A. Planning with a model

This section briefly reviews some parameterizations and solution methods for the classic LQR and related problems from control theory.

Finite horizon LQR. First, consider the finite horizon case. The basic approach is to view it as a dynamic program with the value function $x_t^T P_t x_t$, where

$$P_{t-1} = Q + A^T P_t A - A^T P_t B(R + B^T P_t B)^{-1} B^T P_t A,$$

which in turn gives the optimal control

$$u_t = -K_t x_t = -(R + B^T P_{t+1} B)^{-1} B^T P_{t+1} A x_t,$$

(recursions run backward in time).

Another approach is to view the LQR problem as a linearly-constrained Quadratic Program in all x_t and u_t (where the constraints are given by the dynamics, and the problem size equals the horizon). The QP is clearly a convex problem, but this observation is not useful by itself as the problem size grows with the horizon, and naive use of quadratic programming scales badly. However, the special structure due to the linearity of the dynamics allows for simplifications and a control-theoretic interpretation as follows: the Lagrange multipliers in the QP can be interpreted as "co-state" variables, and they follow a recursion that runs backwards in time known as the "adjoint system" dynamics. Using Lagrange duality, one can show that this approach is equivalent to solving the Riccati recursion mentioned above.

Popular use of the LQR in control practice is often in the receding horizon LQR, (Camacho & Bordons, 2004; Rawlings & Mayne, 2009): at time t, an input sequence is found that minimizes the T-step ahead LQR cost starting at the current time, then only the first input in the sequence is used. The resulting static feedback gain converges to the infinite horizon optimal solution as horizon T becomes longer.

Infinite horizon LQR. Here, the constrained optimization view (QP) is not informative as the problem is infinite dimensional; however the dynamic programming viewpoint readily extends. Suppose the system $A,\ B$ is controllable (which guarantees the optimal cost is finite). It turns out that the value function and the optimal controller are static (i.e., do not depend on t) and can be found by solving the Algebraic Riccati Equation (ARE) given in (1). The optimal K can then be found from equation (2).

The main computational step is solving the ARE, which is extensively studied (e.g. (Lancaster & Rodman, 1995)). One approach due to (Kleinman, 1968) (for continuous time) and (Hewer, 1971) (for discrete time) is to simply run the recursion $P_{k+1} = Q + A^T P_k A - A^T P_k B(R + B^T P_k B)^{-1} B^T P_k A$ where $P_1 = Q$, which converges to the unique positive semidefinite solution of the ARE (since the fixed-point iteration is contractive). Other approaches are direct and based on linear algebra, which carry out an eigenvalue decomposition on a certain block matrix (called the Hamiltonian matrix) followed by a matrix inversion (Lancaster & Rodman, 1995).

Direct computation of the control input has also been considered in the optimal control literature, e.g., gradient updates in function spaces (Polak, 1973). For the linear quadratic setup, direct iterative computation of the feedback gain has been examined in (Mårtensson & Rantzer, 2009), and explored further in (Mårtensson, 2012) with a view towards distributed implementations. There methods are presented as local search heuristics without provable guarantees of reaching the optimal policy.

SDP formulation. The LQR problem can also be expressed as a semidefinite program (SDP) with variable P, as given in (Balakrishnan & Vandenberghe, 2003) (section 5, equation (34), this is for a continuous-time system but there are similar discrete-time versions). This SDP can be derived by relaxing the equality in the Riccati equation to an inequality, then using the Schur complement lemma to rewrite the resulting Riccati inequality as linear matrix inequality. The objective in the case of LQR is the trace of the positive definite matrix variable, and the optimization problem (for the continuous time system) is given as

$$\begin{array}{ll} \text{maximize} & x_0^T P x_0 \\ \text{subject to} & \begin{bmatrix} A^T P + PA + Q & PB \\ B^T P & I \end{bmatrix} \geq 0, \quad P \geq 0, \end{array} \tag{10}$$

where the optimization variable is P This SDP and its dual, and system-theoretic interpretations of its optimality conditions, have been explored in (Balakrishnan & Vandenberghe, 2003). Note that while the optimal solution P^* of this SDP is the unique positive semidefinite solution to the Riccati equation, which in turn gives the optimal policy K^*, other feasible P (not equal to P^*) do not necessarily correspond to a feasible, stabilizing policy K. This means that the feasible set of this SDP is not a convex characterization of all P that correspond to stabilizing K. Thus it also implies that if one uses any optimization algorithm that maintains iterates in the feasible set (e.g., interior point methods), no useful policy can be extracted from the iterates before convergence to P^*. For this reason, this convex formulation is not helpful for parametrizing the space of policies K in a manner that supports the use of local search methods (those that directly lower the cost function of interest as a function of policy K), which is the focus of this work.

B. Non-convexity of the set of stabilizing State Feedback Gains

In this section we prove Lemma 2. Let $\mathcal{K}(A, B)$ denote the set of state feedback gains K such that $A - BK$ is stable, i.e., its eigenvalues are inside the unit circle in the complex plane. This set is generally nonconvex. A concise counterexample to convexity is provided here. Let A and B be 3×3 identity matrices and

$$K_1 = \begin{bmatrix} 1 & 0 & -10 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad K_2 = \begin{bmatrix} 1 & -10 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}.$$

Then the spectra of $A - BK_1$ and $A - BK_2$ are both concentrated at the origin, yet two of the eigenvalues of $A - B\widehat{K}$ with $\widehat{K} = (K_1 + K_2)/2$, are outside of the unit circle in the complex plane.

C. Analysis: the exact case

This section provides the analysis of the convergence rates of the (exact) gradient based methods. First, some helpful lemmas for the analysis are provided.

Throughout, it is convenient to use the following definition:

$$E_K := (R + B^\top P_K B)K - B^\top P_K A\,.$$

The policy gradient can then be written as:

$$\nabla C(K) = 2\left((R + B^\top P_K B)K - B^\top P_K A\right)\Sigma_K = 2E_K\Sigma_K$$

C.1. Helper lemmas

Define the value $V_K(x)$, the state-action value $Q_K(x, u)$, and the advantage $A_K(x, u)$. $V_K(x, t)$ is the cost of the policy starting with $x_0 = x$ and proceeding with K onwards:

$$\begin{aligned} V_K(x) &:= \sum_{t=0}^{\infty} \left(x_t^\top Q x_t + u_t^\top R u_t\right) \\ &= x^\top P_K x\,. \end{aligned}$$

$Q_K(x, u)$ is the cost of the policy starting with $x_0 = x$, taking action $u_0 = u$ and then proceeding with K onwards:

$$Q_K(x, u) := x^\top Q x + u^\top R u + V_K(Ax + Bu)$$

The advantage $A_K(x, u)$ is:

$$A_K(x, u) = Q_K(x, u) - V_K(x)\,.$$

The advantage can be viewed as the change in cost starting at state x and taking a one step deviation from the policy K.

The next lemma is identical to that in (Kakade & Langford, 2002; Kakade, 2003) for Markov decision processes.

Lemma 10. *(Cost difference lemma) Suppose K and K' have finite costs. Let $\{x'_t\}$ and $\{u'_t\}$ be state and action sequences generated by K', i.e. starting with $x'_0 = x$ and using $u'_t = -K'x'_t$. It holds that:*

$$V_{K'}(x) - V_K(x) = \sum_t A_K(x'_t, u'_t)\,.$$

Also, for any x, the advantage is:

$$A_K(x, K'x) = 2x^\top (K' - K)^\top E_K x + x^\top (K' - K)^\top (R + B^\top P_K B)(K' - K)x\,. \tag{11}$$

Proof. Let c'_t be the cost sequence generated by K'. Telescoping the sum appropriately:

$$\begin{aligned} V_{K'}(x) - V_K(x) &= \sum_{t=0} c'_t - V_K(x) \\ &= \sum_{t=0} (c'_t + V_K(x'_t) - V_K(x'_t)) - V_K(x) \\ &= \sum_{t=0} (c'_t + V_K(x'_{t+1}) - V_K(x'_t)) \\ &= \sum_{t=0} A_K(x'_t, u'_t) \end{aligned}$$

which completes the first claim (the third equality uses the fact that $x = x_0 = x'_0$).

For the second claim, observe that:

$$V_K(x) = x^\top \left(Q + K^\top R K\right) x + x^\top (A - BK)^\top P_K (A - BK)x$$

And, for $u = K'x$,

$$\begin{aligned} A_K(x, u) &= Q_K(x, u) - V_K(x) \\ &= x^\top \left(Q + (K')^\top R K'\right) x + x^\top (A - BK')^\top P_K (A - BK')x - V_K(x) \\ &= x^\top \left(Q + (K' - K + K)^\top R (K' - K + K)\right) x + \\ &\quad\; x^\top (A - BK - B(K' - K))^\top P_K (A - BK - B(K' - K))x - V_K(x) \\ &= 2x^\top (K' - K)^\top \left((R + B^\top P_K B)K - B^\top P_K A\right) x + \\ &\quad\; x^\top (K' - K)^\top (R + B^\top P_K B)(K' - K))x\,, \end{aligned}$$

which completes the proof. ☐

This lemma is helpful in proving that $C(K)$ is gradient dominated.

Lemma 11. *(Gradient domination, Lemma 3 and Corollary 5 restated) Let K^* be an optimal policy. Suppose K has finite cost and $\mu > 0$. It holds that:*

$$\begin{aligned} C(K) - C(K^*) &\leq \|\Sigma_{K^*}\| \mathrm{Tr}(E_K^\top (R + B^\top P_K B)^{-1} E_K) \\ &\leq \frac{\|\Sigma_{K^*}\|}{\sigma_{min}(R)} \mathrm{Tr}(E_K^\top E_K) \\ &\leq \frac{\|\Sigma_{K^*}\|}{\sigma_{min}(\Sigma_K)^2 \sigma_{min}(R)} \mathrm{Tr}(\nabla C(K)^\top \nabla C(K)) \\ &\leq \frac{\|\Sigma_{K^*}\|}{\mu^2 \sigma_{min}(R)} \mathrm{Tr}(\nabla C(K)^\top \nabla C(K)) \end{aligned}$$

For a lower bound, it holds that:

$$C(K) - C(K^*) \geq \frac{\mu}{\|R + B^\top P_K B\|} \mathrm{Tr}(E_K^\top E_K)$$

Proof. From Equation 11 and by completing the square,

$$\begin{aligned} & Q_K(x, K'x) - V_K(x) \\ = \;& 2\mathrm{Tr}(xx^\top (K' - K)^\top E_K) + \mathrm{Tr}(xx^\top (K' - K)^\top (R + B^\top P_K B)(K' - K)) \\ = \;& \mathrm{Tr}(xx^\top \left(K' - K + (R + B^\top P_K B)^{-1} E_K\right)^\top (R + B^\top P_K B) \left(K' - K + (R + B^\top P_K B)^{-1} E_K\right)) \\ & -\mathrm{Tr}(xx^\top E_K^\top (R + B^\top P_K B)^{-1} E_K) \\ \geq \;& -\mathrm{Tr}(xx^\top E_K^\top (R + B^\top P_K B)^{-1} E_K) \end{aligned} \tag{12}$$

with equality when $K' = K - (R + B^\top P_K B)^{-1} E_K$.

Let x_t^* and u_t^* be the sequence generated under K_*. Using this and Lemma 10,

$$\begin{aligned}
C(K) - C(K^*) &= -\mathbb{E} \sum_t A_K(x_t^*, u_t^*) \\
&\leq \mathbb{E} \sum_t \mathrm{Tr}(x_t^* (x_t^*)^\top E_K^\top (R + B^\top P_K B)^{-1} E_K) \\
&= \mathrm{Tr}(\Sigma_{K^*} E_K^\top (R + B^\top P_K B)^{-1} E_K) \\
&\leq \|\Sigma_{K^*}\| \mathrm{Tr}(E_K^\top (R + B^\top P_K B)^{-1} E_K) \\
&\leq \|\Sigma_{K^*}\| \|(R + B^\top P_K B)^{-1}\| \, \mathrm{Tr}(E_K^\top E_K) \\
&\leq \frac{\|\Sigma_{K^*}\|}{\sigma_{\min}(R)} \mathrm{Tr}(E_K^\top E_K) \\
&= \frac{\|\Sigma_{K^*}\|}{\sigma_{\min}(R)} \mathrm{Tr}(\Sigma_K^{-1} \nabla C(K)^\top \nabla C(K) \Sigma_K^{-1}) \\
&\leq \frac{\|\Sigma_{K^*}\|}{\sigma_{min}(\Sigma_K)^2 \sigma_{\min}(R)} \mathrm{Tr}(\nabla C(K)^\top \nabla C(K)) \\
&\leq \frac{\|\Sigma_{K^*}\|}{\mu^2 \sigma_{\min}(R)} \mathrm{Tr}(\nabla C(K)^\top \nabla C(K))
\end{aligned}$$

which completes the proof of the upper bound. Here the last step is because $\Sigma_K \succeq \mathbb{E}[x_0 x_0^\top]$.

For the lower bound, consider $K' = K - (R + B^\top P_K B)^{-1} E_K$ where equality holds in Equation 12. Let x_t' and u_t' be the sequence generated under K'. Using that $C(K^*) \leq C(K')$,

$$\begin{aligned}
C(K) - C(K^*) &\geq C(K) - C(K') \\
&= -\mathbb{E} \sum_t A_K(x_t', u_t') \\
&= \mathbb{E} \sum_t \mathrm{Tr}(x_t' (x_t')^\top E_K^\top (R + B^\top P_K B)^{-1} E_K) \\
&\geq \mathrm{Tr}(\Sigma_{K'} E_K^\top (R + B^\top P_K B)^{-1} E_K) \\
&\geq \frac{\mu}{\|R + B^\top P_K B\|} \mathrm{Tr}(E_K^\top E_K)
\end{aligned}$$

which completes the proof. □

Recall that a function f is said to be smooth (or C^1-smooth) if for some finite β, it satisfies:

$$|f(x) - f(y) - \nabla f(y)^\top (x - y)| \leq \frac{\beta}{2} \|x - y\|^2 . \tag{13}$$

for all x, y (equivalently, it is smooth if the gradients of f are continuous).

Lemma 12. *("Almost" smoothness, Lemma 6 restated) $C(K)$ satisfies:*

$$C(K') - C(K) = -2\mathrm{Tr}(\Sigma_{K'} (K - K')^\top E_K) + \mathrm{Tr}(\Sigma_{K'} (K - K')^\top (R + B^\top P_K B)(K - K'))$$

To see why this is related to smoothness (e.g. compare to Equation 13), suppose K' is sufficiently close to K so that:

$$\Sigma_{K'} \approx \Sigma_K + O(\|K - K'\|) \tag{14}$$

and the leading order term $2\mathrm{Tr}(\Sigma_{K'} (K' - K)^\top E_K)$ would then behave as $\mathrm{Tr}((K' - K)^\top \nabla C(K))$. The challenge in the proof (for gradient descent) is quantifying the lower order terms in this argument.

Proof. The claim immediately results from Lemma 10, by using Equation 11 and taking an expectation. □

The next lemma spectral norm bounds on P_K and Σ_K are helpful:

Lemma 13. *It holds that:*

$$\|P_K\| \leq \frac{C(K)}{\mu}, \qquad \|\Sigma_K\| \leq \frac{C(K)}{\sigma_{\min}(Q)}$$

Proof. For the first claim, $C(K)$ is lower bounded as:

$$C(K) = \mathbb{E}_{x_0 \sim \mathcal{D}} x_0^\top P_K x_0 \geq \|P_K\| \sigma_{\min}(\mathbb{E} x_0 x_0^\top)$$

Alternatively, $C(K)$ can be lower bounded as:

$$C(K) = \mathrm{Tr}(\Sigma_K(Q + K^\top R K)) \geq \mathrm{Tr}(\Sigma_K)\sigma_{\min}(Q) \geq \|\Sigma_K\|\sigma_{\min}(Q)\,,$$

which proves the second claim. □

C.2. Gauss-Newton Analysis

The next lemma bounds the one step progress of Gauss-Newton.

Lemma 14. *(Lemma 8 restated) Suppose that:*

$$K' = K - \eta(R + B^\top P_K B)^{-1} \nabla C(K) \Sigma_K^{-1}\,,.$$

If $\eta \leq 1$, then

$$C(K') - C(K^*) \leq \left(1 - \frac{\eta\mu}{\|\Sigma_{K^*}\|}\right)(C(K) - C(K^*))$$

Proof. Observe $K' = K - \eta(R + B^\top P_K B)^{-1} E_K$. Using Lemma 12 and the condition on η,

$$\begin{aligned}
C(K') - C(K) &= -2\eta \mathrm{Tr}(\Sigma_{K'} E_K^\top (R + B^\top P_K B)^{-1} E_K) + \eta^2 \mathrm{Tr}(\Sigma_{K'} E_K^\top (R + B^\top P_K B)^{-1} E_K) \\
&\leq -\eta \mathrm{Tr}(\Sigma_{K'} E_K^\top (R + B^\top P_K B)^{-1} E_K) \\
&\leq -\eta \sigma_{\min}(\Sigma_{K'}) \mathrm{Tr}(E_K^\top (R + B^\top P_K B)^{-1} E_K) \\
&\leq -\eta \mu \mathrm{Tr}(E_K^\top (R + B^\top P_K B)^{-1} E_K) \\
&\leq -\eta \frac{\mu}{\|\Sigma_{K^*}\|}(C(K) - C(K^*))\,,
\end{aligned}$$

where the last step uses Lemma 11. □

With this lemma, the proof of the convergence rate of the Gauss Newton algorithm is immediate.

Proof. (of Theorem 7, Gauss-Newton case) The theorem is due to that $\eta = 1$ leads to a contraction of $1 - \frac{\eta\mu}{\|\Sigma_{K^*}\|}$ at every step. □

C.3. Natural Policy Gradient Descent Analysis

The next lemma bounds the one step progress of the natural policy gradient.

Lemma 15. *Suppose:*

$$K' = K - \eta \nabla C(K) \Sigma_K^{-1}$$

and that $\eta \leq \frac{1}{\|R + B^\top P_K B\|}$. It holds that:

$$C(K') - C(K^*) \leq \left(1 - \eta \sigma_{min}(R) \frac{\mu}{\|\Sigma_{K^*}\|}\right)(C(K) - C(K^*))$$

Proof. Since $K' = K - \eta E_K$, Lemma 12 implies:

$$C(K') - C(K) = -2\eta \mathrm{Tr}(\Sigma_{K'} E_K^\top E_K) + \eta^2 \mathrm{Tr}(\Sigma_{K'} E_K^\top (R + B^\top P_K B) E_K)$$

The last term can be bounded as:

$$\begin{aligned}\mathrm{Tr}(\Sigma_{K'} E_K^\top (R + B^\top P_K B) E_K) &= \mathrm{Tr}((R + B^\top P_K B) E_K \Sigma_{K'} E_K^\top) \\ &\leq \|R + B^\top P_K B\| \mathrm{Tr}(E_K \Sigma_{K'} E_K^\top) \\ &= \|R + B^\top P_K B\| \mathrm{Tr}(\Sigma_{K'} E_K^\top E_K)\,.\end{aligned}$$

Continuing and using the condition on η,

$$\begin{aligned}C(K') - C(K) &\leq -2\eta \mathrm{Tr}(\Sigma_{K'} E_K^\top E_K) + \eta^2 \|R + B^\top P_K B\| \mathrm{Tr}(\Sigma_{K'} E_K^\top E_K) \\ &\leq -\eta \mathrm{Tr}(\Sigma_{K'} E_K^\top E_K) \\ &\leq -\eta \sigma_{\min}(\Sigma_{K'}) \mathrm{Tr}(E_K^\top E_K) \\ &\leq -\eta \mu \mathrm{Tr}(E_K^\top E_K) \\ &\leq -\eta \frac{\mu \sigma_{\min}(R)}{\|\Sigma_{K^*}\|} (C(K) - C(K^*))\end{aligned}$$

using Lemma 11. □

With this lemma, the proof of the natural policy gradient convergence rate can be completed.

Proof. (of Theorem 7, natural policy gradient case) Using Lemma 13,

$$\frac{1}{\|R + B^\top P_K B\|} \geq \frac{1}{\|R\| + \|B\|^2 \|P_K\|} \geq \frac{1}{\|R\| + \frac{\|B\|^2 C(K)}{\mu}}$$

The proof is completed by induction: $C(K_1) \leq C(K_0)$, since Lemma 15 can be applied. The proof proceeds by arguing that Lemma 15 can be applied at every step. If it were the case that $C(K_t) \leq C(K_0)$, then

$$\eta \leq \frac{1}{\|R\| + \frac{\|B\|^2 C(K_0)}{\mu}} \leq \frac{1}{\|R\| + \frac{\|B\|^2 C(K_t)}{\mu}} \leq \frac{1}{\|R + B^\top P_{K_t} B\|}$$

and by Lemma 15:

$$C(K_{t+1}) - C(K^*) \leq \left(1 - \frac{\mu}{\|\Sigma_{K^*}\|} \frac{\sigma_{\min}(R)}{\|R\| + \frac{\|B\|^2 C(K_0)}{\mu}}\right) (C(K_t) - C(K^*))$$

which completes the proof. □

C.4. Gradient Descent Analysis

As informally argued by Equation 14, the proof seeks to quantify how $\Sigma_{K'}$ changes with η. Then the proof bounds the one step progress of gradient descent.

Σ_K PERTURBATION ANALYSIS

This subsections aims to prove the following:

Lemma 16. *(Σ_K perturbation) Suppose K' is such that:*

$$\|K' - K\| \leq \frac{\sigma_{\min}(Q)\mu}{4C(K)\|B\| (\|A - BK\| + 1)}$$

It holds that:

$$\|\Sigma_{K'} - \Sigma_K\| \leq 4 \left(\frac{C(K)}{\sigma_{\min}(Q)}\right)^2 \frac{\|B\| (\|A - BK\| + 1)}{\mu} \|K - K'\|$$

The proof proceeds by starting with a few technical lemmas. First, define a linear operator on symmetric matrices, $\mathcal{T}_K(\cdot)$, which can be viewed as a matrix on $\binom{d+1}{2}$ dimensions. Define this operator on a symmetric matrix X as follows:

$$\mathcal{T}_K(X) := \sum_{t=0}^{\infty}(A-BK)^t X[(A-BK)^\top]^t$$

Also define the induced norm of $\mathcal{T}$ as follows:

$$\|\mathcal{T}_K\| = \sup_X \frac{\|\mathcal{T}_K(X)\|}{\|X\|} \tag{15}$$

where the supremum is over all symmetric matrices X (whose spectral norm is non-zero).

Also, define

$$\Sigma_0 = \mathbb{E} x_0 x_0^\top$$

.

Lemma 17. *($\mathcal{T}_K$ norm bound) It holds that*

$$\|\mathcal{T}_K\| \leq \frac{C(K)}{\mu\, \sigma_{\min}(Q)}$$

Proof. For a unit norm vector $v \in \mathbb{R}^d$ and unit spectral norm matrix X,

$$\begin{aligned}
v^\top(\mathcal{T}_K(X))v &= \sum_{t=0}^{\infty} v^\top (A-BK)^t X[(A-BK)^\top]^t v \\
&= \sum_{t=0}^{\infty} \mathrm{Tr}([(A-BK)^\top]^t v v^\top (A-BK)^t X) \\
&= \sum_{t=0}^{\infty} \mathrm{Tr}([\Sigma_0^{1/2}(A-BK)^\top]^t v v^\top (A-BK)^t \Sigma_0^{1/2} \Sigma_0^{-1/2} X \Sigma_0^{-1/2}) \\
&\leq \sum_{t=0}^{\infty} \mathrm{Tr}([\Sigma_0^{1/2}(A-BK)^\top]^t v v^\top (A-BK)^t \Sigma_0^{1/2}) \|\Sigma_0^{-1/2} X \Sigma_0^{-1/2}\| \\
&= \|\Sigma_0^{-1/2} X \Sigma_0^{-1/2}\| \left(v^\top \mathcal{T}_K(\Sigma_0) v\right) \\
&\leq \frac{1}{\sigma_{\min}(\mathbb{E} x_0 x_0^\top)} \|\mathcal{T}_K(\Sigma_0)\| \\
&= \frac{1}{\mu}\|\Sigma_K\|
\end{aligned}$$

using that $\mathcal{T}_K(\Sigma_0) = \Sigma_K$. The proof is completed using the upper bound on $\|\Sigma_K\|$ in Lemma 13. □

Also, with respect to K, define another linear operator on symmetric matrices:

$$\mathcal{F}_K(X) = (A-BK)X(A-BK)^\top .$$

Let I to denote the identity operator on the same space. Define the induced norm $\|\cdot\|$ of these operators as in Equation 15. Note these operators are related to the operator $\mathcal{T}_K$ as follows:

Lemma 18. *When $(A-BK)$ has spectral radius smaller than 1,*

$$\mathcal{T}_K = (\mathrm{I} - \mathcal{F}_K)^{-1}.$$

Proof. When $(A-BK)$ has spectral radius smaller than 1, $\mathcal{T}_K$ is well defined and is the solution of $\mathcal{T}_K = \mathrm{I} + \mathcal{T}_K \circ \mathcal{F}_K$. Therefore $\mathcal{T}_K \circ (\mathrm{I} - \mathcal{F}_K) = \mathrm{I}$ and $\mathcal{T}_K = (\mathrm{I} - \mathcal{F}_K)^{-1}$. □

Since,

$$\Sigma_K = \mathcal{T}_K(\Sigma_0) = (\mathrm{I} - \mathcal{F}_K)^{-1}(\Sigma_0)\,.$$

The proof of Lemma 16 seeks to bound:

$$\|\Sigma_K - \Sigma_{K'}\| = \|(\mathcal{T}_K - \mathcal{T}_{K'})(\Sigma_0)\| = \|((\mathrm{I} - \mathcal{F}_K)^{-1} - (\mathrm{I} - \mathcal{F}_{K'})^{-1})(\Sigma_0)\|\,.$$

The following two perturbation bounds are helpful in this.

Lemma 19. *It holds that:*

$$\|\mathcal{F}_K - \mathcal{F}_{K'}\| \le 2\|A - BK\|\|B\|\|K - K'\| + \|B\|^2\|K - K'\|^2.$$

Proof. Let $\Delta = K - K'$. For every matrix X,

$$(\mathcal{F}_K - \mathcal{F}_{K'})(X) = (A - BK)X(B\Delta)^\top + (B\Delta)X(A - BK)^\top - (B\Delta)X(B\Delta)^\top.$$

The operator norm of $\mathcal{F}_K - \mathcal{F}_{K'}$ is the maximum possible ratio in spectral norm of $(\mathcal{F}_K - \mathcal{F}_{K'})(X)$ and X. Then the claim follows because $\|AX\| \le \|A\|\|X\|$. □

Lemma 20. *If*

$$\|\mathcal{T}_K\|\|\mathcal{F}_K - \mathcal{F}_{K'}\| \le 1/2\,,$$

then

$$\begin{aligned}\|\left(\mathcal{T}_K - \mathcal{T}_{K'}\right)(\Sigma)\| &\le 2\|\mathcal{T}_K\|\|\mathcal{F}_K - \mathcal{F}_{K'}\|\|\mathcal{T}_K(\Sigma)\|.\\ &\le 2\|\mathcal{T}_K\|^2\|\mathcal{F}_K - \mathcal{F}_{K'}\|\|\Sigma\|.\end{aligned}$$

Proof. Define $\mathcal{A} = \mathrm{I} - \mathcal{F}_K$, and $\mathcal{B} = \mathcal{F}_{K'} - \mathcal{F}_K$. In this case $\mathcal{A}^{-1} = \mathcal{T}_K$ and $(\mathcal{A} - \mathcal{B})^{-1} = \mathcal{T}_{K'}$. Hence, the condition $\|\mathcal{T}_K\|\|\mathcal{F}_K - \mathcal{F}_{K'}\| \le 1/2$ translates to the condition $\|\mathcal{A}^{-1}\|\|\mathcal{B}\| \le 1/2$.

Observe:

$$(\mathcal{A}^{-1} - (\mathcal{A} - \mathcal{B})^{-1})(\Sigma) = (\mathrm{I} - (\mathrm{I} - \mathcal{A}^{-1}\circ\mathcal{B})^{-1})(\mathcal{A}^{-1}(\Sigma)) = (\mathrm{I} - (\mathrm{I} - \mathcal{A}^{-1}\circ\mathcal{B})^{-1})(\mathcal{T}_K(\Sigma))\,.$$

Since $(\mathrm{I} - \mathcal{A}^{-1}\circ\mathcal{B})^{-1} = \mathrm{I} + \mathcal{A}^{-1}\circ\mathcal{B}\circ(\mathrm{I} - \mathcal{A}^{-1}\circ\mathcal{B})^{-1}$,

$$\|(\mathrm{I} - \mathcal{A}^{-1}\circ\mathcal{B})^{-1}\| \le 1 + \|\mathcal{A}^{-1}\circ\mathcal{B}\|\|(\mathrm{I} - \mathcal{A}^{-1}\circ\mathcal{B})^{-1}\| \le 1 + 1/2\|(\mathrm{I} - \mathcal{A}^{-1}\circ\mathcal{B})^{-1}\|$$

which implies $\|(\mathrm{I} - \mathcal{A}^{-1}\circ\mathcal{B})^{-1}\| \le 2$. Hence,

$$\|\mathrm{I} - (\mathrm{I} - \mathcal{A}^{-1}\circ\mathcal{B})^{-1}\| = \|\mathcal{A}^{-1}\circ\mathcal{B}\circ(\mathrm{I} - \mathcal{A}^{-1}\circ\mathcal{B})^{-1}\| \le \|\mathcal{A}^{-1}\|\|\mathcal{B}\|\|(\mathrm{I} - \mathcal{A}^{-1}\circ\mathcal{B})^{-1}\| = 2\|\mathcal{A}^{-1}\|\|\mathcal{B}\|.$$

and so

$$\|\mathrm{I} - (\mathrm{I} - \mathcal{A}^{-1}\circ\mathcal{B})^{-1}\| \le 2\|\mathcal{A}^{-1}\|\|\mathcal{B}\| = 2\|\mathcal{T}_K\|\|\mathcal{F}_K - \mathcal{F}_{K'}\|\,.$$

Combining these two,

$$\|\left(\mathcal{T}_K - \mathcal{T}_{K'}\right)(\Sigma)\| \le \|(\mathrm{I} - (\mathrm{I} - \mathcal{A}^{-1}\circ\mathcal{B})^{-1})\|\|\mathcal{T}_K(\Sigma)\| \le 2\|\mathcal{T}_K\|\|\mathcal{F}_K - \mathcal{F}_{K'}\|\|\mathcal{T}_K(\Sigma)\|.$$

This proves the main inequality. The last step of the inequality is just applying definition of the norm of $\mathcal{T}_K$: $\|\mathcal{T}_K(\Sigma)\| \le \|\mathcal{T}_K\|\|\Sigma\|$. □

With these Lemmas, the proof is completed as follows:

Proof. (of Lemma 16) First, the proof shows $\|\mathcal{T}_K\|\|\mathcal{F}_K - \mathcal{F}_{K'}\| \leq 1/2$, which is the desired condition in Lemma 20. First, observe that under the assumed condition on $\|K - K'\|$, implies that

$$\|B\|\|K' - K\| \leq \frac{\sigma_{\min}(Q)\mu}{4C(K)\left(\|A - BK\| + 1\right)} \leq \frac{1}{4}\frac{\sigma_{\min}(Q)\mu}{C(K)} \leq \frac{1}{4}$$

using that $\frac{\sigma_{\min}(Q)\mu}{C(K)} \leq 1$ due to Lemma 13. Using Lemma 19,

$$\begin{aligned} \|\mathcal{F}_K - \mathcal{F}_{K'}\| &\leq \left(2\|A - BK\|\|B\|\|K - K'\| + \|B\|^2\|K - K'\|^2\right) \\ &\leq 2\|B\|\left(\|A - BK\| + 1\right)\|K - K'\| \end{aligned} \tag{16}$$

Using this and Lemma 17,

$$\|\mathcal{T}_K\|\|\mathcal{F}_K - \mathcal{F}_{K'}\| \leq \frac{C(K)}{\sigma_{\min}(Q)\mu} 2\|B\|\left(\|A - BK\| + 1\right)\|K - K'\| \leq \frac{1}{2}$$

where the last step uses the condition on $\|K - K'\|$.

Thus,

$$\begin{aligned} \|\Sigma_{K'} - \Sigma_K\| &\leq 2\|\mathcal{T}_K\|\|\mathcal{F}_K - \mathcal{F}_{K'}\|\|\mathcal{T}_K(\Sigma_0)\| \\ &\leq 2\frac{C(K)}{\sigma_{\min}(Q)\mu}\left(2\|B\|\left(\|A - BK\| + 1\right)\|K - K'\|\right)\frac{C(K)}{\sigma_{\min}(Q)} \end{aligned}$$

using Lemmas 13 and 19. □

GRADIENT DESCENT PROGRESS

Equipped with these lemmas, the one step progress of gradient descent can be bounded.

Lemma 21. *Suppose that*

$$K' = K - \eta\nabla C(K)\,,$$

where

$$\eta \leq \frac{1}{16}\min\left\{\left(\frac{\sigma_{\min}(Q)\mu}{C(K)}\right)^2 \frac{1}{\|B\|\|\nabla C(K)\|(1 + \|A - BK\|)}\,,\ \frac{\sigma_{\min}(Q)}{2C(K)\|R + B^\top P_K B\|}\right\}. \tag{17}$$

It holds that:

$$C(K') - C(K^*) \leq \left(1 - \eta\sigma_{min}(R)\frac{\mu^2}{\|\Sigma_{K^*}\|}\right)\left(C(K) - C(K^*)\right)$$

Proof. By Lemma 12,

$$\begin{aligned} & C(K') - C(K) \\ = \;& -2\eta\mathrm{Tr}(\Sigma_{K'}\Sigma_K E_K^\top E_K) + \eta^2\mathrm{Tr}(\Sigma_K\Sigma_{K'}\Sigma_K E_K^\top(R + B^\top P_K B)E_K) \\ \leq \;& -2\eta\mathrm{Tr}(\Sigma_K E_K^\top E_K\Sigma_K) + 2\eta\|\Sigma_{K'} - \Sigma_K\|\mathrm{Tr}(\Sigma_K E_K^\top E_K) \\ & +\eta^2\|\Sigma_{K'}\|\|R + B^\top P_K B\|\mathrm{Tr}(\Sigma_K\Sigma_K E_K^\top E_K) \\ \leq \;& -2\eta\mathrm{Tr}(\Sigma_K E_K^\top E_K\Sigma_K) + 2\eta\frac{\|\Sigma_{K'} - \Sigma_K\|}{\sigma_{\min}(\Sigma_K)}\mathrm{Tr}(\Sigma_K E_K^\top E_K\Sigma_K) \\ & +\eta^2\|\Sigma_{K'}\|\|R + B^\top P_K B\|\mathrm{Tr}(\Sigma_K E_K^\top E_K\Sigma_K) \\ = \;& -2\eta\left(1 - \frac{\|\Sigma_{K'} - \Sigma_K\|}{\sigma_{\min}(\Sigma_K)} - \frac{\eta}{2}\|\Sigma_{K'}\|\|R + B^\top P_K B\|\right)\mathrm{Tr}(\nabla C(K)^\top\nabla C(K)) \\ \leq \;& -2\eta\frac{\mu^2\sigma_{\min}(R)}{\|\Sigma_{K^*}\|}\left(1 - \frac{\|\Sigma_{K'} - \Sigma_K\|}{\mu} - \frac{\eta}{2}\|\Sigma_{K'}\|\|R + B^\top P_K B\|\right)\left(C(K) - C(K^*)\right) \end{aligned}$$

where the last step uses Lemma 11.

By Lemma 16,

$$\frac{\|\Sigma_{K'} - \Sigma_K\|}{\mu} \leq 4\eta \left(\frac{C(K)}{\sigma_{\min}(Q)\mu} \right)^2 \|B\| \left(\|A - BK\| + 1 \right) \|\nabla C(K)\| \leq 1/4$$

using the assumed condition on η.

Using this last claim and Lemma 13,

$$\|\Sigma_{K'}\| \leq \|\Sigma_{K'} - \Sigma_K\| + \|\Sigma_K\| \leq \frac{\mu}{4} + \frac{C(K)}{\sigma_{\min}(Q)} \leq \frac{\|\Sigma_{K'}\|}{4} + \frac{C(K)}{\sigma_{\min}(Q)}$$

and so $\|\Sigma_{K'}\| \leq \frac{4C(K)}{3\sigma_{\min}(Q)}$. Hence,

$$1 - \frac{\|\Sigma_{K'} - \Sigma_K\|}{\mu} - \frac{\eta}{2}\|\Sigma_{K'}\|\|R + B^\top P_K B\| \geq 1 - 1/4 - \frac{\eta}{2}\frac{4C(K)}{3\sigma_{\min}(Q)}\|R + B^\top P_K B\| \geq 1/2$$

using the condition on η. □

In order to prove a gradient descent convergence rate, the following bounds are helpful:

Lemma 22. *It holds that*

$$\|\nabla C(K)\| \leq \frac{C(K)}{\sigma_{\min}(Q)} \sqrt{\frac{\|R + B^\top P_K B\|(C(K) - C(K^*))}{\mu}}$$

and that:

$$\|K\| \leq \frac{1}{\sigma_{\min}(R)} \left(\sqrt{\frac{\|R + B^\top P_K B\|(C(K) - C(K^*))}{\mu}} + \|B^\top P_K A\| \right)$$

Proof. Using Lemma 13,

$$\|\nabla C(K)\|^2 \leq \mathrm{Tr}(\Sigma_K E_K^\top E_K \Sigma_K) \leq \|\Sigma_K\|^2 \mathrm{Tr}(E_K^\top E_K) \leq \left(\frac{C(K)}{\sigma_{\min}(Q)} \right)^2 \mathrm{Tr}(E_K^\top E_K)$$

By Lemma 11,

$$\mathrm{Tr}(E_K^\top E_K) \leq \frac{\|R + B^\top P_K B\|(C(K) - C(K^*))}{\mu}$$

which proves the first claim.

Again using Lemma 11,

$$\begin{aligned}
\|K\| &\leq \|(R + B^\top P_K B)^{-1}\| \|(R + B^\top P_K B)K\| \\
&\leq \frac{1}{\sigma_{\min}(R)} \|(R + B^\top P_K B)K\| \\
&\leq \frac{1}{\sigma_{\min}(R)} \left(\|(R + B^\top P_K B)K - B^\top P_K A\| + \|B^\top P_K A\| \right) \\
&= \frac{\|E_K\|}{\sigma_{\min}(R)} + \frac{\|B^\top P_K A\|}{\sigma_{\min}(R)} \\
&\leq \frac{\sqrt{\mathrm{Tr}(E_K^\top E_K)}}{\sigma_{\min}(R)} + \frac{\|B^\top P_K A\|}{\sigma_{\min}(R)} \\
&= \frac{\sqrt{(C(K) - C(K^*))\|R + B^\top P_K B\|}}{\sqrt{\mu}\,\sigma_{\min}(R)} + \frac{\|B^\top P_K A\|}{\sigma_{\min}(R)}
\end{aligned}$$

which proves the second claim. □

With these lemmas, the proof of the gradient descent convergence rate follows:

Proof. (of Theorem 7, gradient descent case) First, the following argues that progress is made at $t = 1$. Based on Lemma 13 and Lemma 22, by choosing η to be an appropriate polynomial in $\frac{1}{C(K_0)}, \frac{1}{\|A\|}, \frac{1}{\|B\|}, \frac{1}{\|R\|}, \sigma_{\min}(R), \sigma_{\min}(Q)$ and μ, the stepsize condition in Equation 17 is satisfied. Hence, by Lemma 21,

$$C(K_1) - C(K^*) \leq \left(1 - \eta\sigma_{\min}(R)\frac{\mu^2}{\|\Sigma_{K^*}\|}\right)(C(K_0) - C(K^*))$$

which implies that the cost decreases at $t = 1$. Proceeding inductively, now suppose that $C(K_t) \leq C(K_0)$, then the stepsize condition in Equation 17 is still satisfied (due to the use of $C(K_0)$ in bounding the quantities in Lemma 22). Thus, Lemma 21 can again be applied for the update at time $t + 1$ to obtain:

$$C(K_{t+1}) - C(K^*) \leq \left(1 - \eta\sigma_{\min}(R)\frac{\mu^2}{\|\Sigma_{K^*}\|}\right)(C(K_t) - C(K^*))\,.$$

Provided

$$T \geq \frac{\|\Sigma_{K^*}\|}{\eta\mu^2\sigma_{\min}(R)} \log\frac{C(K_0) - C(K^*)}{\varepsilon}\,,$$

then $C(K_T) - C(K^*) \leq \varepsilon$, and the result follows. ☐

D. Analysis: the Model-free case

This section shows how techniques from zeroth order optimization allow the algorithm to run in the model-free setting with only black-box access to a simulator. The dependencies on various parameters are not optimized, and the notation h is used to represent different polynomial factors in the relevant factors ($\frac{C(K_0)}{\mu\sigma_{min}(Q)}, \|A\|, \|B\|, \|R\|, 1/\sigma_{min}(R)$). When the polynomial also depend on dimension d or accuracy $1/\epsilon$, this is specified as parameters ($h(d, 1/\epsilon)$).

The section starts by showing how the infinite horizon can be approximated with a finite horizon.

D.1. Approximating $C(K)$ and Σ_K with finite horizon

This section shows that as long as there is an upper bound on $C(K)$, it is possible to approximate both $C(K)$ and $\Sigma(K)$ with any desired accuracy.

Lemma 23. *For any K with finite $C(K)$, let $\Sigma_K^{(\ell)} = \mathbb{E}[\sum_{i=0}^{\ell-1} x_i x_i^\top]$ and $C^{(\ell)}(K) = \mathbb{E}[\sum_{i=0}^{\ell-1} x_i^\top Q x_i + u_i^\top R u_i] = \langle\Sigma_K^{(t)}, Q + K^\top RK\rangle$. If*

$$\ell \geq \frac{d \cdot C^2(K)}{\epsilon\mu\sigma_{min}^2(Q)},$$

then $\|\Sigma_K^{(\ell)} - \Sigma_K\| \leq \epsilon$. Also, if

$$\ell \geq \frac{d \cdot C^2(K)(\|Q\| + \|R\|\|K\|^2)}{\epsilon\mu\sigma_{min}^2(Q)}$$

then $C(K) \geq C^{(\ell)}(K) \geq C(K) - \epsilon$.

Proof. First, the bound on Σ_K is proved. Define the operators $\mathcal{T}_K$ and $\mathcal{F}_K$ as in Section C.4, observe $\Sigma_K = \mathcal{T}_K(\Sigma_0)$ and $\Sigma_K^{(\ell)} = \Sigma_K - (\mathcal{F}_K)^\ell(\Sigma_K)$.

If $X \succeq Y$, then $\mathcal{F}_K(X) \succeq \mathcal{F}_K(Y)$, this follows immediately from the form of $\mathcal{F}_K(X) = (A + BK)X(A + BK)^\top$. If X is PSD then $WXW^\top$ is also PSD for any W.

Now, since

$$\sum_{i=0}^{\ell-1} \mathrm{tr}(\mathcal{F}^\ell(\Sigma_0)) = \mathrm{tr}(\sum_{i=0}^{\ell-1} \mathcal{F}^\ell(\Sigma_0)) \leq \mathrm{tr}(\sum_{i=0}^{\infty} \mathcal{F}^\ell(\Sigma_0)) = \mathrm{tr}(\Sigma_K) \leq \frac{d \cdot C(K)}{\sigma_{min}(Q)}.$$

(Here the last step is by Lemma 13), and all traces are nonnegative, then there must exists $j < \ell$ such that $\mathrm{tr}(\mathcal{F}_K^j(\Sigma_0)) \leq \frac{d \cdot C(K)}{\ell \sigma_{min}(Q)}$.

Also, since $\Sigma_K \preceq \frac{C(K)}{\mu\sigma_{min}(Q)}\Sigma_0$,

$$\mathrm{tr}(\mathcal{F}_K^j(\Sigma_K)) \leq \frac{C(K)}{\mu\sigma_{min}(Q)}\mathrm{tr}(\mathcal{F}_K^j(\Sigma_0)) \leq \frac{d \cdot C^2(K)}{\ell\mu\sigma_{min}^2(Q)}.$$

Therefore as long as

$$\ell \geq \frac{dC^2(K)}{\epsilon\mu\sigma_{min}^2(Q)},$$

it follows that:

$$\|\Sigma_K - \Sigma_K^{(\ell)}\| \leq \|\Sigma_K - \Sigma_K^{(j)}\| = \|\mathcal{F}_K^j(\Sigma_K)\| \leq \epsilon.$$

Here the first step is again because of all the terms are PSD, so using more terms is always better. The last step follows because $\mathcal{F}_K^j(\Sigma_K)$ is also a PSD matrix so the spectral norm is bounded by trace. In fact, it holds that $\mathrm{tr}(\Sigma_K - \Sigma_K^{(\ell)})$ is smaller than ϵ.

Next, observe $C^{(\ell)}(K) = \langle \Sigma_K^{(\ell)}, Q + K^\top RK\rangle$ and $C(K) = \langle \Sigma_K, Q + K^\top RK\rangle$, therefore

$$C(K) - C^{(\ell)}(K) \leq \mathrm{tr}(\Sigma_K - \Sigma_K^{(\ell)})(\|Q\| + \|R\|\|K\|^2).$$

Therefore if

$$\ell \geq \frac{d \cdot C^2(K)(\|Q\| + \|R\|\|K\|^2)}{\epsilon\mu\sigma_{min}^2(Q)},$$

then $\mathrm{tr}(\Sigma_K - \Sigma_K^{(\ell)}) \leq \epsilon/(\|Q\| + \|R\|\|K\|^2)$ and hence $C(K) - C^{(\ell)}(K) \leq \epsilon$.

□

D.2. Perturbation of $C(K)$ and $\nabla C(K)$

The next lemma show that the function value and its gradient are approximate preserved if a small perturbation to the policy K is applied.

Lemma 24. *(C_K perturbation) Suppose K' is such that:*

$$\|K' - K\| \leq \min\left(\frac{\sigma_{\min}(Q)\mu}{4C(K)\|B\|\left(\|A - BK\| + 1\right)}, \|K\|\right)$$

then:

$$\begin{aligned} & |C(K') - C(K)| \\ \leq\ & 6\|K\|\,\|R\|\,\mathbb{E}\|x_0\|^2\left(\frac{C(K)}{\mu\,\sigma_{\min}(Q)}\right)^2\left(\|K\|\|B\|\|A - BK\| + \|K\|\|B\| + 1\right)\|K - K'\| \end{aligned}$$

Proof. As in the proof of Lemma 19, the assumption implies that $\|\mathcal{T}_K\|\|\mathcal{F}_K - \mathcal{F}_{K'}\| \leq 1/2$, and, from Equation 16, that

$$\|\mathcal{F}_K - \mathcal{F}_{K'}\| \leq 2\|B\|\left(\|A - BK\| + 1\right)\|K - K'\|$$

First, observe:

$$\begin{aligned} C(K') - C(K) &\leq \mathrm{Tr}(\mathbb{E}x_0x_0^\top)\|\mathcal{T}_{K'}(Q + (K')^\top RK') - \mathcal{T}_K(Q + K^\top RK)\| \\ &= \mathbb{E}\|x_0\|^2\|\mathcal{T}_{K'}(Q + (K')^\top RK') - \mathcal{T}_K(Q + K^\top RK)\| \\ &= \mathbb{E}\|x_0\|^2\|P_{K'} - P_k\|. \end{aligned}$$

To bound the difference we just need to bound $\|P_{K'} - P_k\|$. For that we have

$$\begin{aligned}
& P_{K'} - P_K \\
= \;& \|\mathcal{T}_{K'}(Q + (K')^\top RK') - \mathcal{T}_K(Q + K^\top RK)\| \\
\leq \;& \|\mathcal{T}_{K'}(Q + (K')^\top RK') - \mathcal{T}_K(Q + (K')^\top RK') \\
& - \left(\mathcal{T}_K(Q + K^\top RK) - \mathcal{T}_K(Q + (K')^\top RK')\right)\| \\
= \;& \|\mathcal{T}_{K'}(Q + (K')^\top RK') - \mathcal{T}_K(Q + (K')^\top RK') - \mathcal{T}_K \circ (K^\top RK - (K')^\top RK')\| \\
\leq \;& 2\|\mathcal{T}_K\|^2 \|\mathcal{F}_K - \mathcal{F}_{K'}\| \|(K')^\top RK')\| + \|\mathcal{T}_K\| \|K^\top RK - (K')^\top RK')\| \\
\leq \;& 2\|\mathcal{T}_K\|^2 \|\mathcal{F}_K - \mathcal{F}_{K'}\| \left(\|(K')^\top RK') - K^\top RK\| + \|K^\top RK)\|\right) \\
& + \|\mathcal{T}_K\| \|K^\top RK - (K')^\top RK')\| \\
\leq \;& \|\mathcal{T}_K\| \|(K')^\top RK') - K^\top RK\| + 2\|\mathcal{T}_K\|^2 \|\mathcal{F}_K - \mathcal{F}_{K'}\| \|K^\top RK\| \\
& + \|\mathcal{T}_K\| \|K^\top RK - (K')^\top RK')\| \\
= \;& 2\|\mathcal{T}_K\| \|(K')^\top RK') - K^\top RK\| + 2\|\mathcal{T}_K\|^2 \|\mathcal{F}_K - \mathcal{F}_{K'}\| \|K^\top RK\|
\end{aligned}$$

For the first term,

$$\begin{aligned}
2\|\mathcal{T}_K\| \|(K')^\top RK') - K^\top RK\| &\leq 2\|\mathcal{T}_K\| \left(2\|K\|\|R\|\|K' - K\| + \|R\|\|K' - K\|^2\right) \\
&\leq 2\|\mathcal{T}_K\| \left(3\|K\|\|R\|\|K' - K\|\right)
\end{aligned}$$

using the assumption that $\|K' - K\| \leq \|K\|$. For the second term,

$$2\|\mathcal{T}_K\|^2 \|\mathcal{F}_K - \mathcal{F}_{K'}\| \|K^\top RK\| \leq 2\|\mathcal{T}_K\|^2 \, 2\|B\| \left(\|A - BK\| + 1\right) \|K - K'\| \, \|K\|^2 \|R\| \, .$$

Combining the two terms completes the proof. □

The next lemma shows the gradient is also stable after perturbation.

Lemma 25. *(∇C_K perturbation) Suppose K' is such that:*

$$\|K' - K\| \leq \min\left(\frac{\sigma_{\min}(Q)\mu}{4C(K)\|B\| \left(\|A - BK\| + 1\right)}, \|K\|\right)$$

then there is a polynomial h_{grad} in $\frac{C(K_0)}{\mu\sigma_{min}(Q)}, \mathbb{E}[\|x_0\|^2], \|A\|, \|B\|, \|R\|, \frac{1}{\sigma_{min}(R)}$ such that

$$\|\nabla C(K') - \nabla C(K)\| \leq h_{grad}\|K' - K\|.$$

Also,

$$\|\nabla C(K') - \nabla C(K)\|_F \leq h_{grad}\|K' - K\|_F.$$

Proof. Recall $\nabla C(K) = 2E_K\Sigma_K$ where $E_K = (R + B^\top P_K B)K - B^\top P_K A$. Therefore

$$\nabla C(K') - \nabla C(K) = 2E_{K'}\Sigma_{K'} - 2E_K\Sigma_K = 2(E_{K'} - E_K)\Sigma_{K'} + 2E_K(\Sigma_{K'} - \Sigma_K).$$

Let's first look at the second term. By Lemma 11,

$$\mathrm{Tr}(E_K^\top E_K) \leq \frac{\|R + B^\top P_K B\|(C(K) - C(K^*))}{\mu},$$

then by Lemma 16

$$\|\Sigma_{K'} - \Sigma_K\| \leq 4\left(\frac{C(K)}{\sigma_{\min}(Q)}\right)^2 \frac{\|B\| \left(\|A - BK\| + 1\right)}{\mu}\|K - K'\|$$

Therefore the second term is bounded by

$$8\left(\frac{C(K)}{\sigma_{\min}(Q)}\right)^2 \frac{\left(\|R + B^\top P_K B\|(C(K) - C(K^*))\right)\|B\| \left(\|A - BK\| + 1\right)}{\mu^2}\|K - K'\|.$$

Next we bound the first term. Since $K' - K$ is small enough, $\|\Sigma_{K'}\| \le \|\Sigma_K\| + \frac{C(K)}{\sigma_{min}(Q)}$.

For $E_{K'} - E_K$, we first need a bound on $P_{K'} - P_K$. By the previous lemma,

$$\|P_K' - P_K\| = 6\left(\left(\frac{C(K)}{\mu\,\sigma_{\min}(Q)}\right)^2 \|K\|^2\|R\|\|B\|\left(\|A-BK\|+1\right) + \left(\frac{C(K)}{\mu\,\sigma_{\min}(Q)}\right)\|K\|\|R\|\right)\|K-K'\|.$$

Therefore

$$E_K' - E_K = R(K'-K) + B^\top(P_{K'}-P_K)A + B^\top(P_{K'}-P_K)BK' + B^\top P_K B(K'-K).$$

Since $\|K'\| \le 2\|K\|$, and $\|K\|$ can be bounded by $C(K)$ (Lemma 22), all the terms can be bounded by polynomials of related parameters multiplied by $\|K - K'\|$. □

D.3. Smoothing and the gradient descent analysis

This section analyzes the smoothing procedure and completes the proof of gradient descent. Although Gaussian smoothing is more standard, the objective $C(K)$ is not finite for every K, therefore technically $\mathbb{E}_{u\sim\mathcal{N}(0,\sigma^2 I)}[C(K+u)]$ is not well defined. This is avoidable by smoothing in a ball.

Let $\mathbb{S}_r$ represent the uniform distribution over the points with norm r (boundary of a sphere), and $\mathbb{B}_r$ represent the uniform distribution over all points with norm at most r (the entire sphere). When applying these sets to matrix a U, the Frobenius norm ball is used. The algorithm performs gradient descent on the following function

$$C_r(K) = \mathbb{E}_{U\sim\mathbb{B}_r}[C(K+U)].$$

The next lemma uses the standard technique (e.g. in (Flaxman et al., 2005)) to show that the gradient of $C_r(K)$ can be estimated just with an oracle for function value.

Lemma 26. $\nabla C_r(K) = \frac{d}{r^2}\mathbb{E}_{U\sim\mathbb{S}_r}[C(K+U)U]$.

This is the same as Lemma 2.1 in Flaxman et al. (2005), for completeness the proof is provided below.

Proof. By Stokes formula,

$$\nabla\int_{\delta\mathbb{B}_r} C(K+U)dx = \int_{\delta\mathcal{S}_r} C(K+U)\frac{U}{\|U\|_F}dx.$$

By definition,

$$C_r(K) = \frac{\int_{\delta\mathbb{B}_r} C(K+U)dx}{\mathrm{vol}_d(\delta\mathbb{B}_r)},$$

Also,

$$\mathbb{E}_{U\sim\mathbb{S}_r}[C(K+U)U] = r\mathbb{E}_{U\sim\mathbb{S}_r}[C(K+U)\frac{U}{r}] = r\cdot\frac{\int_{\delta\mathcal{S}_r} C(K+U)\frac{U}{\|U\|_F}dx}{\mathrm{vol}_{d-1}(\delta\mathbb{S}_r)}.$$

The Lemma follows from combining these equations, and use the fact that

$$\mathrm{vol}_d(\delta\mathbb{B}_r) = \mathrm{vol}_{d-1}(\delta\mathbb{S}_r)\cdot\frac{r}{d}.$$

□

From the lemma above and standard concentration inequalities, it is immediate that it suffices to use a polynomial number of samples to approximate the gradient.

Lemma 27. *Given an ϵ, there are fixed polynomials $h_r(1/\epsilon), h_{sample}(d, 1/\epsilon)$ such that when $r \leq 1/h_r(1/\epsilon)$, with $m \geq h_{sample}(d, 1/\epsilon)$ samples of $U_1, ..., U_n \sim \mathbb{S}_r$, with high probability (at least $1-(d/\epsilon)^{-d}$) the average*

$$\hat{\nabla} = \frac{1}{m}\sum_{i=1}^{m} \frac{d}{r^2} C(K+U_i)U_i$$

is ϵ close to $\nabla C(K)$ in Frobenius norm.

Further, if for $x \sim \mathcal{D}$, $\|x\| \leq L$ almost surely, there are polynomials $h_{\ell,grad}(d, 1/\epsilon)$, $h_{r,trunc}(1/\epsilon)$, $h_{sample,trunc}(d, 1/\epsilon, \sigma, L^2/\mu)$ such that when $m \geq h_{sample,trunc}(d, 1/\epsilon, L^2/\mu)$, $\ell \geq h_{\ell,grad}(d, 1/\epsilon)$, let $x_j^i, u_j^i (0 \leq j \leq \ell)$ be a single path sampled using $K+U_i$, then the average

$$\tilde{\nabla} = \frac{1}{m}\sum_{i=1}^{m} \frac{d}{r^2} [\sum_{j=0}^{\ell-1} (x_j^i)^\top Q x_j^i + (u_j^i)^\top R u_j^i] U_i$$

is also ϵ close to $\nabla C(K)$ in Frobenius norm with high probability.

Proof. For the first part, the difference is broken into two terms:

$$\hat{\nabla} - \nabla C(K) = (\nabla C_r(K) - \nabla C(K)) + (\hat{\nabla} - \nabla C_r(K)).$$

For the first term, choose $h_r(1/\epsilon) = \min\{1/r_0, 2h_{grad}/\epsilon\}$ (r_0 is chosen later). By Lemma 25 when r is smaller than $1/h_r(1/\epsilon) = \epsilon/2h_{grad}$, every point u on the sphere have $\|\nabla C(K+U) - \nabla C(K)\|_F \leq \epsilon/4$. Since $\nabla C_r(K)$ is the expectation of $\nabla C(K+U)$, by triangle inequality $\|\nabla C_r(K) - \nabla C(K)\|_F \leq \epsilon/2$.

The proof also makes sure that $r \leq r_0$ such that for any $U \sim \mathbb{S}_r$, it holds that $C(K+U) \leq 2C(K)$. By Lemma 24, $1/r_0$ is a polynomial in the relevant factors.

For the second term, by Lemma 26, $\mathbb{E}[\hat{\nabla}] = \nabla C_r(K)$, and each individual sample has norm bounded by $2dC(K)/r$, so by Vector Bernstein's Inequality, know with $m \geq h_{sample}(d, 1/\epsilon) = \Theta\left(d\left(\frac{dC(K)}{\epsilon r}\right)^2 \log d/\epsilon\right)$ samples, with high probability (at least $1-(d/\epsilon)^{-d}$) $\|\hat{\nabla} - \mathbb{E}[\hat{\nabla}]\|_F \leq \epsilon/2$.

Adding these two terms and apply triangle inequality gives the result.

For the second part, the proof breaks it into more terms. Let ∇' be equal to $\frac{1}{m}\sum_{i=1}^{m} \frac{d}{r^2} C^{(\ell)}(K+U_i)U_i$ (where $C^{(\ell)}$ is defined as in Lemma 23), then

$$\tilde{\nabla} - \nabla C(K) = (\tilde{\nabla} - \nabla') + (\nabla' - \hat{\nabla}) + (\hat{\nabla} - \nabla C(K)).$$

The third term is just what was bounded earlier, by choosing $h_{r,trunc}(1/\epsilon) = h_r(2/\epsilon)$ and making sure $h_{sample,trunc}(d, 1/\epsilon) \geq h_{sample}(d, 2/\epsilon)$, we guarantees that it is smaller than $\epsilon/2$.

For the second term, choose $\ell \geq \frac{16d^2 \cdot C^2(K)(\|Q\| + \|R\|\|K\|^2)}{\epsilon r \mu \sigma_{min}^2(Q)} =: h_{\ell,grad}(d, 1/\epsilon)$. By Lemma 23, for any K' with $C(K') \leq 2C(K)$, it holds that $\|C^{(\ell)}(K') - C(K')\| \leq \frac{r\epsilon}{4d}$. Therefore by triangle inequality

$$\|\frac{1}{m}\sum_{i=1}^{m} \frac{d}{r^2} C^{(\ell)}(K+U_i)U_i - \frac{1}{m}\sum_{i=1}^{m} \frac{d}{r^2} C(K+U_i)U_i\| \leq \epsilon/4.$$

Finally for the first term it is easy to see that $\mathbb{E}[\tilde{\nabla}] = \nabla'$ where the expectation is taken over the randomness of the initial states x_0^i. Since $\|x_0^i\| \leq L$, $(x_0^i)(x_0^i)^\top \preceq \frac{L^2}{\mu}\mathbb{E}[x_0 x_0^\top]$, as a result the sum

$$[\sum_{j=0}^{\ell-1} (x_j^i)^\top Q x_j^i + (u_j^i)^\top R u_j^i] \leq \frac{L^2}{\mu} C(K+U_i).$$

Therefore, $\tilde{\nabla} - \nabla'$ is again a sum of independent vectors with bounded norm, so by Vector Bernstein's inequality, when $h_{sample,trunc}(d, 1/\epsilon, L^2/\mu)$ is a large enough polynomial, $\|\tilde{\nabla} - \nabla'\| \leq \epsilon/4$ with high probability. Adding all the terms finishes the proof. □

Theorem 28. *There are fixed polynomials $h_{GD,r}(1/\epsilon), h_{GD,sample}(d, 1/\epsilon, L^2/\mu), h_{GD,\ell}(d, 1/\epsilon)$ such that if every step the gradient is computed as Lemma 27 (truncated at step ℓ), pick step size η and T the same as the gradient descent case of Theorem 7, it holds that $C(K_T) - C(K^\star) \le \epsilon$ with high probability (at least $1 - \exp(-d)$).*

Proof. By Lemma 21, when $\eta \le 1/h_{GD,\eta}$ for some fixed polynomial $h_{GD,\eta}$(given in Lemma 21), then

$$C(K') - C(K^*) \le \left(1 - \eta\sigma_{\min}(R)\frac{\mu^2}{\|\Sigma_{K^*}\|}\right)(C(K) - C(K^*))$$

Let $\tilde{\nabla}$ be the approximate gradient computed, and let $K'' = K - \eta\tilde{\nabla}$ be the iterate that uses the approximate gradient. The proof shows given enough samples, the gradient can be estimated with enough accuracy that makes sure

$$|C(K'') - C(K')| \le \frac{1}{2}\eta\sigma_{\min}(R)\frac{\mu^2}{\|\Sigma_{K^*}\|} \cdot \epsilon.$$

This means as long as $C(K) - C(K^*) \ge \epsilon$, it holds that

$$C(K'') - C(K^*) \le \left(1 - \frac{1}{2}\eta\sigma_{\min}(R)\frac{\mu^2}{\|\Sigma_{K^*}\|}\right)(C(K) - C(K^*)).$$

Then the same proof of Theorem 7 gives the convergence guarantee.

Now $C(K'') - C(K')$ is bounded. By Lemma 24, if $\|K'' - K'\| \le \frac{1}{2}\eta\sigma_{\min}(R)\frac{\mu^2}{\|\Sigma_{K^*}\|} \cdot \epsilon \cdot 1/h_{func}$ (h_{func} is the polynomial in Lemma 24), then $C(K'') - C(K')$ is small enough. To get that, observe $K'' - K' = \eta(\nabla - \tilde{\nabla})$, therefore it suffices to make sure

$$\|\nabla - \tilde{\nabla}\| \le \frac{1}{2}\sigma_{\min}(R)\frac{\mu^2}{\|\Sigma_{K^*}\|} \cdot \epsilon \cdot 1/h_{func}$$

By Lemma 25, it suffices to pick $h_{GD,r}(1/\epsilon) = h_{r,trunc}(2h_{func}\|\Sigma_{K^*}\|/(\mu^2\sigma_{min}(R)\epsilon))$, $h_{GD,sample}(d, 1/\epsilon, L^2/\mu) = h_{sample,trunc}(d, 2h_{func}\|\Sigma_{K^*}\|/(\mu^2\sigma_{min}(R)\epsilon), L^2/\mu)$, and $h_{GD,\ell}(d, 1/\epsilon) = h_{\ell,grad}(d, 2h_{func}\|\Sigma_{K^*}\|/(\mu^2\sigma_{min}(R)\epsilon))$. This gives the desired upper-bound on $\|\nabla - \tilde{\nabla}\|$ with high probability (at least $1 - (\epsilon/d)^{-d}$).

Since the number of steps is a polynomial, by union bound with high probability (at least $1 - T(\epsilon/d)^{-d} \ge 1 - \exp(-d)$) the gradient is accurate enough for all the steps, so

$$C(K'') - C(K^*) \le \left(1 - \frac{1}{2}\eta\sigma_{\min}(R)\frac{\mu^2}{\|\Sigma_{K^*}\|}\right)(C(K) - C(K^*)).$$

The rest of the proof is the same as Theorem 7. Note that in the smoothing, because the function value is monotonically decreasing and the choice of radius, all the function value encountered is bounded by $2C(K_0)$, so the polynomials are indeed bounded throughout the algorithm. □

D.4. The natural gradient analysis

Before the Theorem for natural gradient is proven, the following lemma shows the variance can be estimated accurately.

Lemma 29. *If for $x \sim \mathcal{D}$, $\|x\| \le L$ almost surely, there exists polynomials $h_{r,var}(1/\epsilon)$, $h_{varsample,trunc}(d, 1/\epsilon, L^2/\mu)$ and $h_{\ell,var}(d, 1/\epsilon)$ such that if $\hat{\Sigma}_K$ is estimated using at least $m \ge h_{varsample,trunc}(d, 1/\epsilon, L^2/\mu)$ initial points $x_0^1, ..., x_0^m$, m random perturbations $U_i \sim \mathbb{S}_r$ where $r \le 1/h_{r,var}(1/\epsilon)$, all of these initial points are simulated using $\hat{K}_i = K + U_i$ to $\ell \ge h_{\ell,var}(d, 1/\epsilon)$ iterations, then with high probability (at least $1 - (d/\epsilon)^{-d}$) the following estimate*

$$\tilde{\Sigma} = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=0}^{\ell-1} x_j^i (x_j^i)^\top.$$

satisfies $\|\tilde{\Sigma} - \Sigma_K\| \le \epsilon$. Further, when $\epsilon \le \mu/2$, it holds that $\sigma_{min}(\hat{\Sigma}_K) \ge \mu/2$.

Proof. This is broken into three terms: let $\Sigma_K^{(\ell)}$ be defined as in Lemma 23, let $\hat{\Sigma} = \frac{1}{m}\sum_{i=1}^m \Sigma_{K+U_i}$ and $\hat{\Sigma}^{(\ell)} = \frac{1}{m}\sum_{i=1}^m \Sigma_{K+U_i}^{(\ell)}$, then it holds that

$$\tilde{\Sigma} - \Sigma_K = (\tilde{\Sigma} - \hat{\Sigma}^{(\ell)}) + (\hat{\Sigma}^{(\ell)} - \hat{\Sigma}) + (\hat{\Sigma} - \Sigma_K).$$

First, r is chosen small enough so that $C(K+U_i) \le 2C(K)$. This only requires an inverse polynomial r by Lemma 24.

For the first term, note that $\mathbb{E}[\tilde{\Sigma}] = \hat{\Sigma}^{(\ell)}$ where the expectation is taken over the initial points x_0^i. Since $\|x_0^i\| \le L$, $(x_0^i)(x_0^i)^\top \preceq \frac{L^2}{\mu}\mathbb{E}[x_0 x_0^\top]$, and as a result the sum

$$\sum_{j=0}^{\ell-1} x_j^i (x_j^i)^\top Q \preceq \frac{L^2}{\mu}\Sigma_{K+U_i}.$$

Therefore, standard concentration bounds show that when $h_{varsample,trunc}$ is a large enough polynomial, $\|\tilde{\Sigma} - \hat{\Sigma}^{(\ell)}\| \le \epsilon/2$ holds with high probability.

For the second term, Lemma 23 is applied. Because $C(K+U_i) \le 2C(K)$, choosing $\ell \ge h_{\ell,var}(d, 1/\epsilon) = \frac{8d\cdot C^2(K)}{\epsilon\mu\sigma_{min}^2(Q)}$, the error introduced by truncation $\|\hat{\Sigma}^{(\ell)} - \hat{\Sigma}\|$ is then bounded by $\epsilon/4$.

For the third term, Lemma 16 is applied. When $r \le \epsilon \cdot \left(\frac{\sigma_{\min}(Q)}{C(K)}\right)^2 \frac{\mu}{16\|B\|(\|A-BK\|+1)}$, $\|\Sigma_{K+U_i} - \Sigma_K\| \le \epsilon/4$. Since $\hat{\Sigma}$ is the average of Σ_{K+U_i}, by the triangle inequality, $\|\hat{\Sigma} - \Sigma_K\| \le \epsilon/4$.

Adding these three terms gives the result.

Finally, the bound on $\sigma_{min}(\tilde{\Sigma}_K)$ follows simply from Weyl's Theorem. □

Theorem 30. *Suppose $C(K_0)$ is finite and $\mu > 0$. The natural gradient follows the update rule:*

$$K_{t+1} = K_t - \eta \nabla C(K_t)\Sigma_{K_t}^{-1}$$

Suppose the stepsize is set to be:

$$\eta = \frac{1}{\|R\| + \frac{\|B\|^2 C(K_0)}{\mu}}$$

If the gradient and variance are estimated as in Lemma 27, Lemma 29 with $r = 1/h_{NGD,r}(1/\epsilon)$, with $m \ge h_{NGD,sample}(d, 1/\epsilon, L^2/\mu)$ samples, both are truncated to $h_{NGD,\ell}(d, 1/\epsilon)$ iterations, then with high probability (at least $1 - \exp(-d)$) in T iterations where

$$T > \frac{\|\Sigma_{K^*}\|}{\mu}\left(\frac{\|R\|}{\sigma_{min}(R)} + \frac{\|B\|^2 C(K_0)}{\mu\sigma_{min}(R)}\right)\log\frac{2(C(K_0) - C(K^*))}{\varepsilon}$$

then the natural gradient satisfies the following performance bound:

$$C(K_T) - C(K^*) \le \varepsilon$$

Proof. By Lemma 15,

$$C(K') - C(K^*) \le \left(1 - \eta\sigma_{\min}(R)\frac{\mu}{\|\Sigma_{K^*}\|}\right)(C(K) - C(K^*))$$

Let $\tilde{\nabla}$ be the estimated gradient, $\tilde{\Sigma}_K$ be the estimated Σ_K, and let $K'' = K - \eta\tilde{\nabla}\tilde{\Sigma_K}^{-1}$. The proof shows that when both the gradient and the covariance matrix are estimated accurately enough, then

$$|C(K') - C(K'')| \le \frac{\epsilon}{2}\eta\sigma_{\min}(R)\frac{\mu}{\|\Sigma_{K^*}\|}.$$

This implies when $C(K) - C(K^\star) \geq \epsilon$,

$$C(K') - C(K^*) \leq \left(1 - \frac{1}{2}\eta\sigma_{\min}(R)\frac{\mu}{\|\Sigma_{K^*}\|}\right)(C(K) - C(K^*))$$

which is sufficient for the proof.

By Lemma 24, if $\|K'' - K'\| \leq \frac{\epsilon}{2h_{func}}\eta\sigma_{\min}(R)\frac{\mu}{\|\Sigma_{K^*}\|}$ the desired bound on $|C(K') - C(K'')|$ holds. To achieve this, it suffices to have

$$\|\tilde{\nabla}\tilde{\Sigma}_K^{-1} - \nabla C(K)\Sigma_K^{-1}\| \leq \frac{\epsilon}{2h_{func}}\sigma_{\min}(R)\frac{\mu}{\|\Sigma_{K^*}\|}.$$

This is broken into two terms

$$\|\tilde{\nabla}\tilde{\Sigma}_K^{-1} - \nabla C(K)\Sigma_K^{-1}\| \leq \|\tilde{\nabla} - \nabla\|\|\tilde{\Sigma}_K^{-1}\| + \|\nabla C(K)\|\|\tilde{\Sigma}_K^{-1} - \Sigma_K^{-1}\|.$$

For the first term, by Lemma 29 we know when the number of samples is large enough $\|\tilde{\Sigma}_K^{-1}\| \leq 2/\mu$. Therefore it suffices to make sure $\|\tilde{\nabla} - \nabla\| \leq \frac{\epsilon}{8h_{func}}\sigma_{\min}(R)\frac{\mu^2}{\|\Sigma_{K^*}\|}$, this can be done by Lemma 27 by setting $h_{NGD,grad,r}(1/\epsilon) = h_{r,trunc}(\frac{8h_{func}\|\Sigma_K^*\|}{\mu^2\sigma_{min}(R)\epsilon})$,
$h_{NGD,gradsample}(d, 1/\epsilon, L/\mu^2) = h_{sample,trunc}(d, \frac{8h_{func}\|\Sigma_K^*\|}{\mu^2\sigma_{min}(R)\epsilon}, L/\mu^2)$ and
$h_{NGD,\ell,grad}(d, 1/\epsilon) = h_{\ell,grad}(d, \frac{8h_{func}\|\Sigma_K^*\|}{\mu^2\sigma_{min}(R)\epsilon})$.

For the second term, it suffices to make sure $\|\tilde{\Sigma}_K^{-1} - \Sigma_K^{-1}\| \leq \frac{\epsilon}{4h_{func}}\sigma_{\min}(R)\frac{\mu}{\|\Sigma_{K^*}\|\|\nabla C(K)\|}$.. By standard matrix perturbation, if $\sigma_{min}(\Sigma_K) \geq \mu$ and $\|\tilde{\Sigma}_K - \Sigma_K\| \leq \mu/2$, $\|\tilde{\Sigma}_K^{-1} - \Sigma_K^{-1}\| \leq 2\|\tilde{\Sigma}_K - \Sigma_K\|/\mu^2$. Therefore by Lemma 29 it suffices to choose $h_{NGD,var,r}(1/\epsilon) = h_{var,r}(\frac{8h_{func}\|\Sigma_{K^*}\|\|\nabla C(K)\|}{\mu^3\sigma_{min}(R)\epsilon})$, $h_{NGD,varsample}(d, 1/\epsilon, L/\mu^2) = h_{varsample,trunc}(d, \frac{8h_{func}\|\Sigma_{K^*}\|\|\nabla C(K)\|}{\mu^3\sigma_{min}(R)\epsilon}, L/\mu^2)$ and $h_{NGD,\ell,var}(d, 1/\epsilon) = h_{\ell,var}(d, \frac{8h_{func}\|\Sigma_{K^*}\|\|\nabla C(K)\|}{\mu^3\sigma_{min}(R)\epsilon})$. This is indeed a polynomial because $\|\nabla C(K)\|$ is bounded by Lemma 22.

Finally, choose $h_{NGD,r} = \max\{h_{NGD,grad,r}, h_{NGD,var,r}\}$,
$h_{NGD,sample} = \max\{h_{NGD,gradsample}, h_{NGD,varsample}\}$, and $h_{NGD,\ell} = \max\{h_{NGD,\ell,grad}, h_{NGD,\ell,var}\}$. This ensures all the bounds mentioned above hold and that

$$C(K') - C(K^*) \leq \left(1 - \frac{1}{2}\eta\sigma_{\min}(R)\frac{\mu}{\|\Sigma_{K^*}\|}\right)(C(K) - C(K^*))$$

The rest of the proof is the same as Theorem 7. Note again that in the smoothing, because the function value is monotonically decreasing and the choice of radius, all the function values encountered are bounded by $2C(K_0)$, so the polynomials are indeed bounded throughout the algorithm. □

D.5. Standard Matrix Perturbation and Concentrations

In the previous sections, we used several standard tools in matrix perturbation and concentration, which we summarize here. The matrix perturbation theorems can be found in Stewart & Sun (1990). Matrix concentration bounds can be found in Tropp (2012)

Theorem 31 (Weyl's Theorem)**.** *Suppose $B = A + E$, then the singular values of B are within $\|E\|$ to the corresponding singular values of A. In particular $\|B\| \leq \|A\| + \|E\|$ and $\sigma_{min}(B) \geq \sigma_{min}(A) - \|E\|$.*

Theorem 32 (Perturbation of Inverse)**.** *Let $B = A + E$, suppose $\|E\| \leq \sigma_{min}(A)/2$ then $\|B^{-1} - A^{-1}\| \leq 2\|A - B\|/\sigma_{min}(A)$.*

Theorem 33 (Matrix Bernstein)**.** *Suppose $\hat{A} = \sum_i \hat{A}_i$, where $\hat{A}_i$ are independent random matrices of dimension $d_1 \times d_2$ (let $d = d_1 + d_2$). Let $\mathbb{E}[\hat{A}] = A$, the variance $M_1 = \mathbb{E}[\sum_i \hat{A}_i\hat{A}_i^\top]$, $M_2 = \mathbb{E}[\sum_i \hat{A}_i^\top\hat{A}_i]$. If $\sigma^2 = \max\{\|M_1\|, \|M_2\|\}$, and every $\hat{A}_i$ has spectral norm $\|\hat{A}_i\| \leq R$ with probability 1, then with high probability*

$$\|\hat{A} - A\| \leq O(R\log d + \sqrt{\sigma^2 \log d}).$$

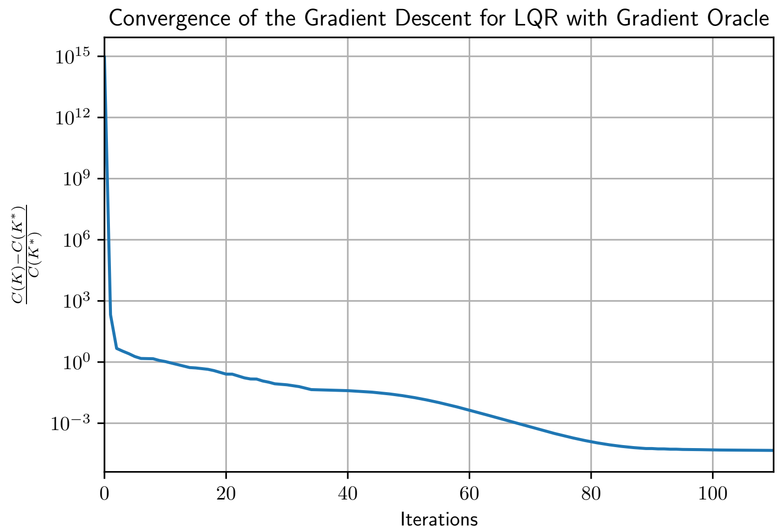

Figure 1. Simulation results with Gradient Descent[4]

In our proof we often treat a matrix as a vector and look at its Frobenius norm, in these cases we use the following corollary:

Theorem 34 (Vector Bernstein)**.** *Suppose $\hat{a} = \sum_i \hat{a}_i$, where $\hat{a}_i$ are independent random vector of dimension d. Let $\mathbb{E}[\hat{a}] = a$, the variance $\sigma^2 = \mathbb{E}[\sum_i \|\hat{a}_i\|^2]$. If every $\hat{a}_i$ has norm $\|\hat{a}_i\| \le R$ with probability 1, then with high probability*

$$\|\hat{a} - a\| \le O(R \log d + \sqrt{\sigma^2 \log d}).$$

E. Simulation Results

Here we give simulations for the gradient descent algorithm (with backtracking step size) to show that the algorithm indeed converges within reasonable time in practice. In this experiment, $x \in \mathbb{R}^{100}$ and $u \in \mathbb{R}^{20}$. We use random matrices A, B. The scaling of A is chosen so that A is stabilizing with high probability ($\lambda_{max}(A) \le 1$). We initialize the solution at $K_0 = 0$, which ensures $C(K_0)$ is finite because A is stabilizing. The distribution of the initial point x_0 is the unit cube. We computed the gradient using Lemma 1. See Figure 1 for the result. Although the example uses the exact gradient for the each iterative step, it provides a glimpse into the potential use of first order methods and their variants for direct feedback gain update in LQR.