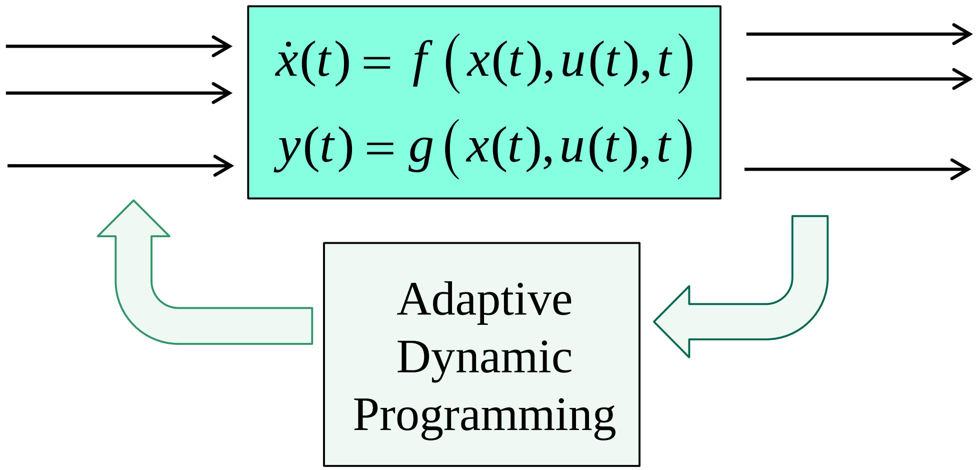
# מערכות מכטרוניות

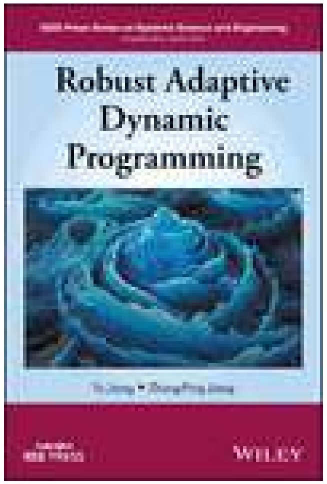
## מבוא לבקרה אופטימלית - חלק 2

## Introduction to Optimal Control




**מרצה: שי ארוגטי**

# בקר משוב מצב אופטימלי




מצגת זו מתבססת (עם תוספות ממקורות אחרים) על פרקים 1, 2, 3 בספר,

Robust Adaptive Dynamic Programming, First Edition
By, Yu Jiang Zhong-Ping Jiang, Wiley, 2017.

CHAPTER 1 - Introduction
CHAPTER 2 - Adaptive Dynamic Programming for Uncertain Linear Systems
CHAPTER 3 - Semi-Global Adaptive Dynamic Programming

מומלץ גם לקרוא את המאמרים (מציגים חומר דומה),

Y. Jiang and Z. P. Jiang. “Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics,” Automatica, 48(10), 2012.

Y. Jiang and Z. P. Jiang. Robust adaptive dynamic programming and feedback stabilization of nonlinear systems. IEEE Transactions on Neural Networks and Learning Systems, 25(5), 2014.

## בקר משוב מצב אופטימלי

פיתוח הבקר האופטימלי הסטציונרי באופן ישיר

תחילה נוכיח את המשפט הבא המשווה בין פתרונות של שתי משוואות ריקטי,

(**Comparison Theorem**)

נניח,
$$A^T P_1 + P_1 A - P_1 \bar{R} P_1 + \bar{Q}_1 = 0$$

$$A^T P_2 + P_2 A - P_2 \bar{R} P_2 + \bar{Q}_2 = 0$$

כאשר $\bar{Q}_2 \geq \bar{Q}_1 \geq 0$

ברור כי עבור $\bar{Q}_2 \neq \bar{Q}_1$ יתקבלו פתרונות שונים, $P_1 = P_1^T, \quad P_2 = P_2^T, \quad P_1 \neq P_2$

בנוסף נניח כי, $\bar{R} \geq 0$ ו- $A - \bar{R} P_2$ היא מטריצה יציבה.

נזכיר כי במקרה של הבקר האופטימלי, $\bar{R} = B R^{-1} B^T$

לכן, $A - \bar{R} P = A - B R^{-1} B^T P$ (החוג הסגור)

ניתן להראות כי ביחס לשתי משוואות הריקטי הנתונות **מתקבל**, $P_2 \geq P_1$ .

# בקר משוב מצב אופטימלי

## פיתוח הבקר האופטימלי הסטציונרי באופן ישיר - המשך

נתון התהליך הבא (מסוג LTI) $\dot{x}(t) = Ax(t) + Bu(t), \quad x(0) = x_0$

נניח כי ידוע שהבקר האופטימלי הוא מסוג משוב מצב, לכן, $u(t) = -Kx(t)$

אנו מעוניינים בהגבר $K$ כך שלפונקציית המחיר הבאה ערך מינימלי,

$$J = \int_0^\infty \left( x^T(t)Qx(t) + u^T(t)Ru(t) \right) dt$$

כאשר, $Q \geq 0, \quad R > 0$

הצמד $(A, B)$ ניתן לייצוב (סטביליזבילי).

מודל החוג הסגור עם בקר משוב מצב כלהו,

$$\dot{x}(t) = (A - BK)x(t), \quad x(0) = x_0$$

לכן עבור כל מטריצת הגברים $K$ , יתקבל,

$$x(t) = e^{(A-BK)t}x_0$$

# בקר משוב מצב אופטימלי

## פיתוח הבקר האופטימלי הסטציונרי באופן ישיר

את הפתרון של החוג הסגור, עבור משוב מצב כלשהו ( $K$ ) ניתן להציב בפונקציית המחיר

$$\begin{aligned} J &= \int_0^\infty \left( x^T(t) Q x(t) + u^T(t) R u(t) \right) dt \\ &= \int_0^\infty \left( x^T(t) Q x(t) + x^T(t) K^T R K x(t) \right) dt \\ &= \int_0^\infty \left( x^T(t) (Q + K^T R K) x(t) \right) dt \\ &= x_0^T \left( \int_0^\infty \left( e^{(A-BK)^T t} (Q + K^T R K) e^{(A-BK)t} \right) dt \right) x_0 \end{aligned}$$

ברור כי רק בקר מייצב, מאפשר ערך סופי של פונקציית המחיר ובמקרה זה ניתן לסמן,

$$P = \int_0^\infty \left( e^{(A-BK)^T t} (Q + K^T R K) e^{(A-BK)t} \right) dt$$

וגם,

$$J = x_0^T P x_0$$

לכן ברור שאנו מחפשים את הבקר $K$ עבורו ה- $P$ היא מינימלית.

# בקר משוב מצב אופטימלי

פיתוח הבקר האופטימלי הסטציונרי באופן ישיר

נזכיר כי הפתרון של משוואות ליאפונוב מהצורה, $\bar{A}^T P + P\bar{A} = -\bar{Q}$

עבור מטריצה $\bar{A}$ יציבה הינו, $P = \int_0^\infty e^{\bar{A}^T t}\bar{Q}e^{\bar{A}t}dt$

לכן ברור שהמטריצה $P$ שהצגנו בשקף הקודם היא פתרון של משוואת ליאפונוב מהצורה,

$$(A - BK)^T P + P(A - BK) = -Q - K^T RK \qquad (*)$$

קל יותר לפתור משוואות ליאפונוב (כלומר למצוא את $P$ ), אבל המשוואה הזאת מנוסחת ביחס למטריצה $\bar{A}$ של החוג הסגור (עם בקר שעדיין לא ידוע).

לכן, נפתח את הסוגריים ונשנה את מבנה המשוואה במטרה למצוא את $P$ .

# <u>בקר משוב מצב אופטימלי</u>

<u>פיתוח הבקר האופטימלי הסטציונרי באופן ישיר</u>

נזכיר,

$$(A-BK)^T P + P(A-BK) = -Q - K^T RK$$

לכן,

$$A^T P + PA - \underbrace{(BK)^T P - PBK + K^T RK}_{\text{נבצע כאן השלמה לריבוע}} + Q = 0$$

$$A^T P + PA - \overbrace{P\underbrace{BR^{-1}}_{\bar{R}}B^T P + \underbrace{(R^{-1}B^T P - K)^T R(R^{-1}B^T P - K)}_{\bar{Q}}} + Q = 0$$

זוהי משוואת ריקטי, וסמנו את ה- $\bar{Q}$ וה- $\bar{R}$ שהופיעו בניסוח של ה- Comparison Theorem.

מתוך ה- Comparison Theorem אנו יודעים ש- $P$ מינימלית תתקבל כאשר $\bar{Q}$ מינימלי.

הבקר האופטימלי הוא הבקר עבורו $P$ מינימלי, לכן הגבר הבקר האופטימלי הוא $K = R^{-1}B^T P$ .

# **בקר משוב מצב אופטימלי**

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

תכנון הבקר האופטימלי דורש פתרון של משוואות ריקטי, זוהי משוואה ריבועית ב- $P$ .

**משוואת ליאפונוב היא משוואה ליניארית** ב- $P$ , לכן יותר קלה לפתרון.

ניתן להחליף את הפתרון של משוואת ריקטי בתהליך איטרטיבי בוא פותרים משוואת ליאפונוב.

הפתרונות בתהליך האיטרטיבי מתכנסים (באופן מונוטוני ועם קצב התכנסות ריבועי **בקרבת הפתרון**) לפתרון של משוואת ריקטי, כלומר לפתרון האופטימלי.

שיטת פתרון זו מתבססת על מאמר של David L. Kleinman מ- 1968, ומהווה בסיס לשיטות תכנון רבות בגישה של Reinforcement learning.

בבסיס השיטה נמצאת משוואת הליאפונוב שהצגנו קודם,

$$(A-BK)^T P + P(A-BK) + Q + K^T RK = 0 \qquad \text{(משוואה (*))}$$

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי (המשך)

עבור $K_i$ מייצב מסוים (יחד עם $(A, Q^{1/2})$ דטקטבילי), ניתן לפתור את המשוואה,

$$(*) \qquad (A - BK_i)^T P_i + P_i(A - BK_i) + Q + K_i^T R K_i = 0$$

ולקבל $P_i > 0$ .

**משפט (Kleinman 1968)**

נניח $P_i, \; i = 1, 2, \cdots$ הם פתרונות (מוגדרים חיובית) של $(*)$ , כאשר באופן רקורסיבי,

$$K_i = R^{-1} B^T P_{i-1}$$

ו- $K_0$ הינו פתרון מייצב (כלומר המטריצה $(A - BK_0)$ יציבה), אז,

1) $\quad P \le P_i \le P_{i-1} \le \cdots, \quad i = 0, 1, \ldots$

2) $\quad \lim_{i \to \infty} P_i = P$

# בקר משוב מצב אופטימלי

חישוב איטרטיבי של הבקר האופטימלי הסטציונרי (המשך)

**דוגמא**

המודל הבא מתאר את שגיאות העקיבה שלי כלי רכב ביחס למסלול רצוי (קו ישר).

(מתוך -Rajesh Rajamani, “Vehicle Dynamics and Control”, 2012, פרק 2 ע"מ 36)

$$\begin{bmatrix} \dot{e}_1 \\ \ddot{e}_1 \\ \dot{e}_2 \\ \ddot{e}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -2\dfrac{C_{\alpha 2}+C_{\alpha 1}}{mv_x} & 2\dfrac{C_{\alpha 2}+C_{\alpha 1}}{m} & -2\dfrac{C_{\alpha 2}l_2 - C_{\alpha 1}l_1}{mv_x} \\ 0 & 0 & 0 & 1 \\ 0 & -2\dfrac{C_{\alpha 2}l_2 - C_{\alpha 1}l_1}{I_{zz}v_x} & 2\dfrac{C_{\alpha 2}l_2 - C_{\alpha 1}l_1}{I_{zz}} & -2\dfrac{C_{\alpha 2}l_2^2 + C_{\alpha 1}l_1^2}{I_{zz}v_x} \end{bmatrix} \begin{bmatrix} e_1 \\ \dot{e}_1 \\ e_2 \\ \dot{e}_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \dfrac{2C_{\alpha 2}}{m} \\ 0 \\ \dfrac{2C_{\alpha 2}l_2}{I_{zz}} \end{bmatrix} \delta$$

במודל , $e_1(t)$ - שגיאת מקום (של כלי הרכב ביחס למסלול הרצוי).

$e_2(t)$ - שגיאת בכיוון (של כלי הרכב ביחס למסלול הרצוי).

$\delta(t)$ - זווית היגוי (אות הבקרה).

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי (המשך)

**דוגמא** (המשך)

בחלק זה של התוכנית, מגדירים את המודל ואת מטריצות המשקל.

```
clc, close all

% Car Parameters

m=1573; Iz=2873; l1=1.58; l2=1.1; lh=7;
Ca1=80000; Ca2=80000; vx=25;

% State Space model
A_22=-2*(Ca2+Ca1)/(m*vx);
A_23= 2*(Ca2+Ca1)/m;
A_24=-2*(Ca2*l2-Ca1*l1)/(m*vx);
A_42=-2*(Ca2*l2-Ca1*l1)/(Iz*vx);
A_43= 2*(Ca2*l2-Ca1*l1)/Iz;
A_44=-2*(-Ca2*((l2)^2)+Ca1*((l1)^2))/(Iz*vx);

B2=2*Ca2/m;
B4=2*Ca2*l2/Iz;

A=[0 1 0 0; 0 A_22 A_23 A_24; 0 0 0 1; 0 A_42 A_43 A_44];
B=[0; B2; 0; B4];

n=4; % state dimention

% Weighting matrices
q11=1; q12=0; q21=0; q22=0;
Q=diag([q11,q12,q21,q22]);
R=1;
```

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי (המשך)

```
% Model-base design with Riccati equation
[K1,P1,E1]=lqr(A,B,Q,R); K1

%%% Model-base design with Lyapunov equation
K_i=K1*0.4; eig(A-B*K_i);
P_i_old=zeros(size(A));
KK=[]; ii=[]; nPi=[];
for i=1:20
   AA=A-B*K_i; QQ=Q+K_i'*R*K_i;
   P_i=lyap(AA',QQ);norm(P_i); nPi=[nPi norm(P_i)];
   K_i=inv(R)*B'*P_i;
   if norm(P_i_old-P_i)<0.01
      KK=[KK K_i']; ii=[ii i];
      break
   end
   P_i_old=P_i;
   KK=[KK K_i']; ii=[ii i];
end
K2=K_i
figure(1)
plot(ii,KK',ii,KK','*',[ii(1) ii(end)],[K1' K1'],'--','linewidth',1)
title('Data-based Off-policy, unknown A,B')
xlabel('Iterations'), ylabel('Control gains'), grid

figure(2)
plot(nPi ,'linewidth',1)
xlabel('Iterations'), ylabel('P-norm'), grid
```

**דוגמא** (המשך)

$K_1$ - מחושב עם הפקודה **lqr** (מייצג את הערכים האופטימליים)

$K_2$ - חושב עם האלגוריתם של Kleinman (**ללא** הפקודה **lqr**). מוצגת התוצאה שהתקבלה אחרי 6 איטרציות.

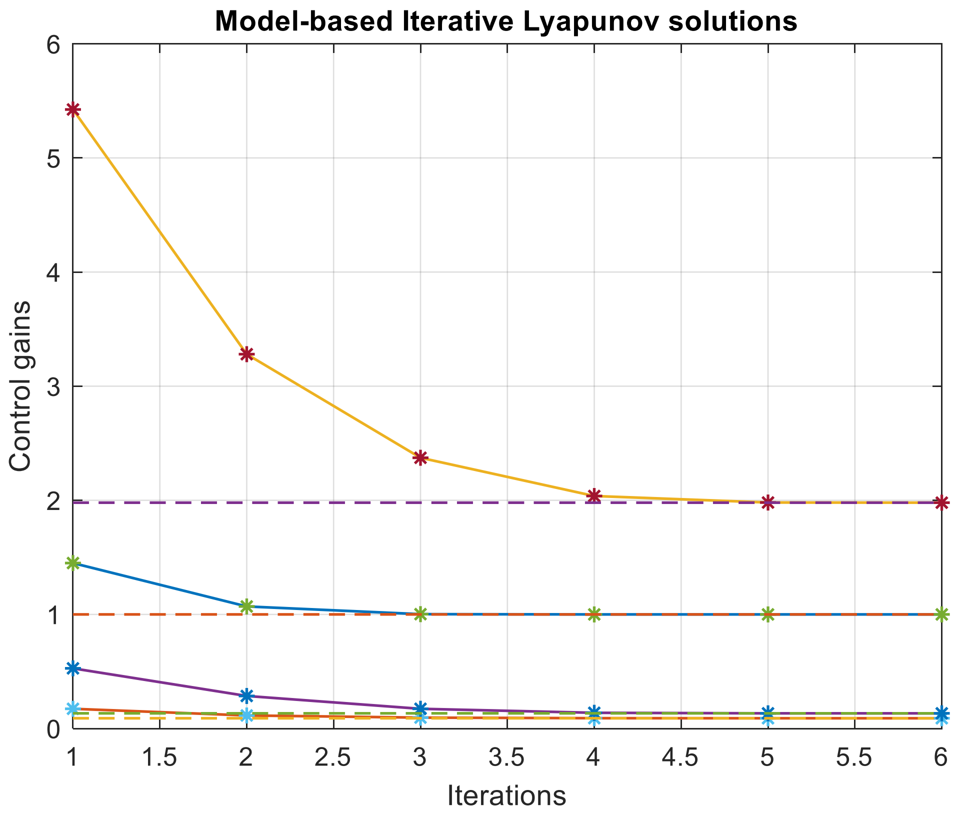
```
K1 =

    1.0000    0.0901    1.9791    0.1333


K2 =

    1.0000    0.0901    1.9791    0.1333

fx >>
```

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי (המשך)

**דוגמא** (המשך)

Figure 1



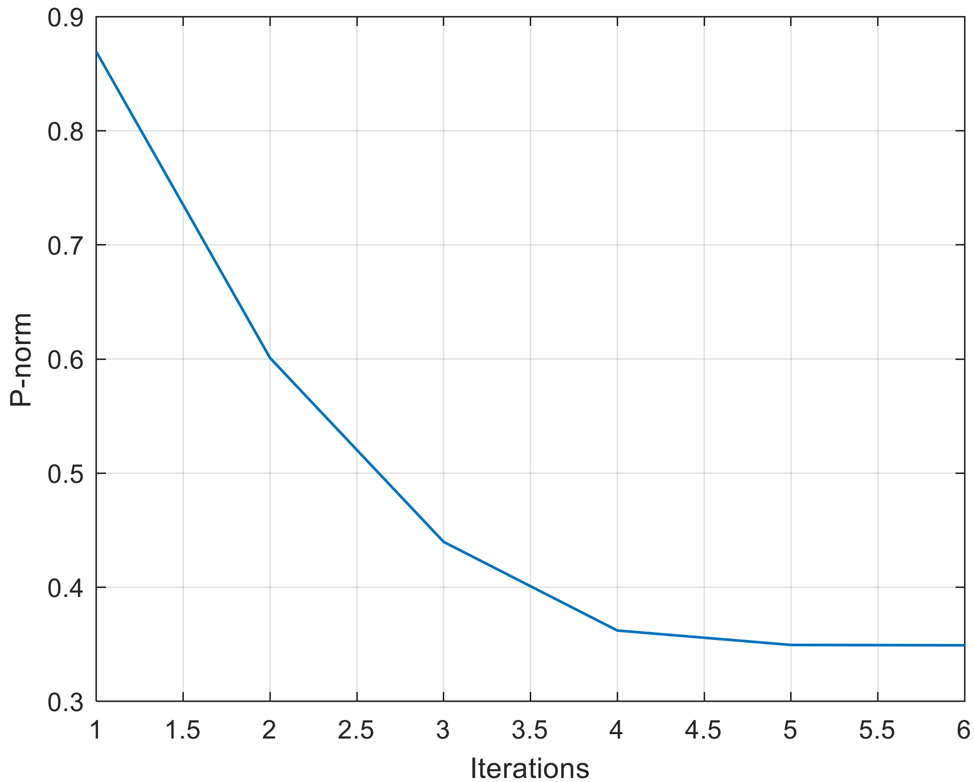

**התכנסות** הגברי הבקר אל הערכים האופטימליים (מוצגות 6 איטרציות).

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי (המשך)

**דוגמא** (המשך)

Figure 2




דעיכה מונוטונית של הנורמה של המטריצה P.

נזכיר כי לפתרון האופטימלי P עם נורמה מינימלית.

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

מכפלת קרוניקר -**Kronecker** product

נניח שתי מטריצות, הממדים
לא בהכרח זהים

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \\ \vdots & & \ddots \end{bmatrix} \quad , \quad B=\begin{bmatrix} b_{11} & b_{12} & \cdots \\ b_{21} & b_{22} & \\ \vdots & & \ddots \end{bmatrix}$$

מכפלת קרוניקר מסומנת ב- $\otimes$ ומוגדרת באופן הבא,

$$A\otimes B=\begin{bmatrix} a_{11}B & a_{12}B & \cdots \\ a_{21}B & a_{22}B & \\ \vdots & & \ddots \end{bmatrix}=\begin{bmatrix} a_{11}\begin{bmatrix} b_{11} & b_{12} & \cdots \\ b_{21} & b_{22} & \\ \vdots & & \ddots \end{bmatrix} & a_{12}\begin{bmatrix} b_{11} & b_{12} & \cdots \\ b_{21} & b_{22} & \\ \vdots & & \ddots \end{bmatrix} & \cdots \\ a_{21}\begin{bmatrix} b_{11} & b_{12} & \cdots \\ b_{21} & b_{22} & \\ \vdots & & \ddots \end{bmatrix} & a_{22}\begin{bmatrix} b_{11} & b_{12} & \cdots \\ b_{21} & b_{22} & \\ \vdots & & \ddots \end{bmatrix} & \cdots \\ \vdots & \vdots & \end{bmatrix}$$

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

מכפלת קרוניקר -**Kronecker** product

לדוגמא,

$$A=\begin{bmatrix}1 & 2\end{bmatrix} \quad , \quad B=\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix} \quad , \quad A\otimes B=\begin{bmatrix}1 & 2 & 2 & 4\\ 3 & 4 & 6 & 8\end{bmatrix}$$

או,

$$A=\begin{bmatrix}1 & 2\end{bmatrix} \quad , \quad B=\begin{bmatrix}3 & 4\end{bmatrix} \quad , \quad A\otimes B=\begin{bmatrix}3 & 4 & 6 & 8\end{bmatrix}$$

תכונות מעניינות של מכפלת קרוניקר (נדגיש כי $A\otimes B\neq B\otimes A$ ),

$$A\otimes(B+C)=A\otimes B+A\otimes C \qquad (A\otimes B)(C\otimes D)=(AC)\otimes(BD)$$

$$(kA)\otimes B=A\otimes(kB)=k(A\otimes B) \qquad (A\otimes B)^{-1}=A^{-1}\otimes B^{-1}$$

$$(A\otimes B)\otimes C=A\otimes(B\otimes C) \qquad (A\otimes B)^{T}=A^{T}\otimes B^{T}$$

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

מכפלת קרוניקר -**Kronecker** product

אנו נשתמש במכפלת קרוניקר, בעיקר כדי להפוך משוואות מטריציות (משוואה שבה הנעלם הוא מטריצה) לסט של משוואות סקלריות (על האיברים של המטריצה שמחפשים).

נגדיר את האופרטור הבא (הופך מטריצה לווקטור),

נניח,

$$A = \begin{bmatrix} | & | & \\ v_1 & v_2 & \cdots \\ | & | & \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \\ \vdots & & \ddots \end{bmatrix}$$

אז,

$$\mathrm{vec}(A) = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{12} \\ a_{22} \\ \vdots \end{bmatrix}$$

לדוגמא,

$$\mathrm{A} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \Rightarrow \mathrm{vec}(A) = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

מכפלת קרוניקר -product **Kronecker**

נגדיר גם אופרטור דומה עבור מטריצות סימטריות.

נניח $P \in R^{n\times n}$ מטריצה סימטרית ( $P = P^T$ ), כך ש

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots \\ * & p_{22} & \\ * & * & \ddots \end{bmatrix}$$

המשולש התחתון במטריצה, מסומן ב- $*$ , נקבע ע"י המשולש העליון)

לכן במטריצה כזאת יש $n^2$ איברים, אבל רק $\dfrac{n(n+1)}{2}$ נעלמים.

האופרטור $\mathrm{svec}(P)$ מסדר את אברי המטריצה $P = P^T$ במערך שהוא ווקטור, אך אינו כולל את האיברים החוזרים בגלל סימטריות.

נגדיר גם את טרנספורמציה הבאה,

$$\mathrm{vec}(P) = N\mathrm{svec}(P), \quad N \in R^{n^2\times(n^2+n)/2}$$

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

מכפלת קרוניקר -product **Kronecker**

לדוגמא,

$$P = P^T = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix}$$

אז,

$$\operatorname{svec}(P) = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{bmatrix}$$

כאן מספר האיברים ב- $\operatorname{svec}(P)$ , הינו $\frac{n(n+1)}{2} = \frac{3(3+1)}{2} = 6$

אבל,

$$\operatorname{vec}(P) = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 2 \\ 4 \\ 5 \\ 3 \\ 5 \\ 6 \end{bmatrix}$$

# בקר משוב מצב אופטימלי

חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

מכפלת קרוניקר -**Kronecker** product

ביחס לדוגמא האחרונה, ניתן להגדיר את מטריצת הטרנספורמציה הבאה, $\mathrm{vec}(P) = N\mathrm{svec}(P)$

$$\underbrace{\begin{bmatrix} 1 \\ 2 \\ 3 \\ 2 \\ 4 \\ 5 \\ 3 \\ 5 \\ 6 \end{bmatrix}}_{\mathrm{vec}(P)} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{N} \underbrace{\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{bmatrix}}_{\mathrm{svec}(P)}$$

# בקר משוב מצב אופטימלי

חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

מכפלת קרוניקר -product **Kronecker**

תכונה חשובה נוספת של מכפלת קרוניקר היא,

$$\mathrm{vec}\left(ABC\right)=\left(C^T\otimes A\right)\mathrm{vec}\left(B\right)$$

על בסיס התכונה הזאת ניתן לכתוב,

$$x^T(t)P_ix(t)=[x^T(t)\otimes x^T(t)]\mathrm{vec}(P_i)$$

לדוגמא, כאשר יש לחשב את אברי המטריצה $P_i$ באלגוריתם איטרטיבי (נציג בהמשך)

אבל מכיוון ש - $P_i$ היא מטריצה סימטרית נכתוב,

$$x^T(t)P_ix(t)=[x^T(t)\otimes x^T(t)]N\mathrm{svec}(P_i)$$

במקרה שלנו, $x(t)$ הינו ווקטור המצב של התהליך המבוקר.

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

נזכיר כי עבור התהליך,

$$\dot{x} = Ax + Bu$$

ניתן להגדיר קריטריון ביצוע,

$$J = \int_0^\infty \left( x^T Q x + u^T Q u \right) dt$$

וגם פונקציית ערך,

$$V(t) = \int_t^\infty \left( x^T Q x + u^T Q u \right) dt$$

אם משתמשים בבקר מיצב מהצורה,

$$u(t) = -K_i x(t)$$

אז פונקציית הערך מקיימת,

$$V(t) = x^T(t) P_i x(t)$$

כאשר הצמד $K_i, P_i$ מקיימים משוואת ליאפונוב מהצורה,

$$\left( A - BK_i \right)^T P_i + P_i \left( A - BK_i \right) + Q + (K_i)^T R K_i = 0$$

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

באלגוריתם של Kleinman מתחילים עם בקר מייצב $K_0$ .

ביחס אליו, פותרים את משוואת ליאפונוב,

$$\left(A - BK_0\right)^T P_0 + P_0\left(A - BK_0\right) + Q + (K_0)^T RK_0 = 0 \quad \Rightarrow \quad P_0$$

עם הפתרון $P_0$ מחשבים את הבקר $K_1$ עבור האיטרציה הבאה,

$$K_1 = R^{-1}B^T P_0$$

והתהליך חוזר על עצמו עד התכנסות, כלומר,

$$\left(A - BK_1\right)^T P_1 + P_1\left(A - BK_1\right) + Q + (K_1)^T RK_1 = 0 \quad \Rightarrow \quad P_1$$

$$K_2 = R^{-1}B^T P_1$$ וכו'

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

כלומר, באופן כללי,

$$(*) \qquad (A-BK_i)^T P_i + P_i (A-BK_i) + Q + (K_i)^T R K_i = 0 \quad \Rightarrow \quad P_i$$

$$(**) \qquad K_{i+1} = R^{-1} B^T P_i$$

כאשר האלגוריתם מתכנס, נסמן, $P_{i+1} \cong P_i = P$

אז, $K = R^{-1} B^T P$

אם נציב את ה- $K$ הזה מעלה (במשוואת ליאפונוב), תתקבל משוואת ריקטי

$$A^T P + PA + Q - PBR^{-1} B^T P = 0 \quad , \quad P > 0$$

הראינו במצגת קודמת, שעם הפתרון של משוואת ריקטי, $u = -R^{-1} B^T P x$ הוא הבקר האופטימלי.

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

הפתרון של משוואת ליאפונוב בכל איטרציה דורש ידיעה של מודל התהליך (המטריצות $A$ ו- $B$ )

נניח כי מודל התהליך אינו ידוע.

אבל יש לנו סט נתונים (DATA) של התהליך, כלומר אנו יודעים $u(t)$ ו- $x(t)$ .

ביחס למודל התהליך,

$$\dot{x} = Ax + Bu$$

ניתן לכתוב,

$$\dot{x} = Ax + Bu_i + B(u - u_i)$$

(הוספנו וחיסרנו $Bu_i$ )

כאשר, $u_i = -K_i x$ (אות הבקר של האיטרציה $i$ )

אם נפעיל את המערכת עם הבקר $u_i = -K_i x$ תתקבל פונקציית הערך

$$V_i(t) = x^T(t) P_i x(t) = \int_t^{\infty} \left( x^T Q x + u_i^T R u_i \right) dt$$

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

נציב את ביטוי הבקר בפונקציית הערך, נקבל,

$$V_i(t) = x^T(t)P_i x(t) = \int_t^{\infty} \left( x^T Q x + x K_i^T Q K_i x \right) dt$$

נגזרת בזמן של פונקציית הערך,

$$\dot{V}_i(t) = 2x^T(t)P_i \dot{x}(t)$$

נציב את $\dot{x}$ (מהשקף הקודם),

$$\dot{V}_i = 2x^T P_i \left( Ax + Bu_i + B(u - u_i) \right)$$

כאשר השמטנו את אינדקס הזמן.

ועם אות הבקרה,

$$\dot{V}_i = 2x^T P_i \left( Ax - BK_i x + B(u + K_i x) \right)$$

$$= \underbrace{2x^T P_i \left( A - BK_i \right) x}_{\text{סקלר}} + \underbrace{2x^T P_i B(u + K_i x)}_{\text{סקלר}}$$

(לביטוי שהתוצאה שלו סקלר, ניתן לבצע טרנספוז ללא שינוי ערכו)

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

בגלל שני הביטויים הסקלריים בשקף הקודם, ניתן לכתוב,

$$\dot{V}_i = x^T \left( (A - BK_i)P_i + P_i(A - BK_i) \right) x + 2(u + K_i x)^T B^T P_i x$$

נשתמש ב- $(*)$ כדי להחליף את $\left(A - BK_i\right)^T P_i + P_i \left(A - BK_i\right)$ עם $-Q - (K_i)^T RK_i$ .

נקבל,
$$\dot{V}_i = -x^T \left( Q + (K_i)^T RK_i \right) x + 2(u + K_i x)^T B^T P_i x$$
$$= -x^T Q_i x + 2(u + K_i x)^T B^T P_i x$$

כאשר $Q_i = Q + (K_i)^T RK_i$

זה הוציא את $A$ מהמשוואה.

גם $B^T$ אינו ידוע, אבל על בסיס $(**)$ ניתן לכתוב, $B^T P_i = RK_{i+1}$

# <u>בקר משוב מצב אופטימלי</u>

<u>חישוב איטרטיבי של הבקר האופטימלי הסטציונרי</u>

קיבלנו,

$$\dot{V}_i = -x^T Q_i x + 2(u + K_i\, x)^T RK_{i+1} x$$

$$= -x^T Q_i x + 2u^T RK_{i+1} x + 2x^T K_i^{\;T} RK_{i+1} x$$

נבצע אינטגרציה על שני הצדדים, מ- $t$ עד $t+T$ .

$$\int_t^{t+T} \dot{V}_i d\tau = -\int_t^{t+T} x^T Q_i x d\tau + 2\int_t^{t+T} u^T RK_{i+1} x d\tau + 2\int_t^{t+T} x^T K_i^{\;T} RK_{i+1} x d\tau$$

נציב,

$$\int_t^{t+T} \dot{V}_i d\tau = V_i(t+T) - V_i(t) = \underbrace{x^T(t+T) P_i x(t+T) - x^T(t) P_i x(t)}$$

הפרש בפונקציית הערך בין
שתי נקודות זמן.

# בקר משוב מצב אופטימלי

חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

קיבלנו,

$$x^T(t+T)P_i x(t+T) - x^T(t)P_i x(t)$$
$$= -\int_t^{t+T} x^T Q_i x d\tau + 2\int_t^{t+T} u^T R K_{i+1} x d\tau + 2\int_t^{t+T} x^T K_i{}^T R K_{i+1} x d\tau$$

הנעלמים כאן הם $P_i$ ו- $K_{i+1}$ .

(שאר המטריצות ידועות, לדוגמא $K_i$ התקבל באיטרציה קודמת)

המשוואה אינה מתבססת על מטריצות של המודל.

זוהי משוואה סקלרית, לכן משוואה יחידה אינה מספיקה כדי לחשב את כל הנעלמים (הסקלריים)

במטריצות $P_i$ ו- $K_{i+1}$ .

ניתן לייצר מספר משוואות, ע"י שימוש בנתונים השייכים למספר קטעי זמן.

למעשה, אם $x \in R^n, u \in R^m$ , יש לייצר לפחות $\frac{n(n+1)}{2} + mn$ משוואות.

# בקר משוב מצב אופטימלי

חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

נשתמש במכפלות קרוניקר כדי לחלץ את הנעלמים,

נזכיר,

$$\mathrm{vec}\left(ABC\right)=\left(C^{T}\otimes A\right)\mathrm{vec}\left(B\right)$$

אם $A$ ו- $C$ הם ווקטורים, כך ש- $ABC$ הוא סקלר אז,

$$A\cdot B\cdot C=\left(C^{T}\otimes A\right)\mathrm{vec}\left(B\right)$$

על בסיס הקשר הזה,

$$2\int_{t}^{t+T}u^{T}R\cdot K_{i+1}\cdot x d\tau=2\int_{t}^{t+T}\left(x^{T}\otimes u^{T}R\right)d\tau\,\mathrm{vec}(K_{i+1})$$

נשתמש גם בתכונה,

$$\left(A\otimes B\right)\left(C\otimes D\right)=\left(AC\right)\otimes\left(BD\right)$$

$$2\int_{t}^{t+T}u^{T}R\cdot K_{i+1}\cdot x d\tau=2\int_{t}^{t+T}\left(x^{T}\otimes u^{T}\right)d\tau\left(I_{n}\otimes R\right)\mathrm{vec}(K_{i+1})$$

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

באופן דומה, נכתוב,

$$2\int_t^{t+T} x^T K_i^{\ T} R \cdot K_{i+1} \cdot x d\tau = 2\int_t^{t+T} \left( x^T \otimes x^T \right) d\tau \left( I_n \otimes K_i^{\ T} R \right) \mathrm{vec}(K_{i+1})$$

וגם,

$$\int_t^{t+T} x^T Q_i \, x d\tau = \left( \int_t^{t+T} x^T \otimes x^T d\tau \right) \mathrm{vec}(Q_i)$$

$$\begin{aligned} & x^T(t+T) P_i x(t+T) - x^T(t) P_i x(t) \\ & \qquad = [x^T(t+T) \otimes x^T(t+T) - x^T(t) \otimes x^T(t)] N \mathrm{svec}(P_i) \end{aligned}$$

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

נניח כי נאספו נתונים לאורך קטע זמן $[t_0, t_l]$

נחלק את קטע הזמן ל- $l$ קטעים, כך ש- $l > \frac{n(n+1)}{2} + mn$ (כדי להשיג את מספר המשוואות הדרוש)

נגדיר,

$$\delta_i = [x(t_{i+1}) \otimes x(t_{i+1}) - x(t_i) \otimes x(t_i)]$$

$$\delta_{xx} = \left[\delta_1,\ \delta_2,\ \cdots,\ \delta_{l-1}\right]^T$$

$$I_{xx} = \left[\int_{t_0}^{t_1} x \otimes x d\tau,\ \int_{t_1}^{t_2} x \otimes x d\tau,\ \cdots,\ \int_{t_{l-1}}^{t_l} x \otimes x d\tau\right]^T$$

$$I_{xu} = \left[\int_{t_0}^{t_1} x \otimes u d\tau,\ \int_{t_1}^{t_2} x \otimes u d\tau,\ \cdots,\ \int_{t_{l-1}}^{t_l} x \otimes u d\tau\right]^T$$

# בקר משוב מצב אופטימלי

חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

על בסיס הגדרות אלה, ניתן לכתוב את סט המשוואות הבא,

$$X_i \Theta_i = Y_i$$

כאשר,

$$\begin{cases} X_i = \left[ -\delta_{xx} N, \quad 2I_{xu}\left(I_n \otimes R\right) + 2I_{xx}\left(I_n \otimes K_i{}^T R\right) \right] \\ Y_i = I_{xx}\mathrm{vec}(Q_i) \end{cases}$$

נתון – מתבסס על נתונים וכל התוצאה של איטרציה קודמת

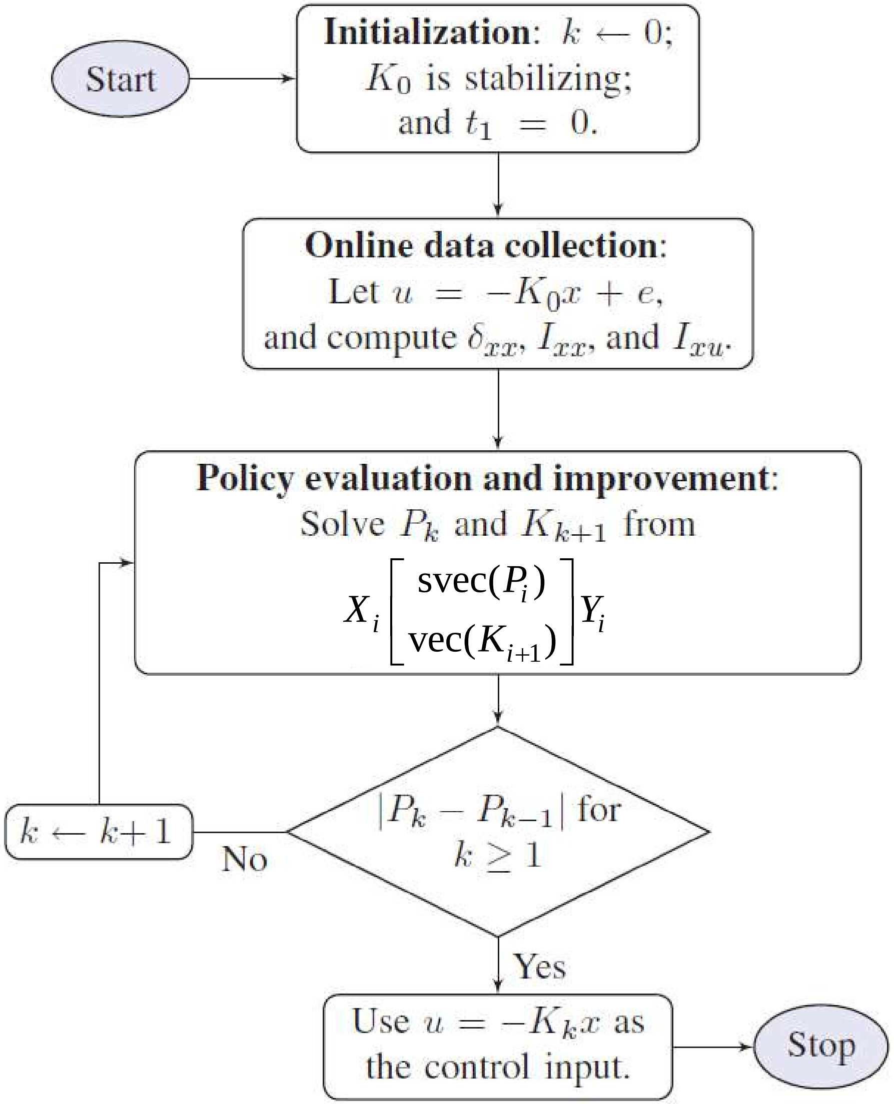
$$\begin{cases} \Theta_i = \begin{bmatrix} \mathrm{svec}(P_i) \\ \mathrm{vec}(K_{i+1}) \end{bmatrix} \end{cases}$$

ווקטור נעלמים, מחושב בכל איטרציה מחדש

רצוי לייצר יותר משוואות מנעלמים (בגלל רעש מדידה), לכן הפתרון הוא במובן של ריבועים פחותים, כלומר

$$\Theta_i = (X_i{}^T X_i)^{-1} X_i{}^T \cdot Y_i$$

# בקר משוב מצב אופטימלי




חישוב איטרטיבי של הבקר האופטימלי הסטצי

אלגוריתם תכנון בקר LQR על בסיס נתונים,

Robust Adaptive Dynamic Programming, Yu Jiang Zhong-Ping Jiang, Wiley, 2017

# בקר משוב מצב אופטימלי

חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

**דוגמא** – נמשיך את הדוגמא שפתרנו קודם, עכשיו נניח שמטריצות המודל אינן ידועות

נזכיר כי מדובר במודל המתאר דינמיקת שגיאות של כלי רכב ביחס למסלול רצוי (קו ישר)

(מתוך -Rajesh Rajamani, "Vehicle Dynamics and Control", 2012, פרק 2 ע"מ 36)

$$\begin{bmatrix} \dot{e}_1 \\ \ddot{e}_1 \\ \dot{e}_2 \\ \ddot{e}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -2\dfrac{C_{\alpha 2}+C_{\alpha 1}}{mv_x} & 2\dfrac{C_{\alpha 2}+C_{\alpha 1}}{m} & -2\dfrac{C_{\alpha 2}l_2-C_{\alpha 1}l_1}{mv_x} \\ 0 & 0 & 0 & 1 \\ 0 & -2\dfrac{C_{\alpha 2}l_2-C_{\alpha 1}l_1}{I_{zz}v_x} & 2\dfrac{C_{\alpha 2}l_2-C_{\alpha 1}l_1}{I_{zz}} & -2\dfrac{C_{\alpha 2}l_2^2+C_{\alpha 1}l_1^2}{I_{zz}v_x} \end{bmatrix} \begin{bmatrix} e_1 \\ \dot{e}_1 \\ e_2 \\ \dot{e}_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \dfrac{2C_{\alpha 2}}{m} \\ 0 \\ \dfrac{2C_{\alpha 2}l_2}{I_{zz}} \end{bmatrix} \delta$$

במודל , $e_1(t)$ - שגיאת מקום (של כלי הרכב ביחס למסלול הרצוי).

$e_2(t)$ - שגיאת בכיוון (של כלי הרכב ביחס למסלול הרצוי).

$\delta(t)$ - זווית היגוי (אות הבקרה).

# בקר משוב מצב אופטימלי

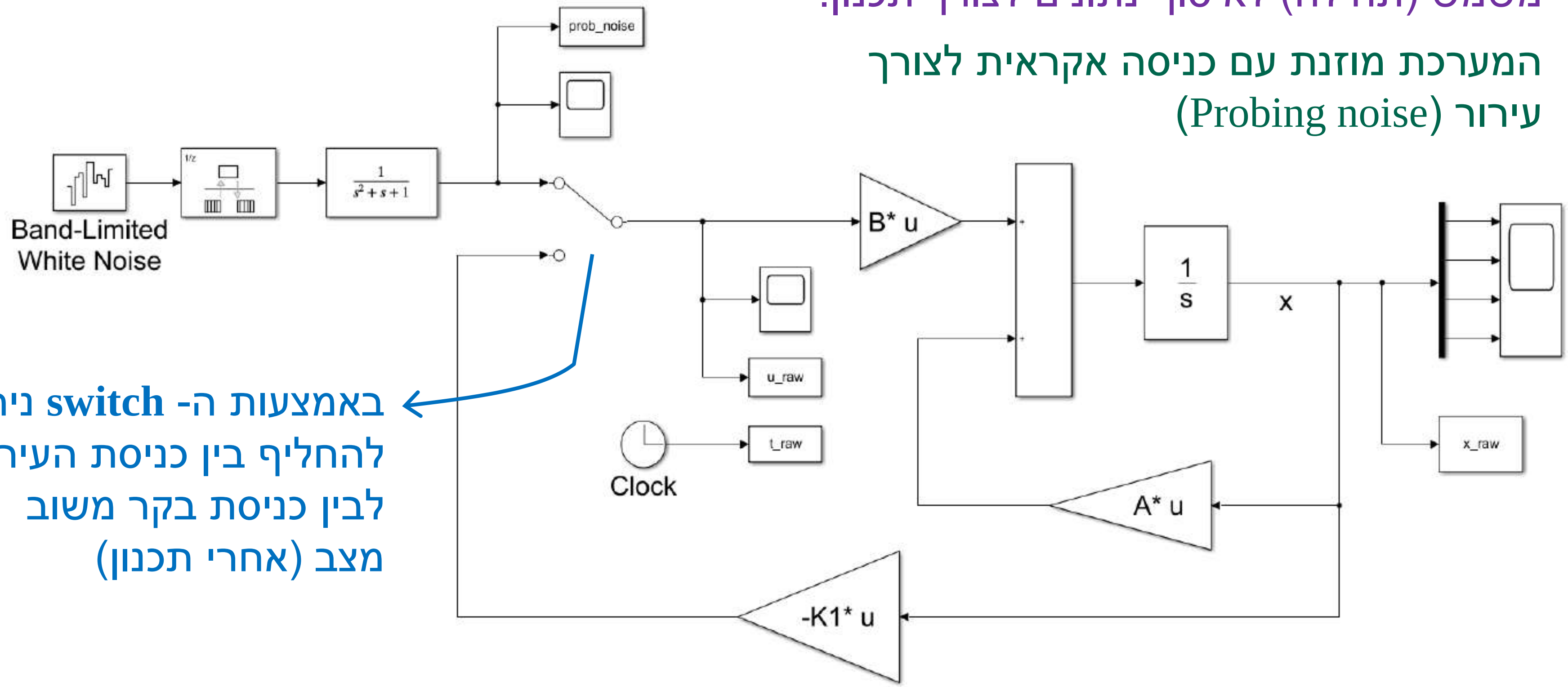
## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

**דוגמא** – המשך

דיאגרמת מלבנים של מודל התהליך.

משמש (תחילה) לאיסוף נתונים לצורך תכנון.

המערכת מוזנת עם כניסה אקראית לצורך עירור (Probing noise)




באמצעות ה- **switch** ניתן להחליף בין כניסת העירור לבין כניסת בקר משוב מצב (אחרי תכנון)

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי (המשך)

**דוגמא** (המשך)

```
% Data-base design ADP

%%% Data Collection (from a Simulink model)
sim('Lane_Follow_ADP_LQR')
strt=4000;
t=t_raw(strt:end);
x=x_raw(strt:end,:);
u=u_raw(strt:end,:);
```

הרצת סימולציה ואיסוף נתונים
(הסימולציה כאן מחליפה הפעלה של
התהליך הממשי לצורך איסוף נתונים)

```
intr=100; %2
delta_xx=[];
for i=intr+1:intr:length(t)-intr;
    x_ti=x(i,:);
    x_tim1=x(i-intr,:);
    delta_xx_i=(kron(x_ti,x_ti)-kron(x_tim1,x_tim1));
    delta_xx=[delta_xx; delta_xx_i];
end
```

$$\delta_i = [x(t_{i+1})\otimes x(t_{i+1}) - x(t_i)\otimes x(t_i)]$$

```
x_kron_x=[]; x_kron_u=[];
for i=1:length(t)
    x_kron_x=[x_kron_x; kron(x(i,:),x(i,:))];
    x_kron_u=[x_kron_u; kron(x(i,:),u(i,:))];
end

I_xu=[]; I_xx=[];
for i=intr+1:intr:length(t)-intr;
    del_i=(i-intr):i;
    I_xu=[I_xu; trapz(t(del_i),x_kron_u(del_i,:))];
    I_xx=[I_xx; trapz(t(del_i),x_kron_x(del_i,:))];
end
```

$$I_{xx} = \left[\int_{t_0}^{t_1} x\otimes x\,d\tau,\ \int_{t_1}^{t_2} x\otimes x\,d\tau,\ \cdots,\ \int_{t_{l-1}}^{t_l} x\otimes x\,d\tau\right]^T$$

$$I_{xu} = \left[\int_{t_0}^{t_1} x\otimes u\,d\tau,\ \int_{t_1}^{t_2} x\otimes u\,d\tau,\ \cdots,\ \int_{t_{l-1}}^{t_l} x\otimes u\,d\tau\right]^T$$

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי (המשך)

**דוגמא** (המשך)

חישוב מבוסס נתונים בשיטת - $K_3$

**Adaptive Dynamic Programing**

הנתונים נאספו ממודל Simulink

```
N=zeros(16,10);
N(1,1)=1;N(2,2)=1;N(3,3)=1;N(4,4)=1;
N(5,2)=1;N(6,5)=1;N(7,6)=1;N(8,7)=1;
N(9,3)=1;N(10,6)=1;N(11,8)=1;N(12,9)=1;
N(13,4)=1;N(14,7)=1;N(15,9)=1;N(16,10)=1;

K_i=K1*Error_K_0;

KK=[]; ii=[]; nPi=[];
for i=1:20
   Q_i=[Q+K_i'*R*K_i];
   Y_i=-I_xx*Q_i(:);
   X_i=[delta_xx*N, -2*I_xx*kron(eye(n),K_i'*R)-2*I_xu*kron(eye(n),R)];

   PK_i=X_i\Y_i;

   vecP_i=N*PK_i(1:10);
   P_i=[vecP_i(1:4) vecP_i(5:8) vecP_i(9:12) vecP_i(13:16)];
   nPi=[nPi norm(P_i)];
   K_i=PK_i(11:14)';
   if norm(P_i_old-P_i)<0.01
      KK=[KK K_i']; ii=[ii i];
      break
   end
   P_i_old=P_i; KK=[KK K_i']; ii=[ii i];
end
K3=K_i
nPi3=nPi;

figure(3)
plot(ii,KK',ii,KK','*',[ii(1) ii(end)],[K1' K1'],'--','linewidth',1)
title('Data-based Off-policy, unknown A,B')
xlabel('Iterations'), ylabel('Control gains'), grid
```

הגדרת המטריצה $N$

$$\Theta_i = (X_i^T X_i)^{-1} X_i^T \cdot Y_i$$

חושב עם הפקודה **lqr**

```
K1 =

    1.0000    0.0901    1.9791    0.1333
```

חושב עם האלגוריתם של **Kleinman**

```
K2 =

    1.0000    0.0901    1.9791    0.1333
```

**חישוב מבוסס נתונים בשיטת ADP**

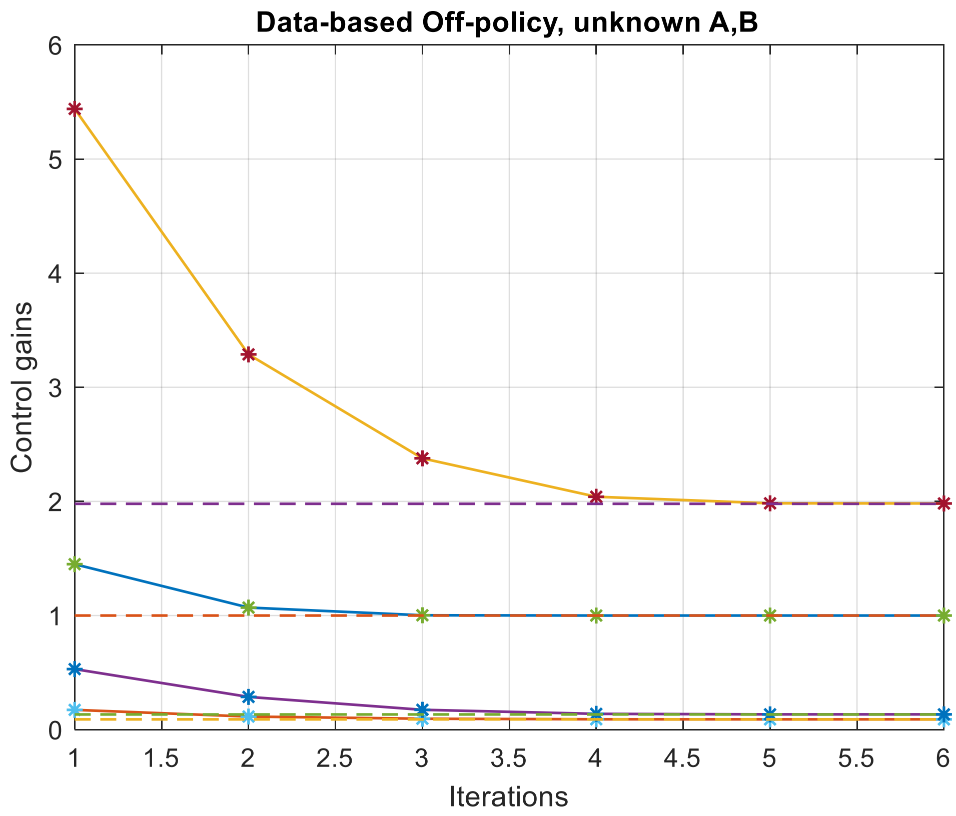
```
K3 =

    1.0000    0.0902    1.9819    0.1340

fx >>
```

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי (המשך)

**דוגמא** (המשך)

Figure 3




**התכנסות** הגברי הבקר אל הערכים האופטימליים (מוצגות 6 איטרציות).

ניתן לראות כי תהליך ההתכנסות כאן (מבוסס נתונים וללא שימוש במודל), זהה לתהליך ההתכנסות של האלגוריתם האיטרטיבי של - **Kleinman** (הוצג קודם ב- Figure 1)

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

נציג את פיתוח משוואות התכנון **למקרה הלא ליניארי** (כאשר המודל אינו ידוע – כלומר תכנון מבוסס נתונים)

מודל התהליך נתון ע"י, $\dot{x} = f(x) + g(x)u$ עם $f(0) = 0$

$x \in R^n, u \in R^m$

מעוניינים בבקר אופטימלי במובן הבא,

$$J(u^*) = \min_u \int_0^\infty \left( h^T(x)h(x) + u^T R u \right) dt$$

עם $h(0) = 0, \;\; R > 0$

הראינו כי,

$$u^*(x) = -\frac{1}{2} R^{-1} g^T(x) \nabla V(x)$$

כאשר,

$$\nabla V^T f(x) + h^T(x)h(x) - \frac{1}{4} \nabla V^T(x) g(x) R^{-1} g^T(x) \nabla V(x) = 0$$

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

תחילה יש צורך בהרחבה של האלגוריתם של Kleinman למקרה הלא ליניארי.

אם נפעיל בקר $u_i$ מייצב וכזה שעבורו פונקציית המחיר מקבלת ערך סופי (admissible),
ניתן לכתוב,

$$V_i(t) = \int_t^{\infty} \left( h^T(x)h(x) + u_i^T R u_i \right) dt$$

מגזירה של המשוואה בזמן,

$$\dot{V}_i = \nabla V_i^T(x)\dot{x} = -h^T(x)h(x) - u_i^T R u_i$$

נציב את מודל התהליך (מתקבלת משוואת Bellman),

$$\nabla V_i^T(x)\left[ f(x) + g(x)u_i \right] + h^T(x)h(x) + u_i^T R u_i = 0 \qquad \text{(Policy evaluation)}$$

עבור $u_i$ נתון, ניתן לפתור משוואה זו לחישוב $V_i(x)$ עם תנאי הסף $V_i(0) = 0$ .

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

המשוואה האחרונה (Policy evaluation) היא המקבילה למשוואת ליאפונוב מהמקרה הליניארי

$$\left(A-BK_i\right)^T P_i + P_i\left(A-BK_i\right)+Q+(K_i)^T RK_i = 0 \quad \Rightarrow \quad P_i$$

שאפשרה לחשב את $P_i$ עבור בקרה מסוים $K_i$ (למעשה לאמוד את טיבו).

עכשיו, עם $V_i(x)$ שחישבנו עבור הבקרה $u_i$ , ניתן "לקדם" את חוק הבקרה באופן הבא,

(Policy improvement) $$u_{i+1}(x) = -\frac{1}{2} R^{-1} g^T(x) \nabla V_i(x)$$

זה שקול, במקרה הליניארי, ל- $$K_{i+1} = R^{-1} B^T P_i$$

הרחבה של התיאוריה של Kleinman קובעת, $$V^*(x) \le V_{i+1}(x) \le V_i(x)$$

וגם ש- $u_i(x)$ הם בקרים מייצבים.

# בקר משוב מצב אופטימלי

חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

ההרחבה של לחישוב הבקר, דורשת ידיעה של מודל התהליך.

נפתח אלגוריתם מבוסס נתונים, דומה לזה שהצגנו עבור המקרה הליניארי.

אלגוריתמים מבוססי נתונים מהסוג שאנו מציגים כאן נקראים,

**Off-policy** reinforcement learning

(כי הנתונים המשמשים לקידום ה- policy אינם התקבלו ע"י הפעלת ה- policy הקודם בתהליך)

כמו במקרה הליניארי, גם כאן, נרשום את המודל באופן הבא,

$$\dot{x} = f(x) + g(x)u_i + g(x)(u - u_i)$$

(הוספנו וחיסרנו $g(x)u_i$ )

כאשר, $u_i$ אות הבקר של האיטרציה $i$ .

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

מגזירה של פונקציית הערך, והצבת מודל התהליך (עם הניסוח החדש),

$$\dot{V}_i(x) = \nabla V_i^T(x)\dot{x} = \nabla V_i^T(x)\left[ f(x) + g(x)u_i + g(x)(u - u_i)\right]$$

$$= \nabla V_i^T(x)\left[ f(x) + g(x)u_i \right] + \nabla V_i^T(x) g(x)(u - u_i)$$

מודל התהליך אינו ידוע, אבל ההרחבה של

התיאוריה של Kleinman מאפשרת לרשום,

(משוואת Bellman) $$\nabla V_i^T(x)\left[ f(x) + g(x)u_i \right] = -h^T(x)h(x) - u_i^T R u_i$$

$$u_{i+1}(x) = -\frac{1}{2} R^{-1} g^T(x) \nabla V_i(x)$$

$$\nabla V_i^T(x) g(x) = -2u_{i+1}^T(x) R$$

# <u>בקר משוב מצב אופטימלי</u>

<u>חישוב איטרטיבי של הבקר האופטימלי הסטציונרי</u>

אחרי הצבות, קיבלנו

$$\dot{V}_i\left(x(x)\right) = -h^T(x)h(x) - u_i^{\ T} R u_i - 2u_{i+1}^{\ T}(x)R(u-u_i)$$

נבצע אינטגרציה על שני הצדדים, מ- $t$ עד $t+T$ .

$$\int_t^{t+T} \dot{V}_i d\tau = -\int_t^{t+T} h^T(x)h(x)d\tau - \int_t^{t+T} u_i^{\ T} R u_i d\tau - 2\int_t^{t+T} u_{i+1}^{\ T}(x)R(u-u_i)d\tau$$

$$V_i\left(x(t+T)\right) - V_i\left(x(t)\right) = -\int_t^{t+T} h^T(x)h(x)d\tau - \int_t^{t+T} u_i^{\ T} R u_i d\tau - 2\int_t^{t+T} u_{i+1}^{\ T}(x)R(u-u_i)d\tau$$

הנעלמים כאן הם $V_i(x)$ ו- $u_{i+1}(x)$ (כמו במקרה הליניארי שם חישבנו $P_i$ ואת $K_{i+1}$ **במקביל**,

המשמעות היא שמבצעים גם את ה- **Policy evaluation** וגם את **Policy improvement** במקביל).

# בקר משוב מצב אופטימלי

חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

במקרה הליניארי, ידענו כי $V_i(x) = x^T P_i x$

(ידענו את הצורה של פונקציית הערך, וחישבנו את המטריצה $P_i$ של הפתרון האופטימלי)

במקרה הלא ליניארי, לא ניתן לקבוע את הצורה של $V(x)$ באופן כללי.

מכיוון שרשרת נוירונים יכולה להוות קירוב של כל פונקציה לא ליניארית כלשהי, נשתמש בקרובים הבאים,

$$\hat{V}_i(x) = \sum_{j=1}^{N_1} \hat{c}_{i,j} \phi_j(x) = \hat{C}_i^T \phi(x)$$

$$\hat{u}_{i+1}(x) = \sum_{j=1}^{N_2} \hat{w}_{i,j} \psi_j(x) = \hat{W}_i^T \psi(x)$$

$\phi(x), \psi(x)$ - ווקטורים של פונקציות בסיס בלתי תלויות (לדוגמא, פונקציות אקטיבציה של NN)

$\hat{C}_i, \hat{W}_i$ - ווקטורים של משקלים לא ידועים.

# בקר משוב מצב אופטימלי

## חישוב איטרטיבי של הבקר האופטימלי הסטציונרי

נציב את הקרובים,

נעלמים

$$\sum_{j=1}^{N_1} \hat{c}_{i,j} \left[ \phi_j \left( x(T+t) \right) - \phi_j \left( x(t) \right) \right] =$$

נעלמים

$$-\int_t^{t+T} \underbrace{\left[ h^T(x) h(x) + u_i^{\ T} R u_i \right]}_{\text{ידוע}} d\tau - 2\int_t^{t+T} \sum_{j=1}^{N_2} \hat{w}_{i,j} \psi_j^{\ T}(x) R(u - u_i) d\tau$$

אות הבקרה שחושב
באיטרציה קודמת

יש לאסוף מספיק נתונים כדי לייצר מספיק משוואות לפתרון הנעלמים $\hat{c}_{i,j}, \hat{w}_{i,j}$ .

כמו במקרה הליניארי, כל משוואה מתקבלת מתוך אינטגרציה על קטע זמן $t$ עד $t+T$ .

(למעשה יש צורך בלפחות $N_1 + N_2$ קטעי זמן)