Lecture 14

Optimal Distributed Control

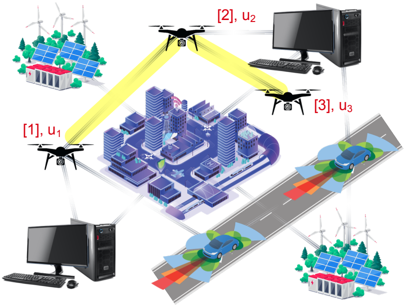
1) The Gradient Descent (GD) algorithm for stochastic LQR

Giancarlo Ferrari Trecate[1]

[1]Dependable Control and Decision Group
École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
giancarlo.ferraritrecate@epfl.ch

Slides: Luca Furieri

Optimal control at the large-scale

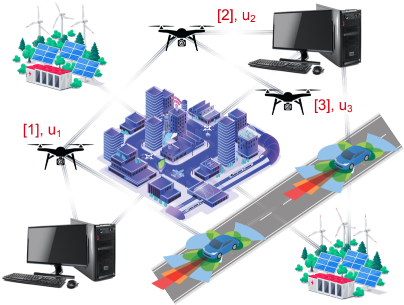


- Multiple, heterogeneous, dynamically coupled

$$x_{t+1}^{[1]} = \tau(x_t^{[2]})$$

- *Local* real-time measurements

Optimal control at the large-scale



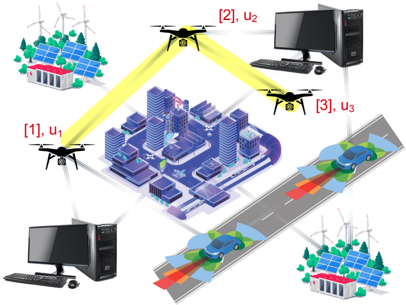
- Multiple, heterogeneous, dynamically coupled

$$x_{t+1}^{[1]} = \tau(x_t^{[2]})$$

- *Local* real-time measurements

Can we apply LQR/LQG? Not really...

Optimal control at the large-scale



- Multiple, heterogeneous, dynamically coupled

$$x_{t+1}^{[1]} = \tau(x_t^{[2]})$$

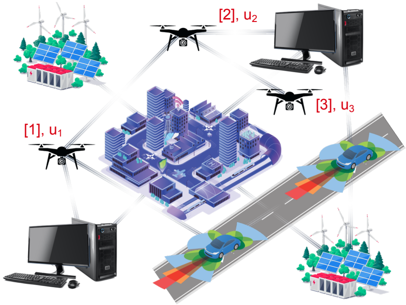
- *Local* real-time measurements

Can we apply LQR/LQG? Not really...

- LQR controller :

$$u_t = \begin{bmatrix} u_t^{[1]} \\ u_t^{[2]} \\ u_t^{[3]} \end{bmatrix} = -\underbrace{\begin{bmatrix} k^{11} & k^{12} & k^{13} \\ k^{21} & k^{22} & k^{23} \\ k^{31} & k^{32} & k^{33} \end{bmatrix}}_{K_{LQR}} \begin{bmatrix} x_t^{[1]} \\ x_t^{[2]} \\ x_t^{[3]} \end{bmatrix}$$

Optimal control at the large-scale



- Multiple, heterogeneous, dynamically coupled

$$x_{t+1}^{[1]} = \tau(x_t^{[2]})$$

- *Local* real-time measurements

Can we apply LQR/LQG? Not really...

- LQR controller :

$$u_t = \begin{bmatrix} u_t^{[1]} \\ u_t^{[2]} \\ u_t^{[3]} \end{bmatrix} = - \begin{bmatrix} k^{11} & k^{12} & k^{13} \\ k^{21} & k^{22} & k^{23} \\ k^{31} & k^{32} & k^{33} \end{bmatrix} \begin{bmatrix} x_t^{[1]} \\ x_t^{[2]} \\ x_t^{[3]} \end{bmatrix}$$

Not implementable! $u^{[3]}$ cannot measure $x^{[1]}$

$$\Rightarrow u_t^{[3]} = k_{32} x_t^{[2]} + k_{33} x_t^{[3]}$$

Optimal control at the large-scale

Optimal distributed control

Find the best control policy $u_t = -Kx_t \ldots K$ subject to "sparsity constraints".

Example

$$K = \begin{bmatrix} * & * & 0 \\ * & * & * \\ 0 & * & * \end{bmatrix}$$

where "$*$" is any real value.

Plan of the next two lectures

1. **Gradient Descent (GD) for stochastic LQR**
   - S-LQR as an instance of LQG
   - Performance evaluation for given K
   - Computing the gradient of S-LQR
   - Convergence results

   centralized

2. **Projected GD for locally optimal distributed controllers**
   - Binary matrices, sparsity subspaces and Distributed S-LQR
   - Projected Gradient Descent (PGD) for DS-LQR
   - Convergence results

Stochastic LQR (S-LQR)

System under control

$$x_{t+1} = Ax_t + Bu_t \qquad x_t \in \mathbb{R}^n \quad x_0 \sim N(\mu, \Sigma_0), \Sigma_0 \geq 0$$

$$x(0) = x_0 \qquad u_t \in \mathbb{R}^m$$

Control cost over an Infinite Horizon (IH)

$$J = \mathbb{E}_{x_0}\left[\sum_{t=0}^{\infty} x_t^\top Q x_t + u_t^\top R u_t\right]$$

where

$$Q = Q^\top \geq 0$$

$$R = R^\top > 0$$

Goal

Find the optimal sequence $u_0, u_1, \ldots, u_\infty$ that minimizes J.

Remarks on S-LQR

1. The cost J is defined as an EXPECTED VALUE over $x_0 \sim N(\mu, \Sigma_0)$
   - Different realizations of $x_0 \Rightarrow$ different $\sum_{t=0}^{\infty} x_t^\top Q x_t + u_t^\top R u_t$
   - The $\mathbb{E}[\cdot]$ in front of $\sum_{t=0}^{\infty} x_t^\top Q x_t + u_t^\top R u_t$ gets rid of stochasticity
2. S-LQR is a special case of LQG with $C = I$, $V = 0$ and $W = 0$!
   - at home : derive KF equations for S-LQR
   - LQG assumes $V \neq 0$, which is violated for S-LQR. Why is this not a problem?

$x^+ = Ax + Bu + w \qquad w \sim WGN(0, W)$

$y = Cx + v \qquad v \sim WGN(0, V)$

$x_0 \sim N(\mu, \Sigma_0)$

Theorem: S-LQR solution

Theorem: S-LQR solution

Assume (A, B) is reachable. The optimal control policy for S-LQR is a linear state-feedback

$$u_t = -K^* x_t$$

where $K^* = \left(R + B^\top P B\right)^{-1} B^\top P A$, and P is such that

$$P = A^\top P A + Q - A^\top P B \left(R + B^\top P B\right)^{-1} B^\top P A.$$

The optimal cost is given by

$$J^* = \mathrm{Trace}(P\Sigma_0) + \mu^\top P \underbrace{\mu}_{\mathbb{E}[x_0]}$$

Theorem: S-LQR solution

Remarks

1. Despite stochasticity on x_0, no need of an observer to reconstruct x_t.
   - $u_t = -K^* x_t$, where x_t is the **actual** state.
   - In LQG, $u_t = -K^* \hat{x}_t$, where $\hat{x}_t$ is the **observer** state.

2. K^* is still **centralized** (dense matrix).

Theorem: S-LQR solution

Proof of the Theorem

- Let $J = \mathbb{E}_{x_0} C(x_0)$, where

$$C(x_0) = \sum_{t=0}^{\infty} x_t^\top Q x_t + u_t^\top R u_t$$

- $C(x_0)$ is the IH-LQR cost of Lecture 8.
- For a fixed x_0 we know that the optimal controller for $C(x_0)$ is

$$K^* = \left(R + B^\top P B\right)^{-1} B^\top P A$$

leading to the optimal LQR cost $C^*(x_0) = x_0^\top P x_0$ where P solves the ARE for LQ control.

$\int_{\mathbb{R}^n} C(x_0) f_{x_0}(x_0)\,dx_0 \geq \int_{\mathbb{R}^n} C^*(x_0) f_{x_0}(x_0)\,dx_0$

Note that $\forall x_0, K^* = \arg\min C(x_0)$, i.e. K^* does not depend upon x_0. Therefore, $K^* = \arg\min \mathbb{E}_{x_0} C(x_0)$, i.e. K^* is optimal for S-LQR (same controller as LQR).

Theorem: S-LQR solution

Proof 2

However, the cost is now

e.g. $\begin{bmatrix} x_0(1) & x_0(2) \end{bmatrix} \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} x_0(1) \\ x_0(2) \end{bmatrix} =$

$\begin{bmatrix} x_0(1) & x_0(2) \end{bmatrix} \begin{bmatrix} P_{11} x_0(1) + P_{12} x_0(2) \\ P_{21} x_0(1) + P_{22} x_0(2) \end{bmatrix}$

$$J = \mathbb{E}_{x_0} x_0^\top P x_0 = \mathbb{E}_{x_0}\left[\sum_{i=1}^n \sum_{j=1}^n P_{ij} x_0(i) x_0(j)\right]$$

$$= \sum_{i=1}^n \sum_{j=1}^n P_{ij} \mathbb{E}_{x_0}\left[x_0(i) x_0(j)\right]$$ By linearity of $\mathbb{E}$

$$= \sum_{i=1}^n \sum_{j=1}^n P_{ij}\left(\Sigma_0(i,j) + \mu_i \mu_j\right)$$ $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \underbrace{\mathbb{E}[X]}_{\mu_X}\underbrace{\mathbb{E}[Y]}_{\mu_Y}$

$n=2 \rightarrow \underbrace{P_{11}\Sigma_0(1,1) + P_{12}\Sigma_0(2,1)}_{i=1} + \underbrace{P_{21}\Sigma_0(1,2) + P_{22}\Sigma_0(2,2)}_{i=2}$

$$= \sum_{i=1}^n \sum_{j=1}^n P_{ij} \Sigma_0(j,i) + \sum_{i=1}^n \sum_{j=1}^n P_{ij} \mu_i \mu_j$$ Σ_0 is symmetric

$$= \sum_{i=1}^n [P\Sigma_0]_{ii} + \mu^\top P \mu = \mathrm{Trace}(P\Sigma_0) + \mu^\top P \mu$$

$n=2$: $\begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} P_{11}\Sigma_{11} + P_{12}\Sigma_{21} & * \\ * & P_{21}\Sigma_{12} + P_{22}\Sigma_{22} \end{bmatrix}$

Theorem: S-LQR solution

Proof 2

Remarks

$$\mathbb{E}_x x^\top A x = \text{Trace}(A\Sigma_x) + \mu_x^\top A \mu_x$$

holds for **any** distribution of the random variable x (i.e not necessarily Gaussian), provided that $\mu_x = \mathbb{E}[x]$ and $\Sigma_x = \mathrm{Var}[x]$.

Plan last two lectures

1. **Gradient Descent (GD) for stochastic LQR**
   - S-LQR as an instance of LQG
   - Performance evaluation for given K
   - Computing the gradient of S-LQR
   - Convergence results
2. **Projected GD for locally optimal distributed controllers**
   - Sparsity structures
   - Projected gradient computation
   - Convergence results

Computing the IH Cost... for a given K

$$u_t = -Kx_t \quad \rightarrow \quad \text{compute} \quad J(K) = \mathbb{E}_{x_0}\left[\sum_{t=0}^{\infty} x_t^\top Q x_t + u_t^\top R u_t\right]$$

Why?

- The company you work for (ControlX) deploys a manually tuned controller.
  - e.g. : pole-placement for A-BK.
- The S-LQR controller $\tilde{K}$ could improve performance...but deploying a new controller costs money.

Question

How much Performance Improvement (PI) can you promise?

$$PI = 100\left(1 - \underbrace{\frac{J(\tilde{K})}{J(K)}}_{\leq 1}\right)\%$$

Computing the IH Cost... for a given K

Challenge: how to evaluate $J(K)$?

Naive approach

for $s = 1, \ldots,$ M experiments
  Sample $x_0 \sim N(\mu, \Sigma_0)$
  Simulate the system controlled by $u_t = -Kx_t$, for ∞ timesteps
  Record the value $C^{[s]}_{(K)} = \sum_{t=0}^{\infty} x_t^\top Q x_t + u_t^\top R u_t$
end

Estimate $\boxed{J(K) \approx \frac{1}{M} \sum_{s=1}^{M} C^{[s]}_{(K)}}$

- Cannot simulate for an infinite time;
- Approximation: cannot perform ∞ experiments.

Computing the IH Cost... for a given K

Theorem: computation of $J(K)$

The cost $J(K) = \mathbb{E}_{x_0}\left[\sum_{t=0}^{\infty} x_t^\top Q x_t + u_t^\top R u_t\right]$, where $u_t = -Kx_t$, is given by

$$J(K) = \mathrm{Trace}(P_K \Sigma_0) + \mu^\top P_K \mu$$

where P_K solves a **Lyapunov Equation**

$$P_K = Q + K^\top R K + (A - BK)^\top P_K (A - BK)$$

Remarks

- On Matlab, solve $P_K = \texttt{dlyap}\left((A - BK)^\top, Q + K^\top R K\right)$
- $PI = 100\left(I - \frac{\mathrm{Trace}(P_{K*}\Sigma_0) + \mu^\top P_{K*}\mu}{\mathrm{Trace}(P_K \Sigma_0) + \mu^\top P_K \mu}\right)\%$

Computing the IH Cost... for a given K

Proof of the theorem

We have $x_1 = (A - BK)x_0$, $x_2 = (A - BK)x_1 = (A - BK)^2 x_0$, $\dots$, $x_t = (A - BK)^t x_0$ and $u_t = -K(A - BK)^t x_0$.

Computing the IH Cost... for a given K

Proof of the theorem

We have $x_1 = (A - BK)x_0$, $x_2 = (A - BK)x_1 = (A - BK)^2 x_0$, $\ldots$, $x_t = (A - BK)^t x_0$ and $u_t = -K(A - BK)^t x_0$. Hence:

$$
\begin{aligned}
J(K) &= \mathbb{E}_{x_0}\left[\sum_{t=0}^{\infty} x_0^\top \left([(A - BK)^t]^\top Q (A - BK)^t + [(A - BK)^t]^\top K^\top R K (A - BK)^t\right) x_0\right] \\
&= \mathbb{E}_{x_0}\left[\sum_{t=0}^{\infty} x_0^\top [(A - BK)^t]^\top (Q + K^\top R K)(A - BK)^t x_0\right] \\
&= \mathbb{E}_{x_0}\left[x_0^\top \left[\underbrace{\sum_{t=0}^{\infty} [(A - BK)^t]^\top (Q + K^\top R K)(A - BK)^t}_{\tilde{P}_K}\right] x_0\right] \\
&= \mathrm{Trace}(\tilde{P}_K \Sigma_0) + \mu^\top \tilde{P}_K \mu
\end{aligned}
$$

using the formula seen before for computing $\mathbb{E}_x[x^T A x]$

Computing the IH Cost... for a given K

Proof of the theorem

We have $x_1 = (A - BK)x_0$, $x_2 = (A - BK)x_1 = (A - BK)^2 x_0$, ..., $x_t = (A - BK)^t x_0$ and $u_t = -K(A - BK)^t x_0$. Hence:

$$
\begin{aligned}
J(K) &= \mathbb{E}_{x_0}\left[\sum_{t=0}^{\infty} x_0^\top \left([(A-BK)^t]^\top Q (A-BK)^t + [(A-BK)^t]^\top K^\top R K (A-BK)^t\right) x_0\right] \\
&= \mathbb{E}_{x_0}\left[\sum_{t=0}^{\infty} x_0^\top [(A-BK)^t]^\top (Q + K^\top R K)(A-BK)^t x_0\right] \\
&= \mathbb{E}_{x_0}\left[x_0^\top \left[\underbrace{\sum_{t=0}^{\infty} [(A-BK)^t]^\top (Q + K^\top R K)(A-BK)^t}_{\tilde{P}_K}\right] x_0 \right] \\
&= \mathrm{Trace}(\tilde{P}_K \Sigma_0) + \mu^\top \tilde{P}_K \mu
\end{aligned}
$$

Remains to show that

$\tilde{P}_K = P_K$ where P_K solves $P_K = Q + K^\top R K + (A - BK)^\top P_K (A - BK)$

Prove it at home! Hint: by substitution...

1h13 $\underline{\downarrow}$

Computing the IH Cost... for a given K

Example

$$x_{t+1} = \overset{A=1}{x_t} + \overset{B=1}{u_t} \qquad u_t = -Kx_t \quad K \in \mathbb{R}$$

$$Q = R = \Sigma_0 = 1 \qquad \mu = 0 \quad \mu = \mathbb{E}x_0$$

closed-loop

1) Compute the cost $J(K=1)$ $(x_{t+1} = x_t - x_t = 0)$

$$P_K = Q + K^\top RK + (A - BK)^\top P_K (A - BK)$$

$$P_K = 1 + 1 \cdot 1 \cdot 1 + (\overbrace{1 - 1 \cdot 1}^{0})^\top P_K (1 - 1 \cdot 1)$$

$$\Rightarrow P_K = 2$$

$$J(K=1) = \mathrm{Trace}(P_K \Sigma_0) + \mu^\top P_K \mu$$

$$= \mathrm{Trace}(2 \cdot 1) + 0 = 2$$

Computing the IH Cost... for a given K

Example

$$x_{t+1} = x_t + u_t \qquad u_t = -Kx_t \quad K \in \mathbb{R}$$

$$Q = R = \Sigma_0 = 1 \qquad \mu = 0 \quad \mu = \mathbb{E}[x_0]$$

2) Compute the PI of using the S-LQR controller K^*

$$P = A^\top PA + Q - A^\top PB(R + B^\top PB)^{-1} BPA$$

$$P = P + 1 - P(1+P)^{-1}P$$

Positive root

$$\frac{P^2}{1+P} = 1 \Rightarrow P^2 = 1 + P \Rightarrow P = \frac{\sqrt{5}+1}{2}$$

$$J(K^*) = \text{Trace}(P\Sigma_0) + \mu^\top P\mu = \frac{\sqrt{5}+1}{2} \cong 1.618$$

$$PI = 100\left(1 - \frac{J(K^*)}{J(K)}\right)\% = 100\left(1 - \frac{1.618}{2}\right)\% = 19\%$$

Plan last two lectures

1. **Gradient Descent (GD) for stochastic LQR**
   - S-LQR as an instance of LQG
   - Performance evaluation for given K
   - Computing the gradient of S-LQR
   - Convergence results
2. **Projected GD for locally optimal distributed controllers**
   - Sparsity structures
   - Projected gradient computation
   - Convergence results

Improving K by gradient descent

Gradient Descent (GD)

Given a continuously differentiable $f : \mathbb{R}^n \to \mathbb{R}$, and a sufficiently small step-size $\eta > 0$, the GD algorithm

$$x_{j+1} = x_j - \eta \cdot \underbrace{\nabla f(x_j)}_{\in \mathbb{R}^n, \text{"gradient of f"}}$$

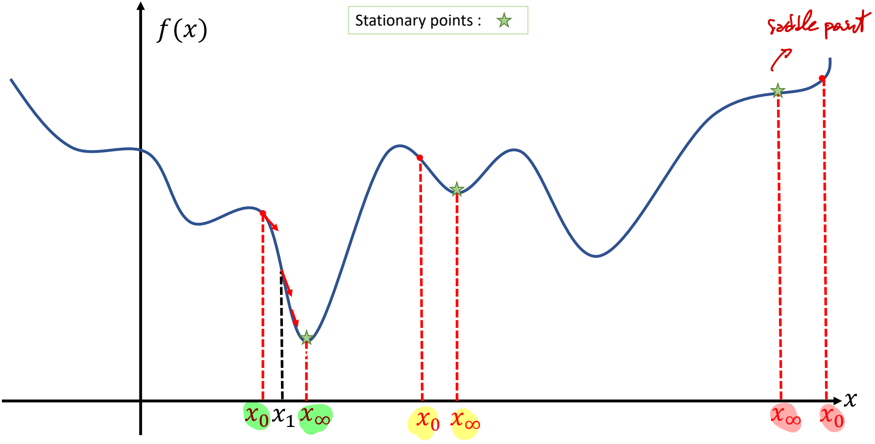
converges to a stationary point of $f(\cdot)$, that is,

$$\lim_{j \to \infty} \nabla f(x_j) = 0$$

Improving K by gradient descent

Remark

In general, convergence to $\nabla f = 0 \nRightarrow$ optimality !



GD for S-LQR

Algorithm

If K_0 not stabilizing
  $J(K_0) = \infty$
INPUT : $K_0 | (A - BK_0)$ is stable, (A, B, Q, R, Σ_0), $\eta > 0$, tolerance $\epsilon > 0$
$K = K_0$
While($\|\nabla J(K)\| > \epsilon$)
  Compute $\nabla J(K)$
  $K \leftarrow K - \eta \nabla J(K)$
end

$J : (\mathbb{R}^m \times \mathbb{R}^n) \to \mathbb{R}$

$$\begin{array}{rc} & \text{column "b"} \\ & \downarrow \\ \nabla J(K) = & \\ \text{row "a"} \rightarrow & \underbrace{\begin{bmatrix} & \vdots & \\ \cdots & \frac{\partial J(K)}{\partial K(a,b)} & \end{bmatrix}}_{m \times n \text{ matrix}} \end{array}$$

element of K in positions (a, b)

GD for S-LQR

Questions

Q1) How do we compute $\nabla J(K)$?

Q2 a) Does the algorithm converge?

Q2 b) If yes (spoiler ...), does $K_j \xrightarrow[j\to\infty]{} K^*$?

Q1) Computing the gradient $\nabla J(K)$

Theorem (gradient evaluation)

Let $x_0 \sim N(0, \Sigma_0)$, and let $J(K)$ denote the cost of the policy $u = -Kx$. Then, we have

$$\nabla J(K) = 2 \underbrace{\left((R + B^\top P_K B)K - B^\top P_K A\right)}_{F_K} \Sigma_K^{CL}$$

where

$$P_K = Q + K^\top RK + (A - BK)^\top P_K (A - BK) \quad (1)$$

$$\Sigma_K^{CL} = \Sigma_0 + (A - BK)\Sigma_K^{CL}(A - BK)^\top \quad (2)$$

Remarks

- First, solve (1) and (2) Lyap. equations. Then, compute $\nabla J(K)$.
- We have seen (1) before. Moreover, one can show that $\mathbb{E}_{x_0}\left[\sum_{t=0}^{\infty} x_t x_t^\top\right]$ is Σ_K^{CL}.
  At home: prove it, i.e. show that $\Sigma_K^{CL} = \mathbb{E}_{x_0}\left[\sum_{t=0}^{\infty} x_t x_t^\top\right]$ is a solution to (2). Hint : similar to P_K computation.

Q1) Computing the gradient $\nabla J(K)$

Example

$A=1 \quad B=1$

$$x_{t+1} = x_t + u_t \qquad u_t = -Kx_t \quad K \in \mathbb{R}$$

$$Q = R = \Sigma_0 = 1 \qquad \mu = 0 \quad \mu = \mathbb{E}[x_0]$$

- compute $\nabla J(K = 1)$
  - We have seen that $P_K = 2$ (solution to (1)). $\mathcal{L}$
  - $\Sigma_K^{CL} = \Sigma_0 + (A - BK)\Sigma_K^{CL}(A - BK)^\top = 1 + (1 - 1)\Sigma_K^{CL}(1 - 1)^\top = 1$
  -

$$\begin{aligned}\nabla J(K) &= 2\left((R + B^\top P_K B)K - B^\top P_K A\right)\Sigma_K^{CL} \\ &= 2\left((1+2)\cdot 1 - 1 \cdot 2 \cdot 1\right) \cdot 1 = 2\end{aligned}$$

Q1) Computing the gradient $\nabla J(K)$

Example

- Since $\nabla J(K=1)=2$, we improve by applying

$$K^1 = K - \eta \nabla J(K) = 1 - 2\eta$$

with $\eta > 0$ small enough.
$\Rightarrow$ we must <u>decrease</u> K to improve the cost.
At home: try to compute $\nabla J(K=0.5)$. Verify that we must <u>increase</u> K to improve the cost.

Plan of the next two lectures

1. **Gradient Descent (GD) for stochastic LQR**
   - The S-LQR problem and its solution
   - Performance evaluation for given K
   - Computing the gradient of S-LQR
   - Convergence results
2. **Projected GD for locally optimal distributed controllers**
   - Sparsity structures
   - Projected gradient descent
   - Convergence results for optimal distributed control

Q2a) Does GD converge?

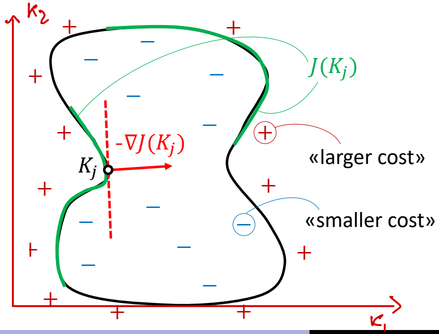
Theorem (full proof not shown)

There exists a constant $L \geq 0$ such that, if the step-size is chosen as $\eta \leq L$, then

$$K_{j+1} = K_j - \eta \nabla J(K_j)$$

converges to $\lim_{j\to\infty} \nabla J(K_j) = 0$.

Intuition: $K = [k_1 \;\; k_2]$



If you move too much in the direction of "$\rightarrow$", you will end up in a region where the cost is larger!

Q2a) Does GD converge?

Arguments used in the theorem proof: how small should η be?

Lemma: GD progress (preliminary result for proving the Theorem)

Suppose that

$$K' = K - \eta \nabla J(K)$$

where

$$\eta \leq \underbrace{\text{poly}(A, B, Q, R, \Sigma_0)}_{\textit{const.}} \cdot \min \left(\frac{1}{J(K) \, \|\nabla J(K)\|}, \frac{1}{J(K)} \right)$$

Then, it holds that

$$J(K') - J(K^*) \leq \underbrace{\underbrace{\left(1 - \eta \cdot \text{poly}_2(R, \overbrace{\Sigma_0}^{>0}, \Sigma_{K^*}^{CL}) \right)}_{<1}}_{1-\alpha} (J(K) - J(K^*))$$

Proof: Lemma 24 of [*"Global Convergence of Policy Gradient Methods for the Linear Quadratic Regulator"*, Maryam Fazel, Rong Ge, Sham M. Kakade, Mehran Mesbahi, 2018].

Q2a) Does GD converge?

$\mathfrak{J}(K') \leq (1-\alpha)\mathfrak{J}(K) + \mathfrak{J}(K^*) - (1-\alpha)\mathfrak{J}(K^*)$
$= (1-\alpha)\mathfrak{J}(K) + \alpha \underbrace{\mathfrak{J}(K^*) \leq (1-\alpha)\mathfrak{J}(K) + \alpha \mathfrak{J}(K)}_{\text{by optimality of } K^*} =$
$= \mathfrak{J}(K)$

Remarks

- $J(K') - J(K^*) \leq (1-\alpha)(J(K) - J(K^*))$, where $\alpha > 0$, implies $J(K') \leq J(K) \Rightarrow$ GD achieves $J(K_{i+1}) \leq J(K_i)$.
- ... however, in the Lemma, η depends on K, which changes in different iterations $\rightarrow \alpha$ in the above inequality also changes.
- *Argument for proving the theorem by leveraging the Lemma*: It can be shown that there is a small enough $L > 0$ *independent of* K such that, for any $\eta \leq L$ the Lemma holds $\rightarrow$ there is $\alpha \in (0, 1)$, independent of K, such that the above inequality holds in all iterations.

2h 13

Mid-lecture summary

1. The problem

Stochastic LQR (S-LQR)

System under control

$$x_{t+1} = Ax_t + Bu_t \qquad x_t \in \mathbb{R}^n \qquad x_0 \sim N(\mu, \Sigma_0),\ \Sigma_0 \geq 0$$

$$x(0) = x_0 \qquad u_t \in \mathbb{R}^m$$

Control cost (I.H.)

$$J = \mathbb{E}_{x_0}\left[\sum_{t=0}^{\infty} x_t^\top Q x_t + u_t^\top R u_t\right]$$

where

$$Q = Q^\top \geq 0$$

$$R = R^\top > 0$$

Goal

Find the optimal sequence $u_0, u_1, \ldots, u_\infty$ that minimizes J.

Mid-lecture summary

2. The solution through the ARE

Theorem: S-LQR solution

Theorem: S-LQR solution

Assume (A, B) is reachable. The optimal control policy for S-LQR is a linear state-feedback

$$u_t = -K^* x_t$$

where $K^* = \left(R + B^\top P B\right)^{-1} B^\top P A$, and P is such that

$$P = A^\top P A + Q - A^\top P B \left(R + B^\top P B\right)^{-1} B^\top P A.$$

The optimal cost is given by

$$J^* = \text{Trace}(P\Sigma_0) + \mu^\top P \underbrace{\mu}_{\mathbb{E}[x_0]}$$

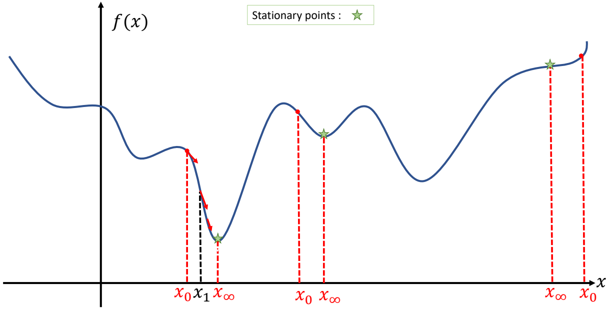
Mid-lecture summary

3. Gradient descent

Improving K by gradient descent

Remark

In general, convergence to $\nabla f = 0 \nRightarrow$ optimality !



Mid-lecture summary

3. Gradient descent

GD for S-LQR

Algorithm

If K_0 not stabilizing
  $J(K_0) = \infty$
INPUT : $K_0 | (A - BK_0)$ is stable, (A, B, Q, R, Σ_0), $\eta > 0$, tolerance $\epsilon > 0$
$K = K_0$
While($\|\nabla J(K)\| > \epsilon$)
  Compute $\nabla J(K)$
  $K \leftarrow K - \eta \nabla J(K)$
end

$J : (\mathbb{R}^m \times \mathbb{R}^n) \rightarrow \mathbb{R}$

$$\underbrace{\begin{array}{cc} & \text{column "b"} \\ & \downarrow \\ \nabla J(K) = & \\ \text{row "a"} \rightarrow & \end{array} \begin{bmatrix} & \vdots & \\ \cdots & \frac{\partial J(K)}{\partial K(a,b)} & \end{bmatrix}}_{m \times n \text{ matrix}}$$

Mid-lecture summary

4. Computation of $J(K)$

Theorem

The cost $J(K) = \mathbb{E}_{x_0}\left[\sum_{t=0}^{\infty} x_t^\top Q x_t + u_t^\top R u_t\right]$, where $u_t = -Kx_t$, is given by

$$J(K) = \mathrm{Trace}(P_K \Sigma_0) + \mu^\top P_K \mu$$

where P_K solves a **Lyapunov Equation**

$$P_K = Q + K^\top R K + (A - BK)^\top P_K (A - BK)$$

Mid-lecture summary

5. Computation of $\nabla J(K)$

Q1) Computing the gradient $\nabla J(K)$

Theorem (gradient evaluation)

Let $x_0 \sim N(0, \Sigma_0)$, and let $J(K)$ denote the cost of the policy $u = -Kx$. Then, we have

$$\nabla J(K) = 2 \underbrace{\left((R + B^\top P_K B)K - B^\top P_K A \right)}_{F_K} \Sigma_K^{CL}$$

where

$$P_K = Q + K^\top R K + (A - BK)^\top P_K (A - BK) \qquad (1)$$

$$\Sigma_K^{CL} = \Sigma_0 + (A - BK)\Sigma_K^{CL}(A - BK)^\top \qquad (2)$$

Remarks

- First, solve (1) and (2) Lyap. equations. Then, compute $\nabla J(K)$.
- We have seen (1) before. Moreover, one can show that $\mathbb{E}_{x_0}\left[\sum_{t=0}^{\infty} x_t x_t^\top\right]$ is Σ_K^{CL}.
  At home: prove it, i.e. show that $\Sigma_K^{CL} = \mathbb{E}_{x_0}\left[\sum_{t=0}^{\infty} x_t x_t^\top\right]$ is a solution to (2). Hint : similar to P_K computation.

Mid-lecture summary

6. Does GD converge?

Theorem

There exists a constant $L \geq 0$ such that, if the step-size is chosen as $\eta \leq L$, then

$$K_{j+1} = K_j - \eta \nabla J(K_j)$$

converges to $\lim_{j\to\infty} \nabla J(K_j) = 0$.

Open question: does $K_j \to K^*$ as $j \to +\infty$?

Q2b) What does GD converge to?

Do we converge to a local or global minimum? $\rightarrow$ Let us check where $\nabla J(K) = 0$, i.e. $\nabla J(K) = 2F_K \Sigma_K^{CL} = 0$.

Q2b) What does GD converge to?

Do we converge to a local or global minimum? $\rightarrow$ Let us check where $\nabla J(K) = 0$, i.e. $\nabla J(K) = 2F_K \Sigma_K^{CL} = 0$.

If $\Sigma_0 > 0$

Theorem: global convergence

The cost function $J(K)$ has a unique stationary point K^* such that $F_{K^*} = \left[(R + B^\top P_{K^*} B)K^* - B^\top P_{K^*} A\right] = 0$.
Hence, if $\eta \leq L$, GD converges to

$$K_\infty = K^* = (R + B^\top P_{K^*} B)^{-1} B^\top P_{K^*} A$$

(the S-LQR solution).

Q2b) What does GD converge to?

Proof of the global convergence theorem

GD converges to K such that $\nabla J(K) = 2F_K \Sigma_K^{CL} = 0$. Either $F_K = (R + B^\top P_K B)K - B^\top P_K A = 0$, or Σ_K^{CL} is not invertible.

$$\Sigma_K^{CL} = \mathbb{E}_{x_0}\left[\sum_{t=0}^{\infty} x_t x_t^\top\right] = \mathbb{E}_{x_0}\left[x_0 x_0^\top\right] + \mathbb{E}_{x_0}\left[\sum_{t=1}^{\infty} x_t x_t^\top\right]$$

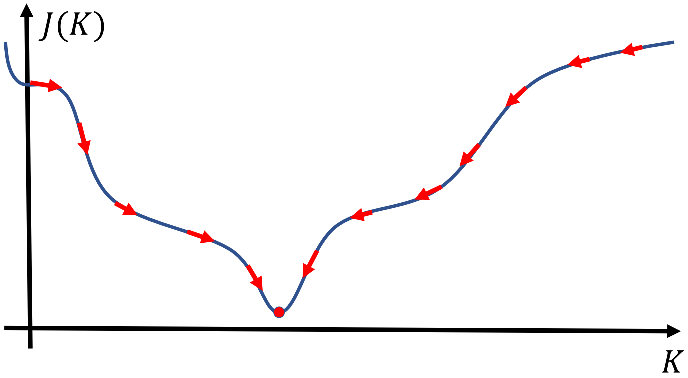
$$= \underbrace{\Sigma_0}_{>0} + \underbrace{\mathbb{E}_{x_0}\left[\sum_{t=1}^{\infty} x_t x_t^\top\right]}_{\geq 0} > 0 \Rightarrow \text{invertible}$$

it can be shown that $F_K = 0 \quad \Rightarrow \quad K = (R + B^\top P_K B)^{-1} B^\top P_K A$

But this K is the optimal S-LQR controller K^*! $\Rightarrow$ if GD converges, it converges to K^*, which is the global optimum.

when P_K solves the corresponding Lyapunov equation (proof at home!)

Q2b) What does GD converge to?



Even if $J(K)$ is non-convex, there is a unique stationary point $\nabla J(K) = 0$, which must be a global minimum.

Summary - Centralized S-LQR

- Given stabilizing K_0 with cost $\text{Trace}(P_{K_0}\Sigma_0) \quad \Rightarrow$ we can improve it through GRADIENT DESCENT

$$K_1 = K_0 - \eta \nabla J(K_0)$$

where $\nabla J(K_0) = 2F_{K_0}\Sigma_{K_0}$.

- η must be small-enough
$\exists L > 0$ such that GD converges with $\eta \leq L$.
  - In practice, decrease η manually.
- By iterating $K_{t+1} = K_t - \eta\nabla J(K_t)$, we converge asymptotically to $K_\infty = K^*$.