# Structured Policy Iteration for Linear Quadratic Regulator (Supplementary Material)

**Youngsuk Park** [1] **Ryan A. Rossi** [2] **Zheng Wen** [3] **Gang Wu** [2] **Handong Zhao** [2]

## 1. Discussion on Non-convexity of Regularized LQR.

From Lemma 2 and appendix in (Fazel et al., 2018), unregularized objective $f(K)$ is known to be not convex, quasi-convex, nor star-convex, but to be gradient dominant, which gives the claim that all the stationary points are optimal as long as $\mathbf{E}[x_0x_0^T] \succ 0$ . However, in regularized LQR, this claim may not hold.

To see this claim that all stationary points may not be global optimal, let's define regularized LQR with $r(K) = \left\|K - K^{\rm lqr}\right\|$ where $K^{\rm lqr}$ is the solution of the Riccati algorithm. We know that $K^{\rm lqr}$ is the global optimal. Assume there is another distinct stationary point (like unregularized LQR) $K'$. Then, $f(K^{\rm lqr}) + \lambda r(K^{\rm lqr}) = f(K^{\rm lqr})$ is always less than $f(K') + \lambda \left\|K' - K^{\rm lqr}\right\|$. If not,i.e., $f(K^{\rm lqr}) \geq f(K') + \lambda \left\|K' - K^{\rm lqr}\right\|$, then $f(K') < f(K^{\rm lqr})$ holds and this is contradiction, showing all stationary points is not global optimal like unregularized LQR. Whether regularized LQR has only one stationary point or not is still an open question.

## 2. Additional Examples of Proximal Operators

Assume $\lambda, \lambda_1, \lambda_2 \in \mathbf{R}_+$ are positive numbers. We denote $(z)_i \in \mathbf{R}$ as its $i$th element or $(z)_j \in \mathbf{R}^{n_j}$ as its $j$th block under an explicit block structure, and $(z)_+ = \max(z, 0)$.

- **Group lasso.** For a group lasso penalty $r(x) = \sum_j^N \|x_j\|_2$ with $x_j \in \mathbf{R}^{n_j}$,

$$\left(\mathbf{prox}_{r,\lambda\eta}(x)\right)_j = \left(1 - \frac{\lambda\eta}{\|x_j\|_2}\right)_+ x_j$$

- **Elastic net** For a elastic net $r(x) = \lambda_1\|x\|_1 + \lambda_2\|x\|_2^2$,

$$\left(\mathbf{prox}_{g,U}(x)\right)_i = \text{sign}(x_i)\left(\frac{1}{\lambda_2\eta+1}|x_i| - \frac{\lambda_1\eta}{\lambda_2\eta+1})\right)_+$$

- **Nonnegative constraint.** Let $r(x) = \mathbf{1}(x \geq 0)$ be the nonnegative constraint. Then

$$\mathbf{prox}_{r,\lambda\eta}(x) = (x)_+$$

- **Simplex constraint** Let $r(x) = \mathbf{1}(x \geq 0, \mathbf{1}^T x = 1)$ be the simplex constraint. Then for $U = \text{Diag}(u)$,

$$\left(\mathbf{prox}_{r,\eta\lambda}(x)\right)_i = (x_i - \eta\lambda\nu)_+,$$

  Here, $\nu$ is the solution satisfying $\sum_i (x_i - \eta\lambda\nu)_+ = 1$, which can be found efficiently via bisection.

## 3. Proof for Convergence Analysis of S-PI

Let's define $\Sigma(K) = \mathbf{E}_{x_0\sim\mathcal{D}}[\sum_{t=0}^\infty x_t x_t^T]$. We often adopt and modify several techincal Lemmas like perturbation analysis from (Fazel et al., 2018).

**Lemma 1** (modification of Lemma 16 in (Fazel et al., 2018))**.** *Suppose $A + BK$ is stable and $K'$ is in the ball $\mathcal{B}(K;\rho_K)$, i.e.,*

$$K' \in \mathcal{B}(K;\rho_K) := \left\{K + \Delta K \in \mathbf{R}^{m\times n} \mid \|\Delta K\| \leq \rho_K\right\}$$

*where the radius $\rho_K$ is*

$$\rho_K = \frac{\sigma_{\min}(\Sigma_0)}{4\|\Sigma(K)\|\left(\|A + BK\| + 1\right)\|B\|}.$$

*Then*

$$\|\Sigma(K') - \Sigma(K)\| \leq \|\Sigma(K)\| \quad (1)$$

**Lemma 2** (Lemma 2 restated)**.** *For $K$ with stable $A + BK$, $f(K)$ is locally smooth with*

$$L_K = 4\|\Sigma(K)\|\|R + B^T P(K)B\| < \infty,$$

*within local ball around $K \in \mathcal{B}(K;\rho_K)$*

*And $f(K)$ is (globally) strongly convex with*

$$m = \sigma_{\min}(\Sigma_0)\sigma_{\min}(R) \geq 0.$$

*In addition, $A + BK'$ is stable for all $K' \in \mathcal{B}(K;\rho_K)$.*

[1]Stanford University [2]Adobe Research [3]Google DeepMind. Correspondence to: Youngsuk Park <youngsuk@stanford.edu>.

*Proceedings of the $37^{th}$ International Conference on Machine Learning*, Vienna, Austria, PMLR 119, 2020. 

*Proof.* First, we describe the terms with Talyor expansion

$$\begin{aligned} f(K') =& f(K) - 2\mathbf{Tr}(\Delta K^T \big( \big(R + B^T P(K) B\big) K \\ &+ B^T P(K) A \big) \Sigma(K)) \\ &+ \underbrace{\mathbf{Tr}\left(\Sigma(K')\Delta K^T (R + B^T P(K)B)\Delta K\right)}_{①} \end{aligned}$$

The second order term ① is (locally) upper bounded by

$$\begin{aligned} ① &\le \|\Sigma(K')\| \|R + B^T P(K) B\| \|\Delta K\|_F^2 \\ &\overset{(a)}{\le} 2\|\Sigma(K)\| \|R + B^T P(K) B\| \|\Delta K\|_F^2 \end{aligned}$$

where (a) holds due to Lemma 1

$$\|\Sigma(K')\| \le \|\Sigma(K)\| + \|\Sigma(K) - \Sigma(K')\| \le 2\|\Sigma(K)\|$$

within a ball $K' \in \mathcal{B}(K; \rho_K)$.

On the other hand,

$$\begin{aligned} ① &\ge \sigma_{\min}(\Sigma(K'))\sigma_{\min}(R + B^T P(K) B)) \|\Delta K\|_F^2 \\ &\overset{(b)}{\ge} \sigma_{\min}(\Sigma_0)\sigma_{\min}(R)\|\Delta K\|_F^2 \end{aligned}$$

where (b) hold due to $\Sigma_0 \preceq \Sigma(K')$ and $R \preceq R + B^T P(K) B$.

Therefore, the second order term is (locally) bounded by

$$\frac{m}{2}\|\Delta K\|_F^2 \le ① \le \frac{L_K}{2}\|\Delta K\|_F^2$$

where

$$\begin{aligned} m =& 2\sigma_{\min}(\Sigma_0)\sigma_{\min}(R) \ge 0, \\ L_K =& 2\|\Sigma(K)\| \|R + B^T P(K) B\| < \infty. \end{aligned}$$

∎

**Lemma 3** (Lemma 3 restated)**.** *Let* $K^+ = \mathbf{prox}_{r,\lambda\eta}(K - \eta\nabla f(K))$*. Then*

$$K^+ \in \mathcal{B}(K; \rho_K)$$

*holds for any* $0 < \eta < \eta_K^{\lambda,r}$ *where* $\eta_K^{\lambda,r}$ *is given as*

$$\eta_K^{\lambda,r} = \begin{cases} \frac{\rho_K}{\|\nabla f(K)\| + \lambda nm} & r(K) = \|K\|_1 \\ \frac{\rho_K}{\|\nabla f(K)\| + \lambda \min(n,m)} & r(K) = \|K\|_* \\ \frac{\rho_K}{2\|\nabla f(K)\| + 2\lambda \|K - K^{\text{ref}}\|} & r(K) = \|K - K^{\text{ref}}\|_F^2 \end{cases}.$$

*Proof.* For lasso, let $S_{\lambda\eta}$ be a soft-thresholding operator.

$$\begin{aligned} \|K^+ - K\| &\le \|(K - \eta\nabla f(K)) - K\| + \|K^+ - (K - \eta\nabla f(K))\| \\ &\le \eta\|\nabla f(K)\| + \|(S_{\eta\lambda}(K - \eta\nabla f(K)) - (K - \eta\nabla f(K))\| \\ &\le \eta\|\nabla f(K)\| + \eta\lambda nm \\ &\le \eta(\nabla f(K) + \lambda nm) \\ &\le \rho_K \end{aligned}$$

where the last inequality holds iff

$$\eta \le \frac{\rho_K}{\|\nabla f(K)\| + \lambda nm}.$$

For nuclear norm,

$$\begin{aligned} &\|K^+ - (K - \eta\nabla f(K))\| \\ &\quad \le \|(\text{TrunSVD}_{\eta\lambda}(K - \eta\nabla f(K)) - (K - \eta\nabla f(K))\| \\ &\quad \le \|U(\mathbf{diag}(S_{\lambda\eta}[\sigma_1, \ldots, \sigma_{\min(n,m)}]) \\ &\qquad - \mathbf{diag}(\sigma_1, \ldots, \sigma_{\min(n,m)}))V^T\| \\ &\quad \le \eta\lambda\min(n,m) \end{aligned}$$

Therefore,

$$\begin{aligned} \|K^+ - K\| &\le \|(K - \eta\nabla f(K)) - K\| + \|K^+ - (K - \eta\nabla f(K))\| \\ &\le \eta(\nabla f(K) + \lambda\min(n,m)) \\ &\le \rho_K \end{aligned}$$

where the last inequality holds iff

$$\eta \le \frac{\rho_K}{\|\nabla f(K)\| + \lambda\min(n,m)}.$$

For the third regulazer,

$$\begin{aligned} &\|K^+ - (K - \eta\nabla f(K))\| \\ &\quad \overset{(a)}{\le} \left\| \frac{2\eta\lambda K^{\text{ref}} + K - \eta\nabla f(K)}{2\eta\lambda + 1} - (K - \eta\nabla f(K)) \right\|_F \\ &\quad = \left\| \frac{2\eta\lambda}{2\eta\lambda + 1}(K^{\text{ref}} - K) - \frac{2\eta\lambda}{2\eta\lambda + 1}\eta\nabla f(K) \right\|_F \\ &\quad \overset{(b)}{\le} 2\eta\lambda \left\|K^{\text{ref}} - K\right\|_F + \eta\left\|\nabla f(K)\right\|_F, \end{aligned}$$

where (a) holds from the closed solution of proximal operator in Lemma 1(in main paper) and (b) holds due to $\frac{2\eta\lambda}{2\eta\lambda+1} \le 2\eta\lambda$ and $\frac{2\eta\lambda}{2\eta\lambda+1} \le 1$. Therefore, using this inequality gives

$$\begin{aligned} &\|K^+ - K\| \\ &\quad \le \|(K - \eta\nabla f(K)) - K\| + \|K^+ - (K - \eta\nabla f(K))\| \\ &\quad \le 2\eta(\|\nabla f(K)\| + \lambda\left\|K^{\text{ref}} - K\right\|_F) \\ &\quad \le \rho_K, \end{aligned}$$

where the last inequality holds iff

$$\eta \le \frac{\rho_K}{2(\|\nabla f(K)\| + \lambda\left\|K^{\text{ref}} - K\right\|_F)}.$$

∎

**Lemma 4.** *For any* $0 < \eta \le \min(\frac{1}{L_K}, \eta_K^{\lambda,r})$*, let* $K^+ = \mathbf{prox}_{r,\lambda\eta}(K - \eta\nabla f(K)) = K - \eta G_\eta(K)$ *where* $G_\eta(K) =$

$\frac{1}{\eta}(K - \mathbf{prox}_{r,\lambda\eta}(K - \eta\nabla f(K)))$. *Then, for any* $Z \in \mathbf{R}^{m\times n}$,

$$F(K^+) \le F(Z) + G_\eta(K)^T(K-Z) - \frac{m}{2}\|K-Z\|_F^2 - \frac{\eta}{2}\|G_\eta(K)\|_F^2 \quad (2)$$

*holds.*

*Proof.* For $K^+ = K - \eta G_\eta(K)$ with any $0 < \eta$ and any $Z \in \mathbf{R}^{m\times n}$, we have

$$\begin{aligned} r(K-\eta G_\eta(K)) &\overset{(a)}{\le} r(Z) - \mathbf{Tr}\left(\partial r(K-\eta G_\eta(K))^T(Z-K+\eta G_\eta(K))\right) \\ &\overset{(b)}{=} r(Z) - \mathbf{Tr}\left((G_\eta(K)-\nabla f(K))^T(Z-K+\eta G_\eta(K))\right) \\ &= r(Z) + \mathbf{Tr}\left(G_\eta(K)^T(K-Z)\right) - \eta\|G_\eta(K)\|_F^2 \\ &\quad + \mathbf{Tr}\left(\nabla f(K)^T(Z-K+\eta G_\eta(K))\right) \end{aligned}$$

where (a) holds due to convexity of $g$, (b) holds due to the property of subgradient on proximal operator. Next, for any $0 < \eta \le \eta_K^{\lambda,r}$, $K^+ \in \mathcal{B}(K;\rho_K)$ holds from Lemma 3 and thus $f(K)$ is locally smooth. Therefore

$$\begin{aligned} f(K-\eta G_\eta(K)) &\overset{(c)}{\le} f(K) - \mathbf{Tr}\left(\nabla f(K)^T\eta G_\eta(K)\right) + \frac{L_K\eta^2}{2}\|G_\eta(K)\|_F^2 \\ &\overset{(d)}{\le} f(K) - \mathbf{Tr}\left(\nabla f(K)^T\eta G_\eta(K)\right) + \frac{\eta}{2}\|G_\eta(K)\|_F^2 \\ &\overset{(e)}{\le} f(Z) - \mathbf{Tr}\left(\nabla f(K)^T(Z-K)\right) - \frac{m}{2}\|Z-K\|_F^2 \\ &\quad - \mathbf{Tr}\left(\nabla f(K)^T\eta G_\eta(K)\right) + \frac{\eta}{2}\|G_\eta(K)\|_F^2 \\ &= f(Z) - \mathbf{Tr}\left(\nabla f(K)^T(Z-K+\eta G_\eta(K))\right) \\ &\quad - \frac{m}{2}\|Z-K\|_F^2 + \frac{\eta}{2}\|G_\eta(K)\|_F^2 \end{aligned} \quad (3)$$

where (c) holds due to $L$-smoothness for $K^+ \in \mathcal{B}(K;\rho_K)$, (d) holds by $\eta \le \frac{1}{L_K}$, (e) holds due to $m$-strongly convexity at $K$. And note that Substituting $Z = K$ in (3) is equivalent to linesearch criterion in Eq. (8) (in main paper), which will be satisfied for small enough stepsize $\eta$ after linesearch iterations.

Adding two inequalities above gives

$$\begin{aligned} F(K^+) =& f(K-\eta G_\eta(K)) + r(K-\eta G_\eta(K)) \\ &\le F(Z) + \mathbf{Tr}\left(G_\eta(K)^T(K-Z)\right) \\ &\quad - \frac{m}{2}\|Z-K\|_F^2 - \frac{\eta}{2}\|G_\eta(K)\|_F^2 \end{aligned} \quad (4)$$

■

**Proposition 1** (Proposition 1 restated)**.** *Assume $A+BK$ is stable. For any stepsize $0 < \eta \le \min(\frac{1}{L_K}, \eta_K^r)$ and next iterate $K^+ = \mathbf{prox}_{r(\cdot),\eta\lambda}(K - \eta\nabla f(K))$,*

$$\rho(A+BK^+) < 1 \quad (5)$$

$$F(K^+) \le F(K) - \frac{1}{2\eta}\|K-K^+\|_F^2 \quad (6)$$

*holds.*

*Proof.* From Lemma 3, (5) comes immediately from Lemma 8 in (Fazel et al., 2018). And applying $Z = K$ in Lemma 4 gives (6). ■

**Lemma 5** (Lemma 4 restated)**.** *Assume that $\{K^i\}_{i=0,\dots}$ is a stabilizing sequence and associated $\{f(K^i)\}_{i=0,\dots}$ and $\{\|K^i - K^\star\|_F\}_{i=0,\dots}$ are decreasing sequences. Then, Lemma 3 holds for*

$$\eta^{\lambda,r} = \begin{cases} \frac{\rho^L}{\rho^f+\lambda nm} & r(K) = \|K\|_1 \\ \frac{\rho^L}{\rho^f+\lambda\min(n,m)} & r(K) = \|K\|_* \\ \frac{\rho^L}{2\rho^f+4\lambda\Delta} & r(K) = \|K-K^{\text{ref}}\|_F^2 \end{cases}. \quad (7)$$

*where*

$$\begin{aligned} \rho^f &= 2\frac{F(K^0)}{\sigma_{\min}(Q)}\Bigg(\|B^T\|\frac{F(K^0)}{\sigma_{\min}(\Sigma_0)}\|A\| + \\ &\quad \left(\|R\| + \|B^T\|\frac{F(K^0)}{\sigma(\Sigma_0)}\|B\|\right)(\Delta + \|K^\star\|)\Bigg), \\ \rho^L &= \frac{\sigma_{\min}(\Sigma_0)^2}{8F(K^0)\|B\|}. \end{aligned}$$

*Proof.* For the proof, we derive the global bound on $\|\nabla f(K^i)\| \le \rho^f$ and $\rho_K \ge \rho^L$, then plug these into Lemma 3 to complete our claim. First, we utilize the derivation of the upperbound on $\|P(K^i)\|$ and $\|\Sigma(K^i)\|$ in (Fazel et al., 2018) under the assumption of decreasing sequence as follows,

$$\|P(K^i)\| \le \frac{F(K^0)}{\sigma_{\min}(\Sigma_0)}, \qquad \|\Sigma(K^i)\| \le \frac{F(K^0)}{\sigma_{\min}(Q)}.$$

From this, we have

$$\rho_K = \frac{\sigma_{\min}(\Sigma_0)}{4\|\Sigma(K)\|\left(\|A+BK\|+1\right)\|B\|} \ge \frac{\sigma_{\min}^2(\Sigma_0)}{8F(K^0)\|B\|}$$

holds where we used the fact that $\|A+BK^i\| < 1$ and $\|\Sigma(K^i)\| \le \frac{F(K^0)}{\sigma_{\min}(Q)}$. Now we complete the proof by also providing $\rho^f$ .Since $\|\nabla f(K^i)\|$ is bounded as

$$\begin{aligned}\|\nabla f(K^i)\| &\le \left\|2\left(\left(R+B^TPB\right)K+B^TPA\right)\Sigma\right\| \\ &\le 2\Big(\left(\|R\|+\left\|B^T\right\|\|P\|\|B\|\right)\|K\| \\ &\quad +\left\|B^T\right\|\|P\|\|A\|\Big)\|\Sigma\| \\ &\le 2\left(\left(\|R\|+\left\|B^T\right\|\frac{F(K^0)}{\sigma_{\min}(\Sigma_0)}\|B\|\right)\|K\|\right. \\ &\quad \left.+\left\|B^T\right\|\frac{F(K^0)}{\sigma_{\min}(\Sigma_0)}\|A\|\right)\frac{F(K^0)}{\sigma(Q)} \\ &\le 2\left(\left(\|R\|+\left\|B^T\right\|\frac{F(K^0)}{\sigma_{\min}(\Sigma_0)}\|B\|\right)(\|K^\star\|+\Delta)\right. \\ &\quad \left.+\left\|B^T\right\|\frac{F(K^0)}{\sigma(\Sigma_0)}\|A\|\right)\frac{F(K^0)}{\sigma_{\min}(Q)}\end{aligned}$$

where the last inequality holds due to $\|K\| \le (\|K^\star\|+\|K-K^\star\|)\le\|K^\star\|+\Delta$.

■

**Proposition 2.** *Let $\eta_i$ be the stepsize from backtracking linesearch at $i$-th iteration. After $N$ iterations, it converges to a stationary point $K^\star$ satisfying*

$$\min_{i=1,\dots,N}\|G_{\eta_i}(K^i)\|_F^2 \le \frac{2(F(K^0)-F^\star)}{\eta_{\min}N}$$

*where $G_{\eta_i}(K^i) \in \nabla f(K^i)+\partial r(K^i-\eta_i\nabla f(K^i))$, $G_{\eta_i}(K^i)=0$ iff $0\in\partial F(K^i)$. Moreover,*

$$\eta_{\min}=h_\eta\Bigg(\sigma_{\min}(\Sigma_0),\sigma_{\min}(Q),\frac{1}{\lambda}, \frac{1}{\|A\|},\frac{1}{\|B\|},\frac{1}{\|R\|},\frac{1}{\Delta},\frac{1}{F(K^0)}\Bigg)$$

*where $h_\eta$ is a function non-decreasing on each argument.*

*Proof.* From Lemma 1,

$$F(K^{i+1})\le F(K^i)-\frac{\eta_i}{2}\|G_{\eta_i}(K^i)\|_F^2\ \forall i$$

Reordering terms and averaging over iterations $i=1\dots N$ give

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^N\|G_{\eta_i}(K^i)\|_2^2 &\le \frac{2}{N}\sum_{i=1}^N\frac{1}{\eta_i}(F(K^i)-F(K^{i+1})) \\ &\le \frac{2(F(K^0)-F(K^\star))}{\min\limits_{i=1,\dots,N}\eta_i N}.\end{aligned}$$

And LHS is lower bounded by

$$\frac{1}{N}\sum_{i=1}^N\|G_{\eta_i}(K^i)\|_F^2 \ge \min_{i=1,\dots,N}\|G_{\eta_i}(K^i)\|_F^2,$$

giving the desirable result. Moreover, it converges to the stationary point since $\lim_{i\to\infty}G_{\eta_i}(K^i)=0$.

Now the remaining part is to bound the stepsize. Note that the stepsize $\eta_i$ after linesearch satisfies

$$\eta_i\ge\frac{1}{\beta}\min\left(\frac{1}{L_{K_i}},\eta^r_{K_i}\right).$$

First we bound $\frac{1}{L_{K_i}}$ as follows,

$$\begin{aligned}\frac{1}{L_{K^i}} &= \frac{1}{4\|\Sigma(K)\|\|R+B^TP(K)B\|} \\ &\ge \frac{1}{4\|\Sigma(K)\|(\|R\|+\|B^T\|\|P(K)\|\|B\|)} \\ &\ge \frac{\sigma_{\min}(\Sigma_0)\sigma_{\min}(Q)}{4F(K^0)(\sigma_{\min}(Q)\|R\|+F(K^0)\|B^T\|\|B\|)}.\end{aligned}$$

Next, about the bound on $\eta^{\lambda,r}_{K_i}$, we already have $\eta^{\lambda,r}_{K_i}\ge\eta^{\lambda,r}$ from Lemma 5.

Note that both of bounds are proportional to $\sigma_{\min}(\Sigma_0)$ and $\sigma_{\min}(Q)$, and inverse-proportional to $\|A\|,\|B\|,\|R\|,\Delta$ and $F(K^0)$.

Therefore

$$\min_{i=1,\dots,}\eta_i\ge\eta_{\min}=h_\eta\Bigg(\sigma_{\min}(\Sigma_0),\sigma_{\min}(Q),\frac{1}{\lambda}, \frac{1}{\|A\|},\frac{1}{\|B\|},\frac{1}{\|R\|},\frac{1}{\Delta},\frac{1}{F(K^0)}\Bigg)$$

for some $h_\eta$ that is non-decreasing on each argument.

■

**Theorem 1** (Theorem 1 restated)**.** *$K^i$ from Algorithm 1 converges to the stationary point $K^\star$. Moreover, it converges linearly, i.e., after $N$ iterations,*

$$\left\|K^N-K^\star\right\|_F^2\le\left(1-\frac{1}{\kappa}\right)^N\left\|K^0-K^\star\right\|_F^2.$$

*Here, $\kappa=1/\left(\eta_{\min}\sigma_{\min}(\Sigma_0)\sigma_{\min}(R)\right))>1$ where*

$$\eta_{\min}=h_\eta\Bigg(\sigma_{\min}(\Sigma_0),\sigma_{\min}(Q),\frac{1}{\lambda}, \frac{1}{\|A\|},\frac{1}{\|B\|},\frac{1}{\|R\|},\frac{1}{\Delta},\frac{1}{F(K^0)}\Bigg),\tag{8}$$

*for some non-decreasing function $h_\eta$ on each argument.*

*Proof.* Substituting $Z = K^\star$ in Lemma 4 gives,

$$\begin{aligned}
&F(K^+) - F^\star \\
&\leq \mathbf{Tr}\left(G_\eta(K)^T(K - K^\star)\right) - \frac{m}{2}\|K - K^\star\|_F^2 - \frac{\eta}{2}\|G_\eta(K)\|_F^2 \\
&= \frac{1}{2\eta}\left(\|K - K^\star\|_F^2 - \|K - K^\star - \eta G_\eta(K)\|_F^2\right) \\
&\quad - \frac{m}{2}\|K - K^\star\|_F^2 \\
&= \frac{1}{2\eta}\left(\|K - K^\star\|_F^2 - \left\|K^+ - K^\star\right\|_F^2\right) - \frac{m}{2}\|K - K^\star\|_F^2 .
\end{aligned}$$

Reordering terms gives

$$\begin{aligned}
\left\|K^+ - K^\star\right\|_F^2 &\leq \|K - K^\star\|_F^2 \\
&\quad - \left(2\eta(F(K^+) - K^\star) + m\eta\|K - K^\star\|_F^2\right) \\
&\leq (1 - m\eta)\|K - K^\star\|_F^2
\end{aligned}$$

where the last inequality holds due to $F(K^+) - F^\star \geq 0$.

Therefore, after $N$ iterations,

$$\begin{aligned}
\left\|K^N - K^\star\right\|_F^2 &\leq (1 - m\eta_N)\cdots(1 - m\eta_1)\left\|K^0 - K^\star\right\|_F^2 \\
&\leq (1 - m\eta_{\min})^N\left\|K^0 - K^\star\right\|^2
\end{aligned}$$

where $\eta_{\min}$ is the same one in Proposition 2 ∎

**Corollary 1.** *Let $K^\star$ be the stationary point from Algorithm 1. Then, after $N$ iterations*

$$N \geq 2\kappa \log\left(\frac{\left\|K^0 - K^\star\right\|_F}{\epsilon}\right),$$

$$\left\|K^N - K^\star\right\|_F \leq \epsilon$$

*holds where $\kappa = 1/\left(\eta_{\min}\sigma_{\min}(\Sigma_0)\sigma_{\min}(R)\right) > 1$ and $\eta_{\min}$ in Eq. (8).*

*Proof.* This is immediate from Theorem 1, using the inequality $(1 - 1/\kappa)^N \leq e^{-N/\kappa}$ and by taking the logarithm. ∎

## 4. Proof for Convergence Analysis of Model-free S-PI

**Lemma 6** (Lemma 30 in (Fazel et al., 2018))**.** *There exists polynomials , $h_{N_{\text{traj}}}$, $h_H$, $h_r$ such that, when $r < 1/h_r(1/\epsilon)$, $N_{\text{traj}} \geq h_{N_{\text{traj}}}(n, 1/\epsilon, \frac{L^2}{\sigma_{\min}(\Sigma_0)})$ and $H \geq h_H(n, 1/\epsilon)$, the gradient estimate $\widehat{\nabla f(K)}$ given in Eq. (13) of Algorithm 3 satisfies*

$$\left\|\widehat{\nabla f(K)} - \nabla f(K)\right\|_F \leq \epsilon$$

*with high probability (at least $1 - (\epsilon/n)^n$.*

**Theorem 2** (Theorem 2 restated)**.** *Suppose $f(K^0)$ is finite, $\Sigma_0 \succ 0$, and that $x_0 \sim \mathcal{D}$ has norm bounded by $L$ almost surly. Suppose the parameters in Algorithm 3 are chosen from*

$$(N_{\text{traj}}, H, 1/r) = h\left(n, \frac{1}{(\sigma_{\min}(\Sigma_0)\sigma_{\min}(R))}, \frac{L^2}{\sigma_{\min}(\Sigma_0)}\right).$$

*for some polynomials $h$. Then, with the same stepsize in Eq. (8), Algorithm 3 converges to its stationary point $K^\star$ with high probability. In particular, there exist iteration $N$ at most $4\kappa\log\left(\frac{\|K^0 - K^\star\|_F}{\epsilon}\right)$ such that $\left\|K^N - K^\star\right\| \leq \epsilon$ with at least $1 - o(\epsilon^{n-1})$ probability. Moreover, it converges linearly,*

$$\left\|K^i - K^\star\right\|^2 \leq \left(1 - \frac{1}{2\kappa}\right)^i \left\|K^0 - K^\star\right\|^2,$$

*for the iteration $i = 1, \ldots, N$, where $\kappa = \eta\sigma_{\min}(\Sigma_0)\sigma_{\min}(R) > 1$.*

*Proof.* Let $\epsilon$ be the error bound we want to obtain, i.e., $\left\|K^N - K^\star\right\| \leq \epsilon$ where $K^N$ is the policy from Algorithm 3 after $N$ iterations.

For a notational simplicity, we denote $K \leftarrow K^i$ and see the contraction of the proximal operator at $i$th iteration. First we use Lemma 6 to claim that, with high probability, $\left\|\widehat{\nabla f(K)} - \nabla f(K)\right\|_F \leq \alpha\epsilon$ for long enough numbers of trajectory $N_{\text{traj}}$ and horizon $H$ where $\alpha$ is specified later.

Second, we bound the error after one iteration of approximated proximal gradient step at the policy $K$, i.e., $\|K' - K^+\|_F$. Here let $K' = \mathbf{prox}(K - \eta\widehat{\nabla f(K)}$ be the next iterate using approximate gradient $\widehat{\nabla f(K)})$ and $K^+ = \mathbf{prox}_{\lambda r}(K - \eta\nabla f(K)$ be the one using the exact gradient $\nabla f(K)$.

$$\begin{aligned}
\left\|K' - K^+\right\|_F &= \left\|\mathbf{prox}(K - \eta\widehat{\nabla f(K)}) - \mathbf{prox}(K - \eta\nabla f(K)\right\|_F \\
&\leq \left\|(K - \eta\widehat{\nabla f(K)}) - (K - \eta\widehat{\nabla f(K)})\right\|_F \\
&= \eta\left\|\widehat{\nabla f(K)} - \nabla f(K)\right\|_F \\
&\leq \eta\alpha\epsilon
\end{aligned}$$

where we use the fact that proximal operator is non-expansive and $\left\|\widehat{\nabla f(K)} - \nabla f(K)\right\|_F \leq \alpha\epsilon$ holds for proper parameter choices (the claim in the previous paragraph).

Third, we find the contractive upperbound after one iteration using approximated proximal gradient.

$$\begin{aligned}\|K' - K^\star\|_F &\le \left\|K' - K^+\right\|_F + \left\|K^+ - K^\star\right\|_F \\ &\le \left\|K' - K^+\right\|_F + \sqrt{(1-1/\kappa)}\,\|K - K^\star\|_F \\ &\le \eta\alpha\epsilon + \sqrt{(1-\kappa)}\,\|K - K^\star\|_F \,.\end{aligned}$$

Let's assume $\|K - K^\star\|_F \ge \epsilon$ under current policy. Then, taking square on both sides gives

$$\begin{aligned}\|K' - K^\star\|_F^2 &\le \eta^2\alpha^2\epsilon^2 + 2\eta\alpha\epsilon\sqrt{(1-1/\kappa)}\,\|K - K^\star\|_F \\ &\quad + (1-1/\kappa)\,\|K - K^\star\|_F^2 \\ &\le \left(1 - 1/\kappa + 2\alpha\eta + \alpha^2\eta^2\right)\|K - K^\star\|_F^2 \,,\end{aligned}$$

where we used $\sqrt{(1-1/\kappa)} \le 1$, $1 \le \kappa$, and the assumption. Choosing $\alpha = \frac{1}{5\eta\kappa} = \frac{1}{5\sigma_{\min}(\Sigma_0)\sigma_{\min}(R))}$ results in

$$\|K' - K^\star\|^2 \le (1 - 1/(2\kappa))\,\|K - K^\star\|^2 \,,$$

with high probability $1 - (\alpha\epsilon/n)^n$. This says the approximate proximal gradient is contractive, decreasing in error after one iteration. Keep applying this inequality, we get

$$\left\|K^i - K^\star\right\|^2 \le (1 - 1/(2\kappa))^i \left\|K^0 - K^\star\right\|^2 .$$

as long as $\epsilon \le \left\|K^{i-1} - K^\star\right\|_F$.

This says that there must exist the iteration $N > 0$ s.t.

$$\left\|K^N - K^\star\right\|_F \le \epsilon \le \left\|K^{N-1} - K^\star\right\|_F \tag{9}$$

Now we claim this $N$ is at most $4\kappa \log\left(\frac{\left\|K^0 - K^\star\right\|_F}{\epsilon}\right)$. To prove this claim, suppose it is not, i.e., $N \ge 4\kappa \log\left(\frac{\left\|K^0 - K^\star\right\|_F}{\epsilon}\right) + 1$. Then, for $N > 1$,

$$\begin{aligned}\left\|K^{N-1} - K^\star\right\|_F^2 &\le (1 - 1/(2\kappa))^{N-1}\left\|K^0 - K^\star\right\|^2 \\ &< e^{-\frac{N-1}{2\kappa}}\left\|K^0 - K^\star\right\|^2 \\ &\le \left(\frac{\left\|K^0 - K^\star\right\|_F}{\epsilon}\right)^{-2}\left\|K^0 - K^\star\right\|^2 \\ &= \epsilon^2,\end{aligned}$$

which is a contradiction to (9).

Finally, we show the probability that this event occurs. Note that all randomness occur when estimating $N$ gradient within $\alpha\epsilon$ error. From union bound, it occurs at least $1 - N(\alpha\epsilon/n)^n$. And this is bounded below by

$$\begin{aligned}1 - N(\alpha\epsilon/n)^n &\ge 1 - \left(4\kappa\log\left(\frac{\left\|K^0 - K^\star\right\|_F}{\epsilon}\right) + 1\right)(\alpha\epsilon/n)^n \\ &\ge 1 - o\left(\epsilon^n \log\left(\frac{\left\|K^0 - K^\star\right\|_F}{\epsilon}\right)\right) \\ &= 1 - o(\epsilon^{n-1}).\end{aligned}$$

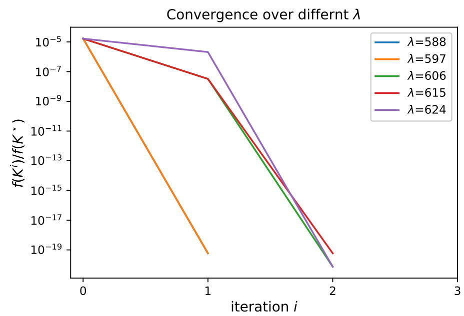


*Figure 1.* Convergence behavior of the Structured Policy Iteration (S-PI) under linesearch for Laplacian system of $(n, m) = (3, 3)$ over various $\lambda$s.

■

## 5. Additional Experiments on Stepsize Sensitivity

In this section, we scrutinize the convergence behaviors of S-PI under some fixed stepsize. For a very small Laplacian system $(n, m) = (3, 3)$ with Lasso penalty $\lambda = 3000$, we run S-PI over a wide range of stepsizes. For stepsize larger than $3.7e - 4$, S-PI diverges and thus is ran under stepsizes smaller than $3.7e - 4$. Let $K^{min}$ be the policy where the objective value attains its minimum among overall iterates and $K^\star$ be the policy from S-PI with linesearch (non-fixed stepsize). Here the cardinality of the optimal policy is 3. For a fixed stepsizes in $[3.7e - 5, 3.7e - 6]$, S-PI converges to the optimal. In Figure 2, the objective value monotonically decreases and the policy converges to optimal one based on errors and cardinality. However, for smaller stepsize like $[3.7e - 7, 3.7e - 8, 3.7e - 9]$, Figure 3 shows that S-PI still converges but does not show monotonic behaviors nor converges to the optimal policy. These figures demonstrate the sensitivity of a stepsize when S-PI is used under a fixed stepsize, rather than linesearch. Like in Figure 3, the algorithm can be unstable under fixed stepsize because the next iterate $K^+$ may not satisfy the stability condition $\rho(A + BK^+) < 1$ and or are not guaranteed for a monotonic decrease. Moreover, this instability may lead to another stationary point even when the iterate falls in some stable policy region after certain iterations. This not only demonstrates the importance of lineasearch due to its sensitivity on the stepsize, but may provide the evidence for why other policy gradient type of methods for LQR did not perform well in practice.

## References


Fazel, M., Ge, R., Kakade, S. M., and Mesbahi, M. Global convergence of policy gradient methods for the linear

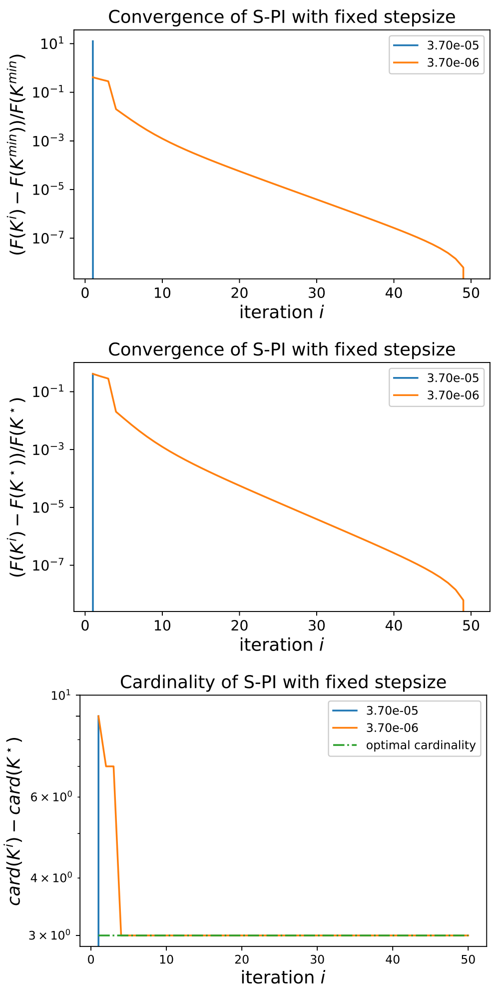


*Figure 2.* Convergence behavior of the Structured Policy Iteration (S-PI) over fixed stepsizes $[3.7e-5, 3.7e-6]$ for Laplacian system of $(n, m) = (3, 3)$ with $\lambda = 3000$.

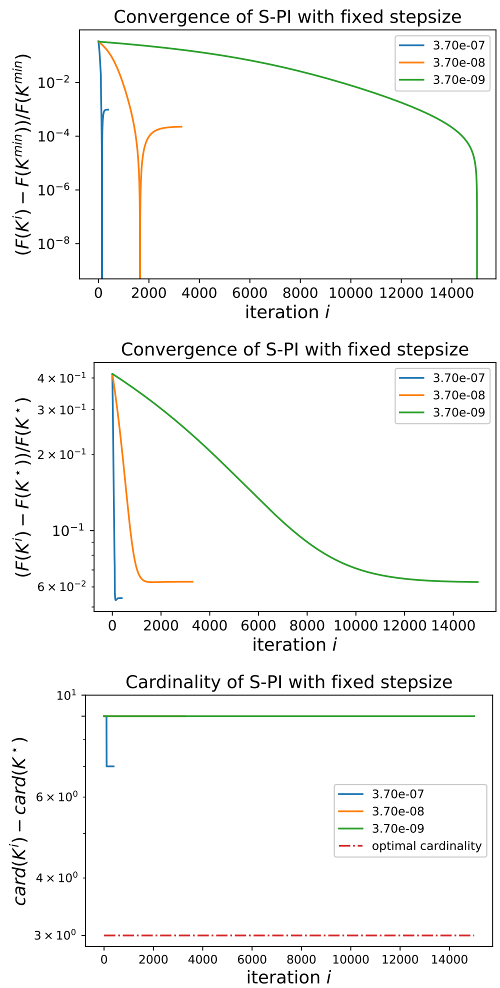


*Figure 3.* Convergence behavior of the Structured Policy Iteration (S-PI) over fixed stepsizes $[3.7e-7, 3.7e-8, 3.7e-9]$ for Laplacian system of $(n, m) = (3, 3)$ with $\lambda = 3000$.

quadratic regulator. *arXiv preprint arXiv:1801.05039*, 2018.