Data-driven control design

linear and nonlinear systems

Lecture 2

C. De Persis°, P. Tesi◇

° Institute of Engineering and Technology
University of Groningen
◇ Department of Information Engineering
University of Firenze

university of groningen

UNIVERSITÀ DEGLI STUDI FIRENZE
DINFO
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Recap from Lecture 1

- We saw a method to design controllers directly from data
- Controller design is based on **data-dependent semi-definite programs (SDP)**

▷ Consider a linear system

$$x(k+1) = Ax(k) + Bu(k)$$

with A, B unknown.

We want to design a control law $u = Kx$ that makes <u>$A + BK$ stable</u>

▷ Run an experiment and collect a dataset $\mathbb{D} = \{\bar{x}(k), \bar{u}(k)\}_{k=0}^{T}$ satisfying

$$\underbrace{\begin{bmatrix} \bar{x}(1) & \bar{x}(2) & \dots & \bar{x}(T) \end{bmatrix}}_{X_1} = \\ A \cdot \underbrace{\begin{bmatrix} \bar{x}(0) & \bar{x}(1) & \dots & \bar{x}(T-1) \end{bmatrix}}_{X_0} + B \cdot \underbrace{\begin{bmatrix} \bar{u}(0) & \bar{u}(1) & \dots & \bar{u}(T-1) \end{bmatrix}}_{U_0}$$

In compact form, $X_1 = AX_0 + BU_0$

- For any K, G that solve

$$\begin{bmatrix} K \\ I \end{bmatrix} = \begin{bmatrix} U_0 \\ X_0 \end{bmatrix} G$$

the matrix $A + BK$ can be parametrized via data

$$A + BK = \begin{bmatrix} B & A \end{bmatrix} \begin{bmatrix} K \\ I \end{bmatrix} = \begin{bmatrix} B & A \end{bmatrix} \begin{bmatrix} U_0 \\ X_0 \end{bmatrix} G = X_1 G$$

- This leads to the design program

$$\begin{cases} \text{find} & K, G, P \succ 0 \\ \text{such that} & (X_1 G)^\top P (X_1 G) - P \prec 0 \\ & \begin{bmatrix} K \\ I \end{bmatrix} = \begin{bmatrix} U_0 \\ X_0 \end{bmatrix} G \end{cases}$$

which can be converted to a convex (SDP) form.

- Persistence of excitation plus reachability make sure that the program is feasible.

Outline of this lecture

We will study data-driven control with noisy data

1. the framework
2. why noisy data are problematic
3. a robust control approach
4. practical considerations
5. summary

The framework
(and why noisy data are problematic)

The framework (general)

Consider a linear system

$$\begin{cases} x(k+1) = Ax(k) + Bu(k) + d(k) \\ y(k) = Cx(k) + Du(k) + n(k) \end{cases}$$

where:

- $x \in \mathbb{R}^n$ state, $u \in \mathbb{R}^m$ control, $y \in \mathbb{R}^p$ output
- $d \in \mathbb{R}^n$ <u>unmeasured</u> disturbance
- $n \in \mathbb{R}^p$ <u>unmeasured</u> noise
- A, B, C, D <u>unknown matrices</u>

We want to design a stabilizing controller based on a dataset of (noisy) input-output data collected from the system.

The framework (this lecture)

Consider a linear system

$$x(k+1) = Ax(k) + Bu(k) + d(k)$$

Problem Suppose we perform an experiment on the system and collect the dataset

$$\mathbb{D} := \{\bar{x}(k), \bar{u}(k)\}_{k=0}^{T}$$

of samples which satisfy

$$\bar{x}(k+1) = A\bar{x}(k) + B\bar{u}(k) + \bar{d}(k), \quad k = 0, 1, \ldots, T$$

where $\bar{d}(k)$ are the unmeasured disturbance samples.
Based on $\mathbb{D}$, design a state-feedback controller K that renders $\underline{A + BK \text{ stable}}$

Note $A + BK$ stable implies that $x^+ = (A + BK)x + d$ is input-to-state stable (ISS)

Why noisy data are problematic

System: $x(k+1) = Ax(k) + Bu(k) + d(k)$

The data-based relation for the system now reads:

$$\underbrace{\begin{bmatrix} \bar{x}(1) & \bar{x}(2) & \dots & \bar{x}(T) \end{bmatrix}}_{X_1} =$$

$$A \cdot \underbrace{\begin{bmatrix} \bar{x}(0) & \bar{x}(1) & \dots & \bar{x}(T-1) \end{bmatrix}}_{X_0} + B \cdot \underbrace{\begin{bmatrix} \bar{u}(0) & \bar{u}(1) & \dots & \bar{u}(T-1) \end{bmatrix}}_{U_0}$$

$$+ \underbrace{\begin{bmatrix} \bar{d}(0) & \bar{d}(1) & \dots & \bar{d}(T-1) \end{bmatrix}}_{D_0}$$

In compact form:

$$X_1 = AX_0 + BU_0 + D_0$$

Note The data matrix D_0 is unknown

Data-based relation: $X_1 = \begin{bmatrix} B & A \end{bmatrix} \begin{bmatrix} U_0 \\ X_0 \end{bmatrix} + D_0$

For any K, G that solve

$$\begin{bmatrix} K \\ I \end{bmatrix} = \begin{bmatrix} U_0 \\ X_0 \end{bmatrix} G$$

the matrix $A + BK$ can still be parametrized via data:

$$A + BK = \begin{bmatrix} B & A \end{bmatrix} \begin{bmatrix} K \\ I \end{bmatrix} = \begin{bmatrix} B & A \end{bmatrix} \begin{bmatrix} U_0 \\ X_0 \end{bmatrix} G = (X_1 - D_0)G$$

The Lyapunov condition now reads

$$\underbrace{((X_1 - D_0)G)^\top P(X_1 - D_0)G - P}_{\mathcal{L}(G,P)} \prec 0$$

but is no longer implementable

and ensuring $(X_1G)^\top PX_1G - P \prec 0$ does not ensure $\mathcal{L}(G, P) \prec 0$.

Classic indirect approach (certainty-equivalence design)

Same issues with classic indirect approach

Data-based relation: $X_1 = \begin{bmatrix} B & A \end{bmatrix} \begin{bmatrix} U_0 \\ X_0 \end{bmatrix} + D_0$

The least-squares problem

$$\min_{\mathcal{A}_e, \mathcal{B}_e} \left\| X_1 - \begin{bmatrix} \mathcal{B}_e & \mathcal{A}_e \end{bmatrix} \begin{bmatrix} U_0 \\ X_0 \end{bmatrix} \right\|_F$$

has solution [1]

$$\begin{bmatrix} \mathcal{B}_e & \mathcal{A}_e \end{bmatrix} = X_1 \begin{bmatrix} U_0 \\ X_0 \end{bmatrix}^\dagger$$

$$\Longrightarrow \underbrace{\begin{bmatrix} \mathcal{B}_e & \mathcal{A}_e \end{bmatrix} - \begin{bmatrix} B & A \end{bmatrix}}_{\text{estimation error}} = X_1 \begin{bmatrix} U_0 \\ X_0 \end{bmatrix}^\dagger - (X_1 - D_0) \begin{bmatrix} U_0 \\ X_0 \end{bmatrix}^\dagger = D_0 \begin{bmatrix} U_0 \\ X_0 \end{bmatrix}^\dagger$$

Simply ensuring $\mathcal{A}_e + \mathcal{B}_e K$ stable does not ensure $A + BK$ stable.

[1] Assume $\begin{bmatrix} U_0 \\ X_0 \end{bmatrix}$ full-row rank.

M. Verhaegen, V. Verdult. Filtering and system identification: a least squares approach. Cambridge Univ. Press, 2007.

A robust control approach

A robust approach to control design

For any K, G that solve

$$\begin{bmatrix} K \\ I \end{bmatrix} = \begin{bmatrix} U_0 \\ X_0 \end{bmatrix} G$$

we have $A + BK = (X_1 - D_0)G$.

The Lyapunov condition reads $((X_1 - D_0)G)^\top P(X_1 - D_0)G - P \prec 0$

The matrix D_0 is unknown but finite (has bounded norm).

The idea is to solve

$$((X_1 - D)G)^\top P(X_1 - D)G - P \prec 0 \quad \text{for all } \|D\| \leq \delta$$

for some $\delta > 0$ that we choose. If $\|D_0\| \leq \delta$ then $A + BK$ will be stable

Controller design problem

Define a disturbance set

$$\mathcal{D} := \left\{ D \in \mathbb{R}^{n \times T} : DD^\top \preceq \Delta \quad \text{for known matrix } \Delta \succeq 0 \right\}$$

- This set specifies norm constraints on possible noise matrices
- $\Delta = \delta^2 I_n$ means $DD^\top \preceq \delta^2 I_n$, equiv. $\|D\| = \sqrt{\lambda_{\max}(DD^\top)} \leq \delta$
- Other Δ can be chosen in specific contexts (later on)

Problem (Controller design)

Based on the dataset $\mathbb{D}$,

$$\begin{aligned} \text{find} \quad & K, G, P \succ 0 \\ \text{such that} \quad & ((X_1 - D)G)^\top P (X_1 - D)G - P \prec 0 \quad \text{for all } D \in \mathcal{D} \\ & \begin{bmatrix} K \\ I \end{bmatrix} = \begin{bmatrix} U_0 \\ X_0 \end{bmatrix} G \end{aligned}$$

Problem (Controller design)

Based on the dataset $\mathbb{D}$,

$$\begin{aligned} \text{find} \quad & K, G, P \succ 0 \\ \text{such that} \quad & ((X_1 - D)G)^\top P (X_1 - D)G - P \prec 0 \quad \text{for all } D \in \mathcal{D} \\ & \begin{bmatrix} K \\ I \end{bmatrix} = \begin{bmatrix} U_0 \\ X_0 \end{bmatrix} G \end{aligned}$$

- Our formulation considers **quadratic stabilization** [2]
- Quadratic stabilization is only <u>sufficient</u> for stabilization

---

[2] <u>Definition</u> (Quadratic stability) A system $x(k+1) = A(\theta)x(k)$ with $\theta \in \Theta$ is quadratically stable if there exists $P \succ 0$ such that

$$A(\theta)^\top P A(\theta) - P \prec 0 \quad \text{for all } \theta \in \Theta$$

Petersen's lemma (matrix elimination method)

Lemma[3] Let V, M and N be given matrices of appropriate dimension, and define the set $\mathcal{D} := \{D : DD^\top \preceq \Delta\}$ where Δ is given. Then,

$$V + MD^\top N + N^\top DM^\top \prec 0 \quad \forall D \in \mathcal{D} \tag{1}$$

if and only if there exists $\varepsilon > 0$ such that

$$V + \varepsilon^{-1}MM^\top + \varepsilon N^\top \Delta N \prec 0 \tag{2}$$

(only $\Longleftarrow$) For every $\varepsilon > 0$

$$\begin{aligned} 0 &\preceq \left(\sqrt{\varepsilon^{-1}}M - \sqrt{\varepsilon}N^\top D\right)\left(\sqrt{\varepsilon^{-1}}M - \sqrt{\varepsilon}N^\top D\right)^\top \\ &= \varepsilon^{-1}MM^\top + \varepsilon N^\top DD^\top N - MD^\top N - N^\top DM^\top \\ &\preceq \varepsilon^{-1}MM^\top + \varepsilon N^\top \Delta N - MD^\top N - N^\top DM^\top \end{aligned}$$

Hence

$$MD^\top N + N^\top DM^\top \preceq \varepsilon^{-1}MM^\top + \varepsilon N^\top \Delta N$$

[3] I. Petersen, C. Hollot. A Riccati equation approach to the stabilization of uncertain linear systems, Automatica, 1986

Main result

Theorem Consider the system $x^+ = Ax + Bu + d$ which generates the dataset from which the matrices U_0, X_0, X_1 are obtained. Let $\mathcal{D} = \{D : DD^\top \preceq \Delta\}$, where Δ is chosen by the designer. Suppose there exist $S \in \mathbb{S}^{n\times n}, Y \in \mathbb{R}^{T\times n}$, and $\varepsilon > 0$ such that

$$S = X_0 Y, \quad \begin{bmatrix} S & (X_1 Y)^\top & Y^\top \\ X_1 Y & S - \varepsilon\Delta & 0 \\ Y & 0 & \varepsilon I \end{bmatrix} \succ 0$$

If $D_0 \in \mathcal{D}$ then $K = U_0 Y S^{-1}$ is stabilizing.

Let $G = YS^{-1}$. The two identities $K = U_0 G$ and $I = X_0 G$ imply $\begin{bmatrix} K \\ I \end{bmatrix} = \begin{bmatrix} U_0 \\ X_0 \end{bmatrix} G$

This implies $A + BK = (X_1 - D_0)G$

Since $D_0 \in \mathcal{D}$ by assumption, it is enough that $(X_1 - D)G$ is stable for all $D \in \mathcal{D}$

We only need to prove that the LMI implies stability of $(X_1 - D)G$ for all $D \in \mathcal{D}$

Suppose the LMI holds, i.e.,

$$\left[\begin{array}{cc|c} S & (X_1Y)^\top & Y^\top \\ X_1Y & S-\varepsilon\Delta & 0 \\ \hline Y & 0 & \varepsilon I \end{array}\right] \succ 0$$

A Schur complement gives

$$\begin{bmatrix} S & (X_1Y)^\top \\ X_1Y & S-\varepsilon\Delta \end{bmatrix} - \varepsilon^{-1} \begin{bmatrix} Y^\top \\ 0 \end{bmatrix} \begin{bmatrix} Y & 0 \end{bmatrix} \succ 0 \quad \Longleftrightarrow$$

$$\underbrace{\begin{bmatrix} S & (X_1Y)^\top \\ X_1Y & S \end{bmatrix}}_{-V} - \varepsilon \begin{bmatrix} 0 \\ I \end{bmatrix} \Delta \underbrace{\begin{bmatrix} 0 & I \end{bmatrix}}_{N} - \varepsilon^{-1} \underbrace{\begin{bmatrix} Y^\top \\ 0 \end{bmatrix}}_{M} \begin{bmatrix} Y & 0 \end{bmatrix} \succ 0$$

By **Petersen's Lemma**, $V + \varepsilon^{-1}MM^\top + \varepsilon N^\top \Delta N \prec 0$ implies $V + MD^\top N + N^\top D M^\top \prec 0$ for all $D \in \mathcal{D} = \{DD^\top \preceq \Delta\}$

$$\begin{bmatrix} S & (X_1Y)^\top \\ X_1Y & S \end{bmatrix} - \begin{bmatrix} Y^\top \\ 0 \end{bmatrix} D^\top \begin{bmatrix} 0 & I \end{bmatrix} - \begin{bmatrix} 0 \\ I \end{bmatrix} D \begin{bmatrix} Y & 0 \end{bmatrix} \succ 0 \quad \forall D \in \mathcal{D}$$

Condition

$$\begin{bmatrix} S & (X_1Y)^\top \\ X_1Y & S \end{bmatrix} - \begin{bmatrix} Y^\top \\ 0 \end{bmatrix} D^\top \begin{bmatrix} 0 & I \end{bmatrix} - \begin{bmatrix} 0 \\ I \end{bmatrix} D \begin{bmatrix} Y & 0 \end{bmatrix} \succ 0 \quad \forall D \in \mathcal{D}$$

is equivalent to

$$\begin{bmatrix} S & ((X_1 - D)Y)^\top \\ (X_1 - D)Y & S \end{bmatrix} \succ 0 \quad \forall D \in \mathcal{D}$$

Another Schur complement gives

$$((X_1 - D)Y)^\top S^{-1}(X_1 - D)Y - S \prec 0 \quad \forall D \in \mathcal{D}$$

Pre- and post-multiplying for S^{-1} gives

$$S^{-1}((X_1 - D)Y)^\top S^{-1}(X_1 - D)YS^{-1} - S^{-1} \prec 0 \quad \forall D \in \mathcal{D}$$

The change of variable $G = YS^{-1}$ gives

$$((X_1 - D)G)^\top S^{-1}(X_1 - D)G - S^{-1} \prec 0 \quad \forall D \in \mathcal{D}$$

which is the Lyapunov condition with $P = S^{-1}$

Remarks

- Provides stability guarantees despite noisy data
- It only requires a finite number of data collected in low sample-complexity experiment(s)
- As simple as the baseline solution: data-dependent SDP
- The data-based problem is solvable via efficient numerical algorithms (cvx)
- No statistics for the disturbance is needed

Example

Consider a randomly generated system

$$A = \begin{bmatrix} -0.3245 & -0.5548 & -0.2793 \\ 0.5906 & -0.4228 & 0.0892 \\ -0.3792 & -0.2863 & -0.0984 \end{bmatrix}, \quad B = \begin{bmatrix} 0.5864 \\ -0.8519 \\ 0.8003 \end{bmatrix}$$

We collect $T = 100$ samples generated with $|u| \leq 1$ and $\|d\| \leq 0.1$. We select $\Delta = I_3$, meaning that $\mathcal{D} = \left\{D : DD^\top \preceq I_3\right\}$. We have

$$D_0 D_0^\top = \begin{bmatrix} d(0) & d(1) & \cdots & d(99) \end{bmatrix} \begin{bmatrix} d(0)^\top \\ d(1)^\top \\ \vdots \\ d(99)^\top \end{bmatrix} \preceq I_3$$

Thus $D_0 \in \mathcal{D}$. The SDP is feasible and returns

$$K = \begin{bmatrix} 0.5980 & 0.0325 & 0.1748 \end{bmatrix}$$

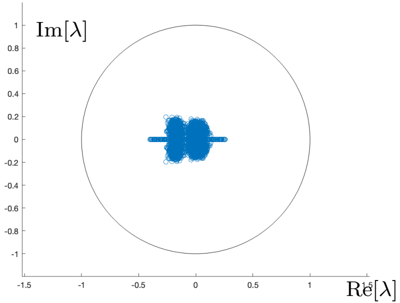
Since $D_0 \in \mathcal{D}$ the controller is stabilizing.

Closed-loop eigenvalues of $A + BK = (X_1 - D_0)G$:

$$\lambda = \begin{bmatrix} -0.3356 \\ -0.0400 \\ -0.0072 \end{bmatrix}$$

Figure Closed-loop eigenvalues of $(X_1 - D)G$ as we vary $D \in \mathcal{D}$

Eigenvalues are far from the border, indicating there is still robustness margin.
Indeed, the design program remains feasible for larger noise values (feasibility is discussed next)



Selected literature

Selected literature

This result appeared in

- C. De Persis, P. Tesi, "Formulas for data-driven control: Stabilization, optimality, and robustness", IEEE TAC, 2019

Other approaches have been proposed using <u>different robust control tools</u>, sometimes considering a purely stochastic setting:

- H. Mania, N. Matni, B. Recht, S. Tu. "On the sample complexity of the linear quadratic regulator". Foundations of Computational Mathematics, 2019
- M. Ferizbegovic, J. Umenberger, H. Hjalmarsson, T. Schön. Learning robust LQ-controllers using application oriented exploration. IEEE L-CSS, 2019
- H. van Waarde, M. Camlibel, M. Mesbahi, "From noisy data to feedback controllers: Nonconservative design via a matrix S-lemma", IEEE TAC, 2020
- J. Berberich, A. Koch, C. Scherer, F. Allgöwer, "Robust data-driven state-feedback design", IEEE American Control Conference (ACC), 2020
- A. Bisoffi, C. De Persis, P. Tesi, "Data-driven control via Petersen's lemma", Automatica, 2022

Some of these approaches are <u>indirect</u>

The indirect approach

Indirect approaches involve explicit sys–ID.

In practice, the uncertainty on D_0 is "projected" onto the system matrices. The idea is to determine all possible systems that are compatible with the data according to the bounds on the noise:

$$\mathcal{C} := \{(\mathcal{A}, \mathcal{B}) : X_1 = \mathcal{A}X_0 + \mathcal{B}U_0 + D, \text{ with } DD^\top \preceq \Delta\}$$

Called **consistency set**, **confidence region**,...

Assuming $W_0 := \begin{bmatrix} U_0 \\ X_0 \end{bmatrix}$ full-row rank, one can explicitly characterize the consistency set:

$$\mathcal{C} = \left\{(\mathcal{A}, \mathcal{B}) : \left(\begin{bmatrix}\mathcal{B} & \mathcal{A}\end{bmatrix} - X_1 W_0^\dagger\right) \Theta \left(\begin{bmatrix}\mathcal{B} & \mathcal{A}\end{bmatrix} - X_1 W_0^\dagger\right)^\top \preceq L\right\}$$

for suitable matrices Θ, L (next slide)

The consistency set has **ellipsoidal form**:

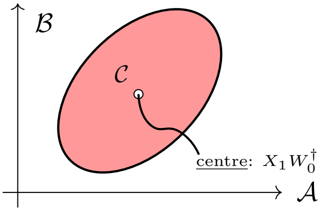
$$\mathcal{C} = \left\{ (\mathcal{A}, \mathcal{B}) : \left(\begin{bmatrix} \mathcal{B} & \mathcal{A} \end{bmatrix} - X_1 W_0^\dagger \right) \Theta \left(\begin{bmatrix} \mathcal{B} & \mathcal{A} \end{bmatrix} - X_1 W_0^\dagger \right)^\top \preceq L \right\}$$

where:

- $\Theta := W_0 W_0^\top$
- $L := H \Theta^{-1} H^\top + \Delta - X_1 X_1^\top$
- $H := X_1 W_0^\top$

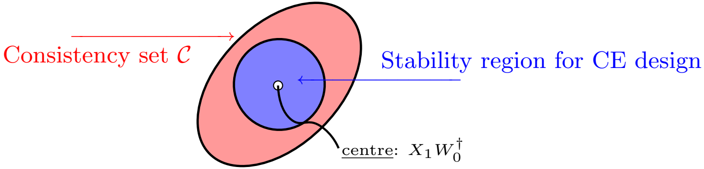
Size and orientation of $\mathcal{C}$ depend on data and Δ

consistency set in scalar case



Stability requires to find a controller that stabilizes the whole set $\mathcal{C}$. Since $\mathcal{C}$ is convex, also this formulation leads to convex design programs (H.J. van Waarde et al., IEEE TAC 2020; A. Bisoffi et al., Automatica 2022)

The centre of the set $\mathcal{C}$ is the least-squares estimate $\begin{bmatrix} \mathcal{B}_e & \mathcal{A}_e \end{bmatrix} = X_1 W_0^\dagger$. We visualize why certainty equivalence can fail if not accompanied by robust controller design.



This also explains why certainty equivalence is popular: by stabilizing the LS estimate we stabilize at least some portion of $\mathcal{C}$. In fact, modern LS–based approaches (S. Dean et al., FCM 2019, M. Ferizbegovic et al., IEEE L-CSS 2019) try to build a controller around the LS estimate which is robust enough to cover $\mathcal{C}$.

Practical considerations

Practical considerations

Recall the main result:

Theorem Consider the system $x^+ = Ax + Bu + d$ which generates the dataset from which the matrices U_0, X_0, X_1 are obtained. Let $\mathcal{D} = \{D : DD^\top \preceq \Delta\}$, where Δ is chosen by the designer. Suppose there exist $S \in \mathbb{S}^{n\times n}, Y \in \mathbb{R}^{T\times n}$, and $\varepsilon > 0$ such that

$$S = X_0 Y, \quad \begin{bmatrix} S & (X_1Y)^\top & Y^\top \\ X_1Y & S - \varepsilon\Delta & 0 \\ Y & 0 & \varepsilon I \end{bmatrix} \succ 0$$

If $D_0 \in \mathcal{D}$ then $K = U_0 Y S^{-1}$ is stabilizing.

We need the problem to be feasible and we need $D_0 \in \mathcal{D}$.

How to choose the set $\mathcal{D}$

The choice of $\mathcal{D} = \{D : DD^\top \preceq \Delta\}$ is context-dependent (no universal recipe)

The choice of $\mathcal{D}$ can involve:

- **Point-wise worst-case bounds**. If we know that $\|d\| \leq \gamma$ point-wise then

$$D_0 D_0^\top = \begin{bmatrix} d(0) & d(1) & \cdots & d(T-1) \end{bmatrix} \begin{bmatrix} d(0)^\top \\ d(1)^\top \\ \vdots \\ d(T-1)^\top \end{bmatrix} \preceq \underbrace{\gamma^2 \cdot T I_n}_{\Delta}$$

  See also Bisoffi et al., "Trade-offs in learning controllers from noisy data", SCL 2021

- **Signal-to-noise ratio conditions**. E.g., $\Delta := \gamma^2 \cdot U_0 U_0^\top$
- **Statistical properties of the noise**

Exploiting statistical properties

If available, statistical properties of noise permit to relax worst-case bounds as long as we accept bounds in probability

Lemma (Almost sure boundedness)[a]

Identically distributed = samples are taken from the same probability distribution
Independent = samples are all independent events: $P(X|Y) = P(X)$ and $P(Y|X) = P(Y)$

Assume $d \in \mathbb{R}^n$ are i.i.d. zero-mean random vectors with covariance matrix Σ and such that $\|d\| \leq \gamma$ almost surely. For all $\mu > 0$,

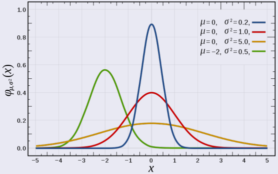
$$D_0 D_0^\top \preceq \underbrace{T\left(\|\Sigma\| + \mu\right) I_n}_{\Delta}$$

with probability at least $1 - 2n \exp\left(-\frac{T\mu^2}{2\gamma^2(\|\Sigma\|+\mu)}\right)$.

(Indeed, $Q_T := \frac{1}{T} D_0 D_0^\top$ is the MLE of the covariance matrix based on T samples)

[a] M. Wainwright. High-dimensional statistics: A non-asymptotic viewpoint. Cambridge University Press, 2019

Lemma (Gaussian noise)[a]
Assume $d \in \mathbb{R}^n$ are i.i.d. random vectors drawn from $\mathcal{N}(0, \Sigma)$.
Then, for all $\mu > 0$,



source: Wikipedia

$$D_0 D_0^\top \preceq \underbrace{T \left(\lambda_{\max}(\Sigma^{1/2})(1+\mu) + \sqrt{\frac{\text{trace}(\Sigma)}{T}} \right)^2 I_n}_{\Delta}$$

with probability at least $1 - \exp(-T\mu^2/2)$.

[a] M. Wainwright. High-dimensional statistics: A non-asymptotic viewpoint. Cambridge University Press, 2019

Results in probability

Theorem Consider the system $x^+ = Ax + Bu + d$ which generates the dataset from which the matrices U_0, X_0, X_1 are obtained. Let $\mathcal{D} = \{D : DD^\top \preceq \Delta\}$, where Δ is chosen by the designer. Suppose there exist $S \in \mathbb{S}^{n\times n}, Y \in \mathbb{R}^{T\times n}$, and $\varepsilon > 0$ such that

$$S = X_0 Y, \qquad \begin{bmatrix} S & (X_1 Y)^\top & Y^\top \\ X_1 Y & S - \varepsilon\Delta & 0 \\ Y & 0 & \varepsilon I \end{bmatrix} \succ 0$$

If $D_0 \in \mathcal{D}$ with probability at least p
then $K = U_0 Y S^{-1}$ is stabilizing with probability at least p.

Remarkably, **distribution–free!**
(more difficult with indirect approaches)

Let $\mathcal{E}$ be an event and let X, Y be two mutually exclusive and collectively exhaustive events, i.e.,

$$P(X \cap Y) = 0 \quad \text{and} \quad P(X \cup Y) = 1$$

By the **law of total probability**,

$$P(\mathcal{E}) = \underbrace{P(\mathcal{E}|X)}_{\text{conditional prob}} \cdot P(X) + P(\mathcal{E}|Y) \cdot P(Y)$$

Applied to our case, suppose that the design program is feasible and let $\mathcal{E}$ denote the event "K is stabilizing". Then,

$$\begin{aligned} P(\mathcal{E}) &= P(\mathcal{E}|D_0 \in \mathcal{D}) \cdot P(D_0 \in \mathcal{D}) + P(\mathcal{E}|D_0 \notin \mathcal{D}) \cdot P(D_0 \notin \mathcal{D}) \\ &\geq \underbrace{P(\mathcal{E}|D_0 \in \mathcal{D})}_{=1} \cdot P(D_0 \in \mathcal{D}) \\ &= P(D_0 \in \mathcal{D}) \end{aligned}$$

Example (cont'd)

Consider again previous system where

$$A = \begin{bmatrix} -0.3245 & -0.5548 & -0.2793 \\ 0.5906 & -0.4228 & 0.0892 \\ -0.3792 & -0.2863 & -0.0984 \end{bmatrix}, \; B = \begin{bmatrix} 0.5864 \\ -0.8519 \\ 0.8003 \end{bmatrix}$$

We collect $T = 100$ samples generated with $|u| \leq 1$ and $d \sim \mathcal{N}(0, 0.01 I_3)$ i.i.d.. With $\mu = 0.4$ we obtain $D_0 D_0^\top \preceq 2.4750 I_3$ with probability at least $\text{p} = 99.97\%$. Hence, we choose $\Delta = 2.4750 I_3$. The design program is feasible and we get

$$K = \begin{bmatrix} 0.5933 & -0.0905 & 0.1460 \end{bmatrix}$$

This controller ensures closed-loop stability with 99.97% probability.

It is indeed stabilizing as $\|D_0\|^2 = 0.9945$.

Feasibility of the design program

The last point regards feasibility. Recall the main result (or the version using probability):

Theorem Consider the system $x^+ = Ax + Bu + d$ which generates the dataset from which the matrices U_0, X_0, X_1 are obtained. Let $\mathcal{D} = \{D : DD^\top \preceq \Delta\}$, where Δ is chosen by the designer. Suppose there exist $S \in \mathbb{S}^{n \times n}, Y \in \mathbb{R}^{T \times n}$, and $\varepsilon > 0$ such that

$$S = X_0 Y, \quad \begin{bmatrix} S & (X_1 Y)^\top & Y^\top \\ X_1 Y & S - \varepsilon\Delta & 0 \\ Y & 0 & \varepsilon I \end{bmatrix} \succ 0$$

If $D_0 \in \mathcal{D}$ then $K = U_0 Y S^{-1}$ is stabilizing.

Under what conditions the design program is feasible?

Can we render the problem feasible? Is the quantity of data important?

Two auxiliary facts

1) Feasibility of

$$S = X_0 Y, \quad \begin{bmatrix} S & (X_1 Y)^\top & Y^\top \\ \star & S - \varepsilon\Delta & 0 \\ \star & \star & \varepsilon I \end{bmatrix} \succ 0$$

with decision variables S, Y, ε is equivalent to the feasibility of

$$\begin{bmatrix} X_0 Y & (X_1 Y)^\top & Y^\top \\ \star & X_0 Y - \varepsilon\Delta & 0 \\ \star & \star & \varepsilon I \end{bmatrix} \succ 0$$

with decision variables Y, ε.

2) Given three symmetric matrices A, B, C with

$$A = B + C, \quad B \succ 0$$

then $A \succ 0$ if $\|C\| < \lambda_{\min}(B)$.

Considerations about feasibility

Decompose the system response into noise-free and noisy:

$$\begin{cases} x^f(k+1) = Ax^f(k) + Bu(k) \\ x^f(0) = x(0) \end{cases} \qquad \begin{cases} x^n(k+1) = Ax^n(k) + d(k) \\ x^n(0) = 0 \end{cases}$$

Define data matrices $X_0^f, X_1^f, X_0^n, X_1^n$ which satisfy the identities $X_0 = X_0^f + X_0^n$, $X_1 = X_1^f + X_1^n$

Decompose the LMI:

$$\underbrace{\begin{bmatrix} X_0 Y & (X_1 Y)^\top & Y^\top \\ \star & X_0 Y - \varepsilon\Delta & 0 \\ \star & \star & \varepsilon I \end{bmatrix}}_{M} = \underbrace{\left[\begin{array}{cc|c} X_0^f Y & (X_1^f Y)^\top & Y^\top \\ \star & X_0^f Y & 0 \\ \hline \star & \star & \varepsilon I \end{array}\right]}_{\text{noise-free term } F} + \underbrace{\begin{bmatrix} X_0^n Y & (X_1^n Y)^\top & 0 \\ \star & X_0^n Y - \varepsilon\Delta & 0 \\ \star & \star & 0 \end{bmatrix}}_{\text{noisy term } N}$$

If the problem with noise-free data is feasible then there exist $Y, \varepsilon > 0$ such that $F \succ 0$.[4]

[4] Recall If $C \succ 0$ then $\left[\begin{array}{c|c} A & B \\ \hline \star & C \end{array}\right] \succ 0 \iff A - BC^{-1}B^\top \succ 0$

Decompose the LMI:

$$\underbrace{\begin{bmatrix} X_0 Y & (X_1 Y)^\top & Y^\top \\ \star & X_0 Y - \varepsilon\Delta & 0 \\ \star & \star & \varepsilon I \end{bmatrix}}_{M} = \underbrace{\begin{bmatrix} X_0^f Y & (X_1^f Y)^\top & Y^\top \\ \star & X_0^f Y & 0 \\ \star & \star & \varepsilon I \end{bmatrix}}_{\text{noise-free term } F} + \underbrace{\begin{bmatrix} X_0^n Y & (X_1^n Y)^\top & 0 \\ \star & X_0^n Y - \varepsilon\Delta & 0 \\ \star & \star & 0 \end{bmatrix}}_{\text{noisy term } N}$$

Assume that the problem with noise-free data is feasible and let $Y, \varepsilon > 0$ such that $F \succ 0$.

Since $F \succ 0$, if

$$\|N\| < \lambda_{\min}(F) \qquad (\sim)$$

then $M \succ 0$. Condition $(\sim)$ can be equivalently expressed as

$$\max\{\|D_0\|, \|\Delta\|\} < c\lambda_{\min}(F) \qquad (\star)$$

for some constant c independent of the noise. To have $(\star)$ fulfilled:

(1) Data quality is more important than quantity

(2) Also the quantity of data can help

Averaging data

Fact (Informal) Consider a zero-mean uncorrelated signal d with variance $E[d^2] = \sigma^2$. Suppose we take N realizations of d. The variance of the average of the realizations is

$$
\begin{aligned}
E\left[\left(\frac{1}{N}\sum_{k=1}^{N} d_k\right)^2\right] &= \frac{1}{N^2}\cdot E\left[\left(\sum_{k=1}^{N} d_k\right)^2\right] \\
&= \frac{1}{N^2}\cdot E\left[(d_1 + d_2 + \ldots + d_N)(d_1 + d_2 + \ldots + d_N)\right] \\
&= \text{uncorrelation implies } E[d_k d_j] = 0 \\
&= \frac{1}{N^2}\cdot E\left[d_1^2 + d_2^2 + \ldots + d_N^2\right] \\
&= \frac{1}{N^2}\cdot (N\cdot\sigma^2) = \frac{\sigma^2}{N}
\end{aligned}
$$

meaning that averaging reduces the variance by a factor of N

Applying this idea to our case, suppose we perform N **experiments**.

Each experiment returns the dataset $(U_{0k}, X_{0k}, X_{1k}, D_{0k})$, $k = 1, \ldots, N$, D_{0k} unknown.

Denote by $(\underline{U}_0, \underline{X}_0, \underline{X}_1, \underline{D}_0)$ the **average dataset**.

We have

$$\begin{aligned} \underline{X}_1 &= \frac{1}{N} \sum_{k=1}^{N} X_{1k} \\ &= \frac{1}{N} \sum_{k=1}^{N} (AX_{0k} + BU_{0k} + D_{0k}) \\ &= A\underline{X}_0 + B\underline{U}_0 + \underline{D}_0 \end{aligned}$$

The average signals still provide a valid trajectory but now $\underline{D}_0$ may have reduced variance, and hence smaller norm. In turn, feasibility becomes easier. This is true under the same (multivariable) conditions as in previous Fact.

Example with Gaussian distribution:

Lemma (Gaussian noise)
Assume $d \in \mathbb{R}^n$ are i.i.d. random vectors drawn from $\mathcal{N}(0, \Sigma)$.
Then, for all $\mu > 0$,

$$\underline{D}_0 \underline{D}_0^\top \preceq \underbrace{\frac{T}{N} \left(\lambda_{\max}(\Sigma^{1/2})(1+\mu) + \sqrt{\frac{\operatorname{trace}(\Sigma)}{T}} \right)^2 I_n}_{\Delta}$$

with probability at least $1 - \exp(-T\mu^2/2)$.

A result on sample–complexity

Theorem Consider a reachable system $x^+ = Ax + Bu + d$ where $d \overset{\text{iid}}{\sim} \mathcal{N}(0, \Sigma)$. Suppose we perform N experiments with the same persistently exciting input u, and determine the average datasets $\underline{U}_0, \underline{X}_0, \underline{X}_1$. (Applying the same input u ensures that the average input is persistently exciting and the feasibility of the noise-free problem.)

Pick any $\mu > 0$ and let

$$\Delta := \frac{T}{N} \left(\lambda_{\max}(\Sigma^{1/2})(1+\mu) + \sqrt{\frac{\text{trace}(\Sigma)}{T}} \right)^2 I_n$$

Consider the design program in the decision variables $S, Y, \varepsilon > 0$:

$$S = \underline{X}_0 Y, \quad \begin{bmatrix} S & (\underline{X}_1 Y)^\top & Y^\top \\ \underline{X}_1 Y & S - \varepsilon\Delta & 0 \\ Y & 0 & \varepsilon I \end{bmatrix} \succ 0$$

If N is sufficiently large then the program is feasible and $K = \underline{U}_0 Y S^{-1}$ is stabilizing with probability at least $1 - \exp(-T\mu^2/2)$.

Example (cont'd)

Consider the same system as before

$$A = \begin{bmatrix} -0.3245 & -0.5548 & -0.2793 \\ 0.5906 & -0.4228 & 0.0892 \\ -0.3792 & -0.2863 & -0.0984 \end{bmatrix}, \quad B = \begin{bmatrix} 0.5864 \\ -0.8519 \\ 0.8003 \end{bmatrix}$$

We take again $|u| \leq 1$ and $T = 100$, but this time we assume $d \sim \mathcal{N}(0, 0.3 I_3)$ i.i.d...

For single experiment, the design program is infeasible ($\|D_0\|^2 \approx 34$)

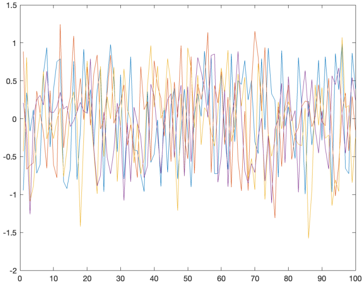
Next, we considera $N = 100$ repeated experiments (same u) and we obtain $\underline{D}_0 \underline{D}_0^\top \preceq 0.7425 I_3$ with probability at least p $= 99.97\%$. Hence, we choose $\Delta = 0.7425 I_3$. This time the design program is feasible and we get

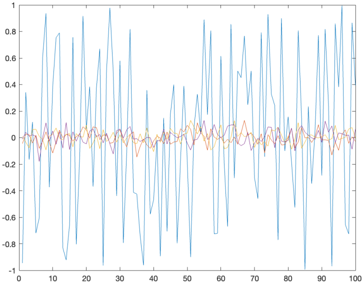
$$K = \begin{bmatrix} -0.1604 & -0.1705 & -0.3730 \end{bmatrix}$$

The controller ensures closed-loop stability with 99.97% probability.
It is indeed stabilizing as $\|\underline{D}_0\|^2 = 0.3437$. which is approximately $\frac{1}{N}\|D_0\|^2$.

(Left) Input and disturbance signals for one of the experiments

(Right) Input and disturbance signals after averaging $N = 100$ experiments





Summary and research topics

- We saw a method for designing controllers from noisy data
- Statistics–free but amenable to sample-complexity analysis
- Computationally effective (convex programming)
- Offers stability guarantees and is interpretable
- We saw stabilization problems and state-feedback design
- Important extensions exist: I/O data and dynamic controllers, input and state constraints, network implementation (including resource-aware control), sparse design,...

  Some of these extensions are very preliminary
- Optimal control not well understood (C. De Persis, P. Tesi, "Low-complexity learning of linear quadratic regulators from noisy data", Automatica 2021)

Additional material

Measurement noise

Measurement noise

Consider the following setting:

$$x(k+1) = Ax(k) + Bu(k), \quad y(k) = x(k) + n(k)$$

The relation between data and dynamics now reads:

$$X_1 = AX_0 + BU_0, \quad Y_0 = X_0 + N_0, \quad Y_1 = X_1 + N_1$$

This gives

$$\begin{aligned} Y_1 &= X_1 + N_1 \\ &= AX_0 + BU_0 + N_1 \\ &= AY_0 + BU_0 + \underbrace{N_1 - AN_0}_{Q_0 = Q_0(N_0, N_1, A)} \end{aligned}$$

Same structure as before

Let

$$\mathcal{Q} := \left\{Q :\ QQ^\top \preceq \Delta\right\}$$

represent a bound on $Q_0 = N_1 - AN_0$. This set specifies how much the noise may deviate from a genuine system trajectory $n^+ = An$.

Theorem Consider a system $x^+ = Ax + Bu$, $y = x + n$ which generates the dataset from which the matrices U_0, X_0, X_1 are obtained. Let $\mathcal{Q} = \left\{Q :\ QQ^\top \preceq \Delta\right\}$, where Δ is chosen by the designer. Suppose there exist $S \in \mathbb{S}^{n\times n}, Z \in \mathbb{R}^{T\times n}$, and $\varepsilon > 0$ such that

$$S = Y_0 Z, \qquad \begin{bmatrix} S & (Y_1 Z)^\top & Z^\top \\ Y_1 Z & S - \varepsilon\Delta & 0 \\ Z & 0 & \varepsilon I \end{bmatrix} \succ 0$$

If $Q_0 \in \mathcal{D}$ then $K = U_0 Z S^{-1}$ is stabilizing.

Petersen's lemma

Petersen's lemma

Lemma Let V, M and N be given matrices of appropriate dimension, and define the set $\mathcal{D} := \{D : DD^\top \preceq \Delta\}$ where Δ is given. Then,

$$V + MD^\top N + N^\top DM^\top \prec 0 \quad \forall D \in \mathcal{D}$$

if and only if there exists $\varepsilon > 0$ such that

$$V + \varepsilon^{-1}MM^\top + \varepsilon N^\top \Delta N \prec 0$$

Necessity is the difficult part, we give some intuitions considering the scalar case

We want to show that

$$V + MD^\top N + N^\top DM^\top \prec 0 \quad \text{for all } D : DD^\top \preceq \Delta$$
$$\implies \quad \exists\varepsilon > 0 : V + \varepsilon^{-1}MM^\top + \varepsilon N^\top \Delta N \prec 0$$

Assume $M, N, \Delta \neq 0$. By hypothesis

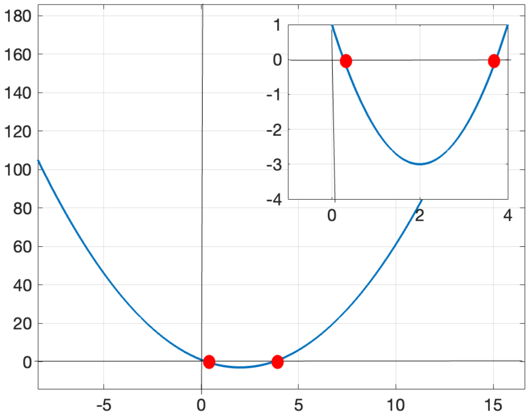
$$V + 2\sqrt{\Delta}|M||N| < 0$$

We also have $V - 2\sqrt{\Delta}|M||N| < 0$, so

$$V^2 - 4\Delta M^2 N^2 > 0$$

This can be viewed as the discriminant of the second-order polynomial $\Delta N^2 \lambda^2 + V\lambda + M^2$, which has two positive roots $\lambda_\pm$, with $\lambda_+ > \lambda_-$ (note that $\Delta N^2 > 0$, $V < 0$, $M^2 > 0$).

The polynomial $\Delta N^2 \lambda^2 + V\lambda + M^2$ has two positive roots $\lambda_+ > \lambda_-$, so there exists a value $\varepsilon \in (\lambda_-, \lambda_+)$ such that $\Delta N^2 \varepsilon^2 + V\varepsilon + M^2 < 0$. This is equivalent to $V + \varepsilon^{-1} M^2 + \varepsilon \Delta N^2 < 0$.



Continuous-time results

Continuous-time results

Discrete-time results can be extended to continuous-time systems with X_1 matrix of state derivatives (adding noise if necessary).

Theorem Consider the system $\dot{x} = Ax + Bu + d$ which generates the dataset from which the matrices U_0, X_0, X_1 are obtained. Let $\mathcal{D} = \left\{D : DD^\top \preceq \Delta\right\}$, where Δ is chosen by the designer. Suppose there exist $S \in \mathbb{S}^{n\times n}, Y \in \mathbb{R}^{T\times n}$, and $\varepsilon > 0$ such that

$$S = X_0 Y, \quad \begin{bmatrix} (X_1 Y) + (X_1 Y)^\top + \varepsilon\Delta & Y^\top \\ Y & -\varepsilon I \end{bmatrix} \prec 0$$

If $D_0 \in \mathcal{D}$ then $K = U_0 Y S^{-1}$ is stabilizing.