}}_t(x_\theta^*, y^*) \\
&\leq \sum_{t=1}^{T} \bar{\mathcal{L}}_t(x_\theta^*, y_p^*) + G_{\bar{\mathcal{L}}} T \|y_p^* - y^*\|_1 \\
&\leq \sum_{t=1}^{T} \bar{\mathcal{L}}_t(x_\theta^*, y_\theta^*) + G_{\bar{\mathcal{L}}} T \|y_p^* - y^*\|_1.
\end{aligned}
$$

To show the first inequality in the statement of the lemma notice that

$$
\begin{aligned}
\sum_{t=1}^{T} \bar{\mathcal{L}}_t(x^*, y^*) &\geq \sum_{t=1}^{T} \bar{\mathcal{L}}_t(x^*, y_\theta^*) \\
&\geq \sum_{t=1}^{T} \bar{\mathcal{L}}_t(x_p^*, y_\theta^*) - G_{\bar{\mathcal{L}}} T \|x_p^* - x^*\|_1
\end{aligned}
$$

$$\geq \sum_{t=1}^{T} \bar{\mathcal{L}}_t(x_\theta^*, y_\theta^*) - G_{\bar{\mathcal{L}}} T \|x_p^* - x^*\|_1.$$

This concludes the proof. □

Proof of Theorem 3. For convenience set $\bar{\mathcal{L}}_t(x,y) = x^\top A_t y$. Let (x^*, y^*) be any saddle point of $\min_{x\in\Delta}\max_{y\in\Delta}\sum_{t=1}^T x^\top A_t y$, let (x_p^*, y_p^*) be the respective projections onto Δ_θ using $\|\cdot\|_\infty$ norm. By the choice of θ we have that $|\ln(\theta)| > 1$ additionally, notice that $\max_{z\in\Delta_\theta}\sum_{i=1}^d z_i \ln(z_i) + \ln(d) \leq 0 + \ln(d)$ by Jensen's inequality.

$$\begin{aligned}
&\sum_{t=1}^T x_t^\top A_t y_t - \min_{x\in\Delta}\max_{y\in\Delta}\sum_{t=1}^T x^\top A_t y \\
&\leq \sum_{t=1}^T x_t^\top A_t y_t - \min_{x\in\Delta_\theta}\max_{y\in\Delta_\theta}\sum_{t=1}^T x^\top A_t y + G_{\bar{\mathcal{L}}} T\|x^* - x_p^*\|_1 \quad \text{by Lemma 7} \\
&\leq \sum_{t=1}^T x_t^\top A_t y_t - \min_{x\in\Delta_\theta}\max_{y\in\Delta_\theta}\sum_{t=1}^T x^\top A_t y + 2G_{\bar{\mathcal{L}}} T\theta(d_1 - 1) \quad \text{by Lemma 6} \\
&\leq 8\eta[G_{\bar{\mathcal{L}}} + \frac{1}{\eta}\max(G_{R_X}, G_{R_Y})]^2(1+\ln(T)) + \frac{T}{\eta}\max_{y\in\Delta_\theta} R_Y(y) + \frac{T}{\eta}\max_{x\in\Delta_\theta} R_X(x) + 2G_{\bar{\mathcal{L}}} T\theta(d_1-1) \quad \text{by Theorem 2} \\
&\leq 8\eta[G_{\bar{\mathcal{L}}} + \frac{|\ln(\theta)|}{\eta}]^2(1+\ln(T)) + \frac{T}{\eta}\max_{y\in\Delta_\theta} R_Y(y) + \frac{T}{\eta}\max_{x\in\Delta_\theta} R_X(x) + 2G_{\bar{\mathcal{L}}} T\theta(d_1-1) \\
&\leq 32\eta G_{\bar{\mathcal{L}}}^2(1+\ln(T)) + \frac{T}{\eta}\max_{y\in\Delta_\theta} R_Y(y) + \frac{T}{\eta}\max_{x\in\Delta_\theta} R_X(x) + 2G_{\bar{\mathcal{L}}} T e^{-\eta G_{\bar{\mathcal{L}}}}(d_1-1) \quad \text{by the choice of } \theta \\
&\leq 32\eta G_{\bar{\mathcal{L}}}^2(1+\ln(T)) + \frac{T}{\eta}\ln(d_2) + \frac{T}{\eta}\ln(d_1) + 2G_{\bar{\mathcal{L}}} T e^{-\eta G_{\bar{\mathcal{L}}}}(d_1-1) \\
&\leq 32 G_{\bar{\mathcal{L}}}\sqrt{T}(1+\ln(T)) + \sqrt{T}(\ln d_1 + \ln d_2) + 2d_1 G_{\bar{\mathcal{L}}} T e^{-\sqrt{T}} \\
&= O\left(\ln(T)\sqrt{T} + \sqrt{T}\max\{\ln d_1, \ln d_2\}\right) + \quad o(1)\max\{d_1, d_2\}.
\end{aligned}$$

The last line follows because $G_{\bar{\mathcal{L}}} \leq 1$, since each entry of A is bounded between $[-1, 1]$. A symmetrical argument yields the other side of the inequality. □

Proof of Lemma 9.

$$\begin{aligned}
&\mathbb{E}[\sum_{t=1}^T x_t^\top \hat{A}_t y_t] \\
&= \mathbb{E}[\sum_{t=1}^{T-1} x_t^\top \hat{A}_t y_t] + \mathbb{E}[x_T^\top \hat{A}_T y_T] \\
&= \mathbb{E}[\sum_{t=1}^{T-1} x_t^\top \hat{A}_t y_t] + \mathbb{E}[\mathbb{E}[x_T^\top \hat{A}_T y_T | \tau = 1, ..., T-1]] \\
&= \mathbb{E}[\sum_{t=1}^{T-1} x_t^\top \hat{A}_t y_t] + \mathbb{E}[x_T^\top \mathbb{E}[\hat{A}_T | \tau = 1, ..., T-1] y_T] \\
&= \mathbb{E}[\sum_{t=1}^{T-1} x_t^\top \hat{A}_t y_t] + \mathbb{E}[x_T^\top A_T y_T] \quad \text{by Theorem 4.}
\end{aligned}$$

Repeating the argument $T-1$ more times yields the result. □

Proof of Lemma 10. Let us fist bound $|\sum_{t=1}^T x^\top A_t y - \sum_{t=1}^T x^\top \hat{A}_t y|$ for any $x \in \Delta_X$ and $y \in \Delta_Y$ with probability 1.

$$
\begin{aligned}
&|\sum_{t=1}^T x^\top A_t y - \sum_{t=1}^T x^\top \hat{A}_t y| \\
&= |x^\top(\sum_{t=1}^T A_t y - \sum_{t=1}^T \hat{A}_t y)| \\
&\le \|x\|_2 \|\sum_{t=1}^T A_t y - \hat{A}_t y\|_2 \\
&\le \|\sum_{t=1}^T A_t y - \hat{A}_t y\|_2
\end{aligned}
$$

It now follows that

$$
\begin{aligned}
&\sum_{t=1}^T x^\top \hat{A}_t y \le \sum_{t=1}^T x^\top A_t y + \|\sum_{t=1}^T A_t y - \hat{A}_t y\|_2 \\
\implies &\min_{x\in\Delta_{X,\delta}} \sum_{t=1}^T x^\top \hat{A}_t y \le \sum_{t=1}^T x^\top A_t y + \|\sum_{t=1}^T A_t y - \hat{A}_t y\|_2 \quad \forall x \in \Delta_{X,\delta}, y \in \Delta_{Y,\delta} \\
\implies &\min_{x\in\Delta_{X,\delta}} \sum_{t=1}^T x^\top \hat{A}_t y \le \max_{y\in\Delta_{Y,\delta}} \sum_{t=1}^T x^\top A_t y + \|\sum_{t=1}^T A_t y - \hat{A}_t y\|_2 \quad \forall x \in \Delta_{X,\delta}, y \in \Delta_{Y,\delta} \\
\implies &\max_{y\in\Delta_{Y,\delta}} \min_{x\in\Delta_{X,\delta}} \sum_{t=1}^T x^\top \hat{A}_t y \le \min_{x\in\Delta_{X,\delta}} \max_{y\in\Delta_{Y,\delta}} \sum_{t=1}^T x^\top A_t y + \|\sum_{t=1}^T A_t y - \hat{A}_t y\|_2 \quad \forall x \in \Delta_{X,\delta}, y \in \Delta_{Y,\delta}.
\end{aligned}
$$

This concludes the proof as $\max_{y\in\Delta_{Y,\delta}} \min_{x\in\Delta_{X,\delta}} \sum_{t=1}^T x^\top \hat{A}_t y = \min_{x\in\Delta_{X,\delta}} \max_{y\in\Delta_{Y,\delta}} \sum_{t=1}^T x^\top \hat{A}_t y$ (since the function is convex-concave and the sets Δ_Y^δ and Δ_X^δ are convex and compact), the other side of the inequality can be obtained using the other inequality follows from applying the same reasoning. □

Proof of Lemma 11. For any y define $\alpha_t \triangleq A_t y - \hat{A}_t y$. We first show that for all t, t' such that $t < t'$ it holds that $\mathbb{E}[\alpha_t^\top \alpha_{t'}] = 0$. Indeed

$$
\begin{aligned}
\mathbb{E}[\alpha_t^\top \alpha_{t'}] &= \mathbb{E}[(A_t y - \hat{A}_t y)^\top (A_{t'} y - \hat{A}_{t'} y)] \\
&= \mathbb{E}[(A_t y)^\top A_{t'} y - (A_t y)^\top \hat{A}_{t'} y - (\hat{A}_t y)^\top A_{t'} y + (\hat{A}_t y)^\top \hat{A}_{t'} y] \\
&= (A_t y)^\top A_{t'} y - (A_t y)^\top A_{t'} y - (A_t y)^\top A_{t'} y + \mathbb{E}[(\hat{A}_t y)^\top \hat{A}_{t'} y] \\
&= (A_t y)^\top A_{t'} y - (A_t y)^\top A_{t'} y - (A_t y)^\top A_{t'} y + (A_t y)^\top A_{t'} y \\
&= 0,
\end{aligned}
$$

where the second to last line follows since

$$
\begin{aligned}
\mathbb{E}[(\hat{A}_t y)^\top \hat{A}_{t'} y] &= \mathbb{E}_{1,\dots,t'-1}[\mathbb{E}[(\hat{A}_t y)^\top \hat{A}_{t'} y | \tau = 1, \dots, t'-1]] \\
&= \mathbb{E}_{1,\dots,t'-1}[(\hat{A}_t y)^\top \mathbb{E}[\hat{A}_{t'} y | \tau = 1, \dots, t'-1]] \\
&= \mathbb{E}_{1,\dots,t'-1}[(\hat{A}_t y)^\top A_{t'} y] \\
&= (A_t y)^\top A_{t'} y.
\end{aligned}
$$

Now,

$$
\mathbb{E}[\|\sum_{t=1}^T A_t y - \hat{A}_t y\|_2] = \sqrt{\mathbb{E}[\|\sum_{t=1}^T \alpha_t\|_2]^2}
$$

$$
\begin{aligned}
&\leq \sqrt{\mathbb{E}[\|\sum_{t=1}^{T}\alpha_t\|_2^2]} \quad \text{by Jensen's Inequality} \\
&= \sqrt{\sum_{t=1}^{T}\mathbb{E}[\|\alpha_t\|_2^2] + 2\sum_{t<t'}\mathbb{E}[\alpha_t^\top \alpha_{t'}]} \\
&= \sqrt{\sum_{t=1}^{T}\mathbb{E}[\|A_t y - \hat{A}_t y\|_2^2]} \\
&\leq \sqrt{\sum_{t=1}^{T}\mathbb{E}[2\|A_t y\|^2 + 2\|\hat{A}_t y\|_2^2]}
\end{aligned}
$$

We proceed to bound $\|\hat{A}_t y\|_2$, the upper bound we obtain will also bound $\|A_t y\|$ because of the following fact. If the random vector $\tilde{a}$ satisfies $\|\tilde{a}\| \leq c$ for some constant c with probability 1 then $\|\mathbb{E}\tilde{a}\| \leq c$. Indeed by Jensen's inequality we have that $\|\mathbb{E}\tilde{a}\| \leq \mathbb{E}\|\tilde{a}\| \leq c$. Let us omit the subscript t for the rest of the proof. Let $\hat{A}_{[i,:]}$ be the i-th row of matrix $\hat{A}$.

$$
\begin{aligned}
\|\hat{A}y\|_2 &= \sqrt{\sum_{i=1}^{d_1}\big[\sum_{j=1}^{d_2}\hat{a}_{i,j}y_j\big]^2} \\
&\leq \sum_{i=1}^{d_1}\sqrt{\big[\sum_{j=1}^{d_2}\hat{a}_{i,j}y_j\big]^2} \\
&= \sum_{i=1}^{d_1}\big|\sum_{j=1}^{d_2}\hat{a}_{i,j}y_j\big| \\
&\leq \sum_{i=1}^{d_1}\|\hat{A}_{[i,:]}\|_\infty\|y\|_1 \quad \text{by generalized Cauchy Schwartz} \\
&\leq d_1 \max_{i,j}\big|\frac{A_{i,j}}{\delta^2}\big| \quad \text{by definition of } \hat{A} \text{ and using the fact that } x_t \in \Delta_{X,\delta} \text{ and } y_t \in \Delta_{Y,\delta} \\
&\leq \frac{d_1}{\delta^2}.
\end{aligned}
$$

Notice the upper bound $\frac{d_2}{\delta^2}$ can also be obtained by interchanging the summations and repeating the argument. This yields the desired result. □

Proof of Theorem 5. We first focus on one side of the inequality,

$$
\begin{aligned}
&\mathbb{E}[\sum_{t=1}^{T} e_{x,t}^\top A_t e_{y,t} - \min_{x\in X}\max_{y\in Y}\sum_{t=1}^{T}x^\top A_t y] \\
&= \mathbb{E}[\sum_{t=1}^{T} e_{x,t}^\top A_t e_{y,t}] - \mathbb{E}[\min_{x\in X}\max_{y\in Y}\sum_{t=1}^{T}x^\top A_t y] \\
&= \mathbb{E}[\sum_{t=1}^{T} x_t^\top A_t y_t] - \mathbb{E}[\min_{x\in X}\max_{y\in Y}\sum_{t=1}^{T}x^\top A_t y] \quad \text{by Lemma 8} \\
&= \mathbb{E}[\sum_{t=1}^{T} x_t^\top A_t y_t] - \mathbb{E}[\min_{x\in \Delta_X^\delta}\max_{y\in \Delta_Y^\delta}\sum_{t=1}^{T}x^\top A_t y] + 2\delta G_{\tilde{\mathcal{L}}}^{\|\cdot\|_1}(d_1-1)T \quad \text{by Lemmas 6 and 7} \\
&\leq \mathbb{E}[\sum_{t=1}^{T} x_t^\top A_t y_t] - \mathbb{E}[\min_{x\in \Delta_X^\delta}\max_{y\in \Delta_Y^\delta}\sum_{t=1}^{T}x^\top \hat{A}_t y] + \frac{2\sqrt{T}\min(d_1,d_2)}{\delta^2} + 2\delta G_{\tilde{\mathcal{L}}}^{\|\cdot\|_1}(d_1-1)T \quad \text{by Lemmas 10 and 11}
\end{aligned}
$$

$$\leq \mathbb{E}[\sum_{t=1}^{T} x_t^\top \hat{A}_t y_t] - \mathbb{E}[\min_{x \in \Delta_X^\delta} \max_{y \in \Delta_Y^\delta} \sum_{t=1}^{T} x^\top \hat{A}_t y] + \frac{2\sqrt{T}\min(d_1, d_2)}{\delta^2} + 2\delta G_{\bar{\mathcal{L}}}^{\|\cdot\|_1}(d_1 - 1)T \quad \text{by Lemma 9}$$

$$\leq 8\eta[G_{\hat{\mathcal{L}}}^{\|\cdot\|_1} + \frac{|\ln(\delta)|}{\eta}]^2(1 + \ln(T)) + \frac{T}{\eta}(\ln(d_1) + \ln(d_2))$$

$$+ \frac{2\sqrt{T}\min(d_1, d_2)}{\delta^2} + 2\delta G_{\bar{\mathcal{L}}}^{\|\cdot\|_1}(d_1 - 1)T \quad \text{as in the proof of Theorem 3}$$

$$= 8\eta[\frac{1}{\delta^2} + \frac{|\ln(\delta)|}{\eta}]^2(1 + \ln(T)) + \frac{T}{\eta}(\ln(d_1) + \ln(d_2)) + \frac{2\sqrt{T}\min(d_1, d_2)}{\delta^2} + 2\delta(d_1 - 1)T \quad \text{by Lemma 1}$$

$$= O((d_1 + d_2)\ln(T)T^{5/6}) \quad \text{after plugging in } \delta = \frac{1}{T^{1/6}}, \eta = T^{1/6}$$

The other side of the inequality follows by a symmetrical argument. □