A. Proofs

Proof of Lemma 3. Consider the Bellman equation

$$\lambda + V_{\pi,\ell}(x,a) = \ell(x,a) + V_{\pi,\ell}(Ax+Ba, \pi(Ax+Ba)) \,.$$

We prove the lemma by showing that the given quadratic form is the unique solution of the Bellman equation.

Let $z = (x \quad a)$ and

$$z' = \begin{pmatrix} Ax+Ba \\ -K(Ax+Ba)+c \end{pmatrix} = \begin{pmatrix} I \\ -K \end{pmatrix} \begin{pmatrix} A & B \end{pmatrix} \begin{pmatrix} x \\ a \end{pmatrix} + \begin{pmatrix} 0 \\ c \end{pmatrix} \,.$$

We guess a quadratic form for the value functions and write

$$\lambda + z^\top P z + L^\top z = (x-g_*)^\top Q (x-g_*) + a^\top a + z'^\top P z' + L^\top z' \,.$$

The above equation has a solution if

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} = \begin{pmatrix} A^\top \\ B^\top \end{pmatrix} \begin{pmatrix} I & -K^\top \end{pmatrix} P \begin{pmatrix} I \\ -K \end{pmatrix} \begin{pmatrix} A & B \end{pmatrix} + \begin{pmatrix} Q & 0 \\ 0 & I \end{pmatrix} , \tag{11}$$

and

$$L^\top = \begin{pmatrix} L_1^\top & L_2^\top \end{pmatrix} = \left(L^\top + 2 \begin{pmatrix} 0 & c^\top \end{pmatrix} P \right) \begin{pmatrix} I \\ -K \end{pmatrix} \begin{pmatrix} A & B \end{pmatrix} - \begin{pmatrix} 2g_*^\top Q & 0 \end{pmatrix} , \tag{12}$$

and

$$\lambda = g_*^\top Q g_* + c^\top P_{22} c + L_2^\top c \,.$$

We have that

$$\left\| \begin{pmatrix} A & B \end{pmatrix} \begin{pmatrix} I \\ -K \end{pmatrix} \right\| = \|A - BK\| < 1 \,.$$

This implies that iterative equations (11) and (12) have a unique solution. Thus, the quadratic form is the solution of the Bellman equation. □

Proof of Lemma 4. From Lemma 3, we have that

$$P_t = \begin{pmatrix} A^\top \\ B^\top \end{pmatrix} \begin{pmatrix} I & -K_t^\top \end{pmatrix} P_t \begin{pmatrix} I \\ -K_t \end{pmatrix} \begin{pmatrix} A & B \end{pmatrix} + \begin{pmatrix} Q & 0 \\ 0 & I \end{pmatrix}$$

and

$$L_t^\top = \left(L_t^\top + 2 \begin{pmatrix} 0 & c_t^\top \end{pmatrix} P_t \right) \begin{pmatrix} I \\ -K_t \end{pmatrix} \begin{pmatrix} A & B \end{pmatrix} - \begin{pmatrix} 2g_t^\top Q & 0 \end{pmatrix} \,.$$

Notice that the value of P_t depends only on the values of A, B, and K_t, which in turn, by Lemma 2, depend only on $\{K_1, P_1, \dots, P_{t-1}\}$. Thus, matrix P_t is determined by K_1 independently of the adversarial choices $\{g_1, \dots, g_t\}$.

In the absence of adversarial vectors, the optimal policy has the form of $\pi(x) = -K_* x$, where $K_* = (I + B^\top S B)^{-1} B^\top S A$ and S is the solution of the Riccati equation. Consider a problem where $g_1 = g_2 = \cdots = 0$, $c_1 = c_2 = \cdots = 0$, and $K_1 = K_*$ is the gain matrix of the optimal policy. Then, V_1 is the value function of the optimal policy. Because π_2 is the greedy policy with respect to V_1, it is the optimal policy and thus K_2 is also the gain matrix of the optimal policy, and so $K_2 = K_1$. Repeating the same argument shows that all gain matrices are the same. Thus, if we choose K_1 to be the optimal gain matrix in the non-adversarial problem, we will get $K_1 = \cdots = K_t$ and hence $P_1 = P_2 = \cdots = P_t$. □

Proof of Lemma 7. First we prove (i). Under policy $\pi_t(x) = -K_* x + c_t$, we have that

$$\left(x_\infty^{\pi_t},\, \pi_t(x_\infty^{\pi_t}) \right) = \left(A x_\infty^{\pi_t} + B\pi_t(x_\infty^{\pi_t}),\, \pi_t(A x_\infty^{\pi_t} + B\pi_t(x_\infty^{\pi_t})) \right) \,.$$

Thus, by (1) and (7),

$$\begin{aligned}\lambda &= (x_\infty^{\pi_t} - g_t)^\top Q(x_\infty^{\pi_t} - g_t) + (-K_* x_\infty^{\pi_t} + c_t)^\top (-K_* x_\infty^{\pi_t} + c_t) \\ &= g_t^\top Q g_t + c_t^\top (I + B^\top (I - A + BK_*)^{-\top}(Q + K_*^\top K_*)(I - A + BK_*)^{-1} B) c_t \\ &\quad + 2(-g_t^\top Q - c_t^\top K_*)(I - A + BK_*)^{-1} B c_t \,.\end{aligned}$$

Then (5) implies that

$$\begin{aligned}L_{t,2}^\top &= 2(-g_t^\top Q - c_t^\top K_*)(I - A + BK_*)^{-1} B \,, \\ P_{*,22} &= I + B^\top (I - A + BK_*)^{-\top}(Q + K_*^\top K_*)(I - A + BK_*)^{-1} B \,.\end{aligned}$$

By Lemmas 2 and 4, $c_t = -\frac{1}{2} P_{*,22}^{-1} \left(\frac{1}{t-1} \sum_{s=1}^{t-1} L_{s,2} \right)$. Thus,

$$\begin{aligned}c_t &= -P_{*,22}^{-1} \left(\frac{1}{t-1} \sum_{s=1}^{t-1} L_{s,2} \right) \\ &= -\frac{P_{*,22}^{-1} B^\top}{t-1} (I - A + BK_*)^{-\top} \sum_{s=1}^{t-1} (-Q g_s - K_*^\top c_s) \\ &= \frac{1}{t-1} \left(D \sum_{s=1}^{t-1} g_s + H \sum_{s=1}^{t-1} c_s \right) \,, \qquad (13)\end{aligned}$$

where $H = P_{*,22}^{-1} B^\top (I - A + BK_*)^{-\top} K_*^\top$. To obtain a bound on $\max_t \|c_t\|$ from the above equation, we need to show that $\|H\|$ is sufficiently smaller than one. Let $N = (I - A + BK_*)^{-1}$, $M = K_* N B$, and $L = (I + M^\top M)^{-1} M^\top$. We have that

$$\begin{aligned}H &= (I + B^\top N^\top (Q + K_*^\top K_*) N B)^{-1} M^\top \\ &\prec (I + B^\top N^\top K_*^\top K_* N B)^{-1} M^\top \\ &= (I + M^\top M)^{-1} M^\top \\ &= L \,, \qquad (14)\end{aligned}$$

and

$$\begin{aligned}LL^\top &= (I + M^\top M)^{-1} M^\top M (I + M^\top M)^{-1} \\ &= (I + M^\top M)^{-1} (M^\top M + I - I)(I + M^\top M)^{-1} \\ &= (I + M^\top M)^{-1} \left(I - (I + M^\top M)^{-1} \right) \,.\end{aligned}$$

Because $\left\|M^\top M\right\| = \lambda_{\max}(M^\top M)$, $\|N\| \le 1/(1-\rho)$, and $\left\|M^\top M\right\| \le \|K_*\|^2 \|B\|^2 / (1-\rho)^2$, we get that

$$\begin{aligned}\left\|LL^\top\right\| &\le \left\|(I + M^\top M)^{-1}\right\| \left\|I - (I + M^\top M)^{-1}\right\| \\ &\le 1 - \frac{1}{1 + \|M^\top M\|} \\ &\le 1 - \frac{1}{1 + \|K_*\|^2 \|B\|^2 / (1-\rho)^2} \\ &= \frac{\|K_*\|^2 \|B\|^2 / (1-\rho)^2}{1 + \|K_*\|^2 \|B\|^2 / (1-\rho)^2} \,.\end{aligned}$$

By (14) and the above inequality, we get that

$$\begin{aligned}\|H\| &\le \|L\| = \left\|L^\top\right\| = \sqrt{\lambda_{\max}(LL^\top)} = \sqrt{\|LL^\top\|} \\ &\le \frac{\|K_*\| \|B\| / (1-\rho)}{\sqrt{1 + \|K_*\|^2 \|B\|^2 / (1-\rho)^2}} \,.\end{aligned}$$

Let $v = 1/(1-\|H\|)$. We get that

$$
\begin{aligned}
v &\le \frac{1}{1-\frac{\|K_*\|\|B\|/(1-\rho)}{\sqrt{1+\|K_*\|^2\|B\|^2/(1-\rho)^2}}}\\
&= \frac{\sqrt{1+\|K_*\|^2\|B\|^2/(1-\rho)^2}}{\sqrt{1+\|K_*\|^2\|B\|^2/(1-\rho)^2}-\|K_*\|\|B\|/(1-\rho)}\\
&= \sqrt{1+\|K_*\|^2\|B\|^2/(1-\rho)^2}\left(\sqrt{1+\|K_*\|^2\|B\|^2/(1-\rho)^2}+\frac{\|K_*\|\|B\|}{1-\rho}\right)\\
&= H' \,.
\end{aligned}
$$

Now we are ready to bound $\|c_t\|$. By (13), we get that for any $t \ge 1$,

$$
\|c_t\| \le \|D\| G + \frac{1}{t-1}\sum_{s=1}^{t-1}\|c_s\| \le \|D\| G + \|H\| \max_{s\ge 1}\|c_s\| \ .
$$

Thus, $\max_{t\ge 1}\|c_t\| \le \|D\| G + \|H\| \max_{t\ge1}\|c_t\|$ and thus, $\max_{t\ge1}\|c_t\| \le \frac{\|D\|G}{1-\|H\|} \le \|D\| GH' = C$.

Proof of (ii). First we write c_t in terms of c_{t-1}:

$$
\begin{aligned}
c_t &= \frac{1}{t-1}\left(D\sum_{s=1}^{t-1} g_s + H\sum_{s=1}^{t-1} c_s\right)\\
&= \frac{Dg_{t-1}}{t-1} + \frac{Hc_{t-1}}{t-1} + \frac{t-2}{t-1}\left(\frac{D}{t-2}\sum_{s=1}^{t-2} g_s + \frac{H}{t-2}\sum_{s=1}^{t-2} c_s\right)\\
&= \frac{Dg_{t-1}}{t-1} + \frac{Hc_{t-1}}{t-1} + \frac{t-2}{t-1}c_{t-1}\\
&= \frac{1}{t-1}(Dg_{t-1} + ((t-2)I + H)c_{t-1}) \,.
\end{aligned}
$$

This implies that $c_t - c_{t-1} = \frac{1}{t-1}(Dg_{t-1} - (I-H)c_{t-1})$. Then we use the facts that $\|c_t\| \le C$ and $\|H\| < 1$ to obtain

$$
\|c_t - c_{t-1}\| \le \frac{\|D\| G + 2C}{t-1} \ .
$$

□

Proof of Lemma 8. Let $f^\pi : \mathcal{X} \to \mathcal{X}$ be the transition function under policy $\pi = (K, c)$, i.e. $f^\pi(x) = (A - BK)x + Bc$. Let $\epsilon_{k,t} = \|x_k - x_\infty^{\pi_t}\|$ and $\epsilon_t = \|x_t - x_\infty^{\pi_t}\|$ denote the difference between the state variable and the limiting state under the chosen policy. We write[4]

$$
\begin{aligned}
\epsilon_{k,t} &= \|f^{\pi_k}(x_{k-1}) - f^{\pi_t}(x_{k-1}) + f^{\pi_t}(x_{k-1}) - x_\infty^{\pi_t}\|\\
&\le \|f^{\pi_k}(x_{k-1}) - f^{\pi_t}(x_{k-1})\| + \|f^{\pi_t}(x_{k-1}) - f^{\pi_t}(x_\infty^{\pi_t})\| \ .
\end{aligned}
$$

From this decomposition, we get that

$$
\begin{aligned}
\epsilon_{k,t} &\le \|B\| \|c_k - c_t\| + \|f^{\pi_t}(x_{k-1}) - f^{\pi_t}(x_\infty^{\pi_t})\|\\
&\le \|B\| \|c_k - c_t\| + \rho \|x_{k-1} - x_\infty^{\pi_t}\|\\
&\le \|B\| (\|D\| G + 2C)\sum_{s=k}^{t-1}\frac{1}{s} + \rho \|x_{k-1} - x_\infty^{\pi_t}\| \ .
\end{aligned}
$$

[4] A similar decomposition, but with a different norm, was used in (Even-Dar et al., 2009, proof of Lemma 5.2.) to bound the difference between the stationary distribution of the chosen policy and the distribution of the state variable in a finite MDP problem.

Thus,

$$\begin{aligned}\epsilon_t &\le \|B\| \left(\|D\| G + 2C\right) \sum_{k=1}^{t} \rho^{t-k} \sum_{s=k}^{t-1} \frac{1}{s} + \rho^{t-1} \left\|x_1 - x_\infty^{\pi_t}\right\| \\ &= \|B\| \left(\|D\| G + 2C\right) \sum_{s=1}^{t-1} \frac{1}{t-s} \sum_{k=s}^{t-1} \rho^k + \rho^{t-1} \frac{\|B\| C}{1-\rho} \\ &\le \frac{\|B\| \left(\|D\| G + 2C\right)}{1-\rho} \sum_{s=1}^{t-1} \frac{\rho^s}{t-s} + \rho^{t-1} \frac{\|B\| C}{1-\rho} \,,\end{aligned}$$

where the second step follows from Equation (7), Lemma 7, and the fact that $x_1 = 0$. If $t > \lceil \log(T-1)/\log(1/\rho) \rceil$, we get that

$$\begin{aligned}\sum_{s=1}^{t-1} \frac{\rho^s}{t-s} &= \sum_{s:\rho^s \le 1/(t-1)} \frac{\rho^s}{t-s} + \sum_{s:1>\rho^s>1/(t-1)} \frac{\rho^s}{t-s} \\ &\le \frac{1}{t-1} \sum_{s=1}^{t-1} \frac{1}{t-s} + \frac{\log(t-1)}{\log(1/\rho)} \left(\frac{1}{t - \log(t-1)/\log(1/\rho)} \right) \\ &\le \frac{1+\log(t-1)}{t-1} + \frac{\log(t-1)}{\log(1/\rho)} \left(\frac{1}{t - \log(t-1)/\log(1/\rho)} \right) \,.\end{aligned}$$

Thus,

$$\begin{aligned}\epsilon_t \le &\frac{\|B\| \left(\|D\| G + 2C\right)}{1-\rho} \left(\frac{1+\log(t-1)}{t-1} + \frac{\log(t-1)}{\log(1/\rho)} \left(\frac{1}{t - \log(t-1)/\log(1/\rho)} \right) \right) \\ &+ \rho^{t-1} \frac{\|B\| C}{1-\rho} \,.\end{aligned}$$

To prove the second part of lemma, let $u_T = \lceil \log(T-1)/\log(1/\rho) \rceil$. We have that

$$\sum_{t>u_T} \frac{1}{t - \log(T-1)/\log(1/\rho)} \le \sum_{t>u_T} \frac{1}{t-u_T} \le \sum_{t=1}^{T-u_T} \frac{1}{t} \le \sum_{t=1}^{T} \frac{1}{t} \le 1 + \log(T) \,. \tag{15}$$

Thus, by (8) and (15),

$$\begin{aligned}\sum_{t=1}^{T} \epsilon_t &\le \sum_{t \le u_T} \epsilon_t + \sum_{t>u_T} \epsilon_t \\ &\le \frac{1}{1-\rho} \Bigg(4\|B\| C \left\lceil \frac{\log T}{\log(1/\rho)} \right\rceil + \frac{\|B\| C}{1-\rho} \\ &\qquad + \|B\| \left(\|D\| G + 2C\right)(1+\log T) \left(1 + \log T + \frac{\log T}{\log(1/\rho)} \right) \Bigg) \,.\end{aligned}$$

$\square$

The fact that all gain matrices are identical greatly simplifies the boundedness proof.

Proof of Lemma 11. First, it is easy to verify that $P_{*,22} \succ I$ and thus, $H(V_t) = P_{*,22} \succ 2I$. The gradient of the value function can be written as

$$\nabla_a V_t(x_\infty^\pi, a) = 2P_{*,22} a + P_{*,21} x_\infty^\pi + L_{t,2}^\top \,.$$

Thus, $\|\nabla_a V_t(x_\infty^\pi, a)\| \le F$ for any $\|a\| \le U$.

Proof of (i). By (8), $\|x_t\| \le X$, and by Lemma 7, $\|c_t\| \le C$. Thus, all actions are bounded by

$$\|a_t\| = \|-K_* x_t + c_t\| \le \|K_*\| X + C \le U \ .$$

Proof of (ii) and (iii). By Lemma 6,

$$\|-K x_\infty^\pi + c\| \le K' X' + C' \le U \ .$$

Similarly,

$$\|-K_* x_\infty^\pi + c_t\| \le \|K_*\| X' + C \le U \ .$$

Proof of (iv). By (4) and the fact that $K_t = K_*$ and $P_t = P_*$, we get that

$$\|L_t\| \le \frac{2}{1-\rho} \left(G \|Q\| + \rho C \|P_*\| \right) \ .$$

Further, by (2), for any policy $\pi \in \Pi$ and any action satisfying $\|a\| \le U$, the value functions are bounded by

$$\begin{aligned} V_t(x_\infty^\pi, a) &= \begin{pmatrix} x_\infty^{\pi\top} & a^\top \end{pmatrix} P_* \begin{pmatrix} x_\infty^\pi \\ a \end{pmatrix} + L_t^\top \begin{pmatrix} x_\infty^\pi \\ a \end{pmatrix} \\ &\le \|P_*\| (X' + U)^2 + \frac{2}{1-\rho} \left(G \|Q\| + \rho C \|P_*\| \right) (X' + U) \\ &= V \ . \end{aligned}$$

□

Proof of Lemma 13. For policy $\pi = (K, c)$, we have $\ell_t(x, \pi) = x^\top (Q + K^\top K) x - 2(c^\top K + g_t^\top Q) x + c^\top c + g_t^\top Q g_t$. Define $S = Q + K^\top K$ and $d_t = 2(c^\top K + g_t^\top Q)$. We write

$$\begin{aligned} \gamma_T &= \sum_{t=1}^T \left(x_\infty^{\pi\top} S x_\infty^\pi - d_t x_\infty^\pi \right) - \sum_{t=1}^T \left(x_t^{\pi\top} S x_t^\pi - d_t x_t^\pi \right) \\ &= \sum_{t=1}^T d_t (x_t^\pi - x_\infty^\pi) + \sum_{t=1}^T \left(\left\| S^{1/2} x_\infty^\pi \right\| - \left\| S^{1/2} x_t^\pi \right\| \right) \left(\left\| S^{1/2} x_t^\pi \right\| + \left\| S^{1/2} x_\infty^\pi \right\| \right) \ . \end{aligned}$$

Thus,

$$\begin{aligned} \gamma_T &\le \sum_{t=1}^T d_t (x_t^\pi - x_\infty^\pi) + \sum_{t=1}^T \left\| S^{1/2} (x_t^\pi - x_\infty^\pi) \right\| \left(\left\| S^{1/2} x_t^\pi \right\| + \left\| S^{1/2} x_\infty^\pi \right\| \right) \\ &\le \sum_{t=1}^T \left(\|d_t\| + \left\| S^{1/2} \right\| \left(\left\| S^{1/2} x_t^\pi \right\| + \left\| S^{1/2} x_\infty^\pi \right\| \right) \right) \|x_t^\pi - x_\infty^\pi\| \\ &\le Z_1' \sum_{t=1}^T \|x_t^\pi - x_\infty^\pi\| \ . \end{aligned}$$

We get the desired result by Lemma 6. □