Dual Control Revisited

Anders Rantzer

Lund University, Sweden

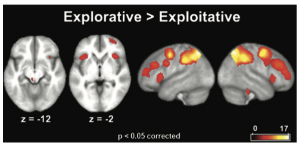
How did you decide what to eat last night?

Did you take the opportunity to try something new?

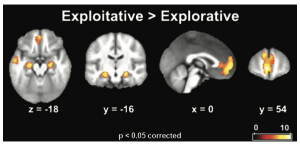
Or did you stay safe and order one of your old favorites?

The Exploration-exploitation Dilemma

In the first case, you were probably using this part of your brain:



If the second case, this is the part that you used:



[*Understanding the exploration–exploitation dilemma: An fMRI study...*, Laureiro-M, et al. (2015)]

The Exploration-Exploitation Dilemma

Exploration:
Trying out new options that may lead to better future outcomes.

Exploitation:
Choosing the best-known option based on past experiences

In Evolutionary biology: What mutation rate is good for survival?

In Management: How much should a company spend on R&D?

In Science: How much time should you spend reading past work?

Model from Computer Science: Analysis of multi-armed bandits.

Dual Control - Alexander A. Feldbaum 1913-69

Control should be probing as well as directing

- A. A. Feldbaum, Dual control theory, Avtomat. Telemekh., 1960, 21:9, 21:11
- R. E. Bellman Dynamic Programming, Academic Press 1957

Important differences from bandit problems:

- Control action can impact future learning opportunities
- Measurements often incomplete
- Unmodelled dynamics

Outline of Today's Presentation

- **Background**
- **A Data Driven Riccati Equation**

$$\Sigma_t(Q-I)\Sigma_t = \hat{\Sigma}_t^\top \min_K \left(\begin{bmatrix} I \\ K \end{bmatrix}^\top Q \begin{bmatrix} I \\ K \end{bmatrix}\right)\hat{\Sigma}_t$$

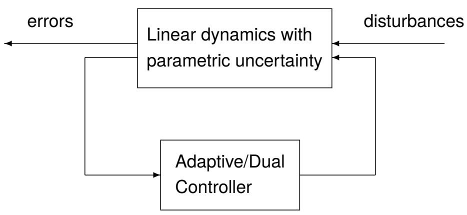
- **A Linear Quadratic Dual Controller**
- **Quantitative analysis**

(Robustness degree) $\geq$ (Excitation level) $\times$ (Degree of stabilizability)

Outline

- Background
  - **Adaptive control**
  - Learning theory
- A Data Driven Riccati Equation
- A Linear Quadratic Dual Controller
- Quantitative analysis

Dual Control for Robustness



Large parameter variations could be too much for a single linear time-invariant controller.

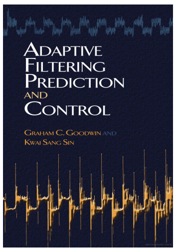
A nonlinear controller can do much better!

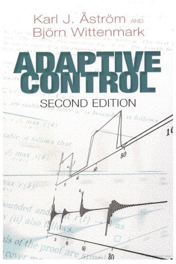
The tradeoff between exploration and exploitation leads to dual control.

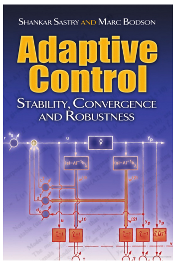
A Brief History of Adaptive Control

- Learn enough about a process and its environment for control
- Early work driven adaptive flight control 1950-1970.
  - Several adaptive schemes, but no analysis
  - Disasters in flight tests - the X-15 crash nov 15 1967
- Emergence of adaptive theory 1970-1980
  - Model reference adaptive control (servo problem) from flight control
  - The self tuning regulator (regulation problem) from process control
- Self tuning controllers on the market since 1985
- Relay Autotuning 1984

Adaptive Control — What Can We Learn?







Åstrom & Wittenmark 1995:
"*Unfortunately, there is no collection of results that can be called a theory of adaptive control in the sense specified.*"

Outline

- Background
  - Adaptive control
  - **Learning theory**
- A Data Driven Riccati Equation
- A Linear Quadratic Dual Controller
- Quantitative analysis

Statistical Machine Learning

Tail and concentration inequalites in common with

- Mathematics (measure theory, combinatorics, analysis)
- Compressed sensing
- Statistical model selection
- Network Routing
- Pattern recognition

  ⋮

Very promising for use in system identification and adaptive control!

[Abbasi-Yadkori, Faradonbeh, Hazan, Dean, Jedra, Mania, Matni, Michailidis, Pappas, Proutiere, Recht, Sandberg, Simchowitz, Szepesvari, Tu,Tsiamis, Tewari, Ziemann, ...]

(See Review in IEEE Control Systems Magazine December 2023!)

Optimal Control

Given functions f and $g \geq 0$, find a control policy μ^* to

$$\text{minimize} \quad \sum_{k=0}^{\infty} g(x_k, \mu^*(x_k))$$
$$\text{subject to} \quad x_{k+1} = f(x_k, \mu^*(x_k))$$

The infinite horizon optimal cost J^* solves the *Bellman equation*

$$J^*(x) = \min_u \Big[\underbrace{g(x,u)}_{\text{first step}} + \underbrace{J^*(f(x,u))}_{\text{future cost}}\Big]$$

The optimal policy is

$$\mu^*(x) = \arg\min_u \left[g(x,u) + J^*(f(x,u))\right].$$

The Bellman equation in terms of Q-function

$$\text{minimize} \quad \sum_{k=0}^{\infty} g(x_k, u_k) \quad \text{subject to} \quad x_+ = f(x,u)$$

The infinite horizon optimal cost J^* solves the *Bellman equation*

$$J^*(x) = \min_u \underbrace{[g(x,u) + J^*(f(x,u))]}_{Q^*(x,u)}$$

A **model free** writing of the Bellman equation is

$$Q^*(x,u) = g(x,u) + \min_v Q^*(x_+, v)$$

A large number of reinforcement learning algorithms are based on approximations of this equation.

Outline

- Background
  - Adaptive and dual control
  - Learning theory
- **A Data Driven Riccati Equation**
- A Linear Quadratic Dual Controller
- Quantitative analysis

Linear Quadratic Optimal Control

Given matrices A, B, find a control policy $u = Kx$ to

$$\text{minimize} \quad \sum_{k=0}^{\infty} \left(|x|^2 + |u|^2\right)$$
$$\text{subject to} \quad x_{k+1} = Ax + Bu$$

The infinite horizon optimal cost $|x|_P^2$ solves

$$|x|_P^2 = \min_u \left(|x|^2 + |u|^2 + |Ax + Bu|_P^2\right) \qquad \text{(Riccati equation)}$$

The minimizing argument gives a linear policy: $u = Kx$.

The Riccati equation in terms of Q-function

$$|x|_P^2 = \min_u \underbrace{\left(|x|^2 + |u|^2 + |Ax + Bu|_P^2\right)}_{\begin{bmatrix} x \\ u \end{bmatrix}^\top Q \begin{bmatrix} x \\ u \end{bmatrix}}$$

can be rewritten in terms of Q:

$$\begin{bmatrix} x \\ u \end{bmatrix}^\top (Q - I) \begin{bmatrix} x \\ u \end{bmatrix} = (Ax + Bu)^\top \underbrace{\min_K \left(\begin{bmatrix} I \\ K \end{bmatrix}^\top Q \begin{bmatrix} I \\ K \end{bmatrix} \right)}_{P} (Ax + Bu)$$

and in **model free** form

$$\begin{bmatrix} x \\ u \end{bmatrix}^\top (Q - I) \begin{bmatrix} x \\ u \end{bmatrix} = x_+^\top \min_K \left(\begin{bmatrix} I \\ K \end{bmatrix}^\top Q \begin{bmatrix} I \\ K \end{bmatrix} \right) x_+$$

The minimizing K gives an optimal policy $u = Kx$.

The Riccati equation in terms of Q-function

$$\begin{bmatrix} x \\ u \end{bmatrix}^\top (Q - I) \begin{bmatrix} x \\ u \end{bmatrix} = x_+^\top \min_K \left(\begin{bmatrix} I \\ K \end{bmatrix}^\top Q \begin{bmatrix} I \\ K \end{bmatrix} \right) x_+$$

can be solved by collecting data:

$$\begin{bmatrix} x_0 & \dots & x_t \\ u_0 & \dots & u_t \end{bmatrix}^\top (Q - I) \begin{bmatrix} x_0 & \dots & x_t \\ u_0 & \dots & u_t \end{bmatrix}$$
$$= \begin{bmatrix} x_1 & \dots & x_{t+1} \end{bmatrix}^\top \min_K \left(\begin{bmatrix} I \\ K \end{bmatrix}^\top Q \begin{bmatrix} I \\ K \end{bmatrix} \right) \begin{bmatrix} x_1 & \dots & x_{t+1} \end{bmatrix}$$

If $(x_0, u_0), \dots, (x_t, u_t)$ span all dimensions in $\mathbb{R}^{n+m}$, then this gives the optimal control law! Can we stop here? No. This is the start![1]

[1] For linear quadratic Q-learning, see [Bradtke (1992)] and [Rizvi/Lin (2019)].

A Data Driven Riccati Equation

Multiply

$$\begin{bmatrix} x_0 \dots x_{t-1} \\ u_0 \dots u_{t-1} \end{bmatrix}^\top (Q - I) \begin{bmatrix} x_0 \dots x_{t-1} \\ u_0 \dots u_{t-1} \end{bmatrix} = \begin{bmatrix} x_1 \dots x_t \end{bmatrix}^\top \min_K \left(\begin{bmatrix} I \\ K \end{bmatrix}^\top Q \begin{bmatrix} I \\ K \end{bmatrix} \right) \begin{bmatrix} x_1 \dots x_t \end{bmatrix}$$

from the left by

$$\begin{bmatrix} \lambda^t x_0 & \dots & x_{t-1} \\ \lambda^t u_0 & \dots & u_{t-1} \end{bmatrix},$$

its transpose from the right. This gives a **data driven Riccati equation**:

$$\Sigma_t (Q - I) \Sigma_t = \hat{\Sigma}_t^\top \min_K \left(\begin{bmatrix} I \\ K \end{bmatrix}^\top Q \begin{bmatrix} I \\ K \end{bmatrix} \right) \hat{\Sigma}_t$$

where λ is a forgetting factor and

$$\Sigma_t = \sum_{k=0}^{t-1} \lambda^{t-1-k} \begin{bmatrix} x_k \\ u_k \end{bmatrix} \begin{bmatrix} x_k \\ u_k \end{bmatrix}^\top, \qquad \hat{\Sigma}_t = \sum_{k=0}^{t-1} \lambda^{t-1-k} x_{k+1} \begin{bmatrix} x_k^\top & u_k^\top \end{bmatrix}.$$

Comments on the Data Driven Riccati Equation

- Unlike most reinforcement learning algorithms, memory is in Σ_t and $\hat{\Sigma}_t$, not in Q. Linear dynamics of Σ_t and $\hat{\Sigma}_t$ simplifies analysis.
- When Σ_t is invertible, the data driven Riccati equation is algebraically equivalent to the standard Riccati equation for $\begin{bmatrix} \hat{A}_t & \hat{B}_t \end{bmatrix} := \hat{\Sigma}_t \Sigma_t^{-1}$.
- Hard to enforce stabilizability of $(\hat{A}_t, \hat{B}_t)$. Easy to bound Q.
- **Excitation directions** of Σ_t determine the accuracy of Q and K. However, only controllable state directions matter.

Excitation and Dual Control

1985 Bai/Sastry **Persistency of excitation**, *sufficient richness and parameter convergence in discrete time adaptive control*

1986 Green/Moore, **Persistence of excitation** *in linear systems*

1988 Mareels/Gevers **Persistency of excitation** *criteria for linear, multivariable, time-varying systems*

2005 Willems et.al, *A note on* **persistency of excitation**
(With the "fundamental lemma" recently used for data-driven control)

1986 Åström/Helmersson, ***Dual control*** *of an integrator with unknown gain*

1995 Wittenmark, *Adaptive* ***dual control*** *methods: An overview*

2018 Mesbah, *Stochastic model predictive control with active uncertainty learning: A Survey on* ***dual control***

2021 Flayac/Nair/Shames, *Nonlinear* ***dual control*** *based on fast moving horizon estimation and model predictive control with an observability constraint*

Outline

Problem Formulation

Define $\mathcal{M}_\beta$ as the set of all pairs (A, B) such that there exists Q with $I \preceq Q \preceq \beta^2 I$ and

$$Q - I = \begin{bmatrix} A & B \end{bmatrix}^\top \min_K \left(\begin{bmatrix} I \\ K \end{bmatrix}^\top Q \begin{bmatrix} I \\ K \end{bmatrix} \right) \begin{bmatrix} A & B \end{bmatrix}.$$

Find a controller $\mu : (x_0, \dots, x_t) \mapsto u_t$, that stabilizes the system

$$x_{t+1} = Ax_t + Bu_t + w_t \qquad t \geq 0$$

for all $(A, B) \in \mathcal{M}_\beta$ subject to a bound on disturbances w_t.

Optimal state feedback behavior is expected as $\lim_{t \to \infty} w_t = 0$.

The Linear Quadratic Dual Controller

$$\Sigma_{t+1} = \lambda \underbrace{\begin{bmatrix} \Sigma_t^{xx} & \Sigma_t^{xu} \\ \Sigma_t^{ux} & \Sigma_t^{uu} \end{bmatrix}}_{\Sigma_t} + \begin{bmatrix} x_t \\ u_t \end{bmatrix} \begin{bmatrix} x_t \\ u_t \end{bmatrix}^\top \qquad \Sigma_0 = 0$$

$$\hat{\Sigma}_{t+1} = \lambda \hat{\Sigma}_t + x_{t+1} \begin{bmatrix} x_t^\top & u_t^\top \end{bmatrix} \qquad \hat{\Sigma}_0 = 0$$

$$K_t = \mathbf{K}(\Sigma_t, \hat{\Sigma}_t)$$

$$u_t = \underbrace{K_t x_t}_{\text{Exploitation}} + \underbrace{\epsilon |x_t| \mathbf{E}\left(\Sigma_t, K_t x_t - \Sigma_t^{ux} (\Sigma_t^{xx})^\dagger x_t\right)}_{\text{Exploration}}$$

The states Σ_t, $\hat{\Sigma}_t$, collect correlation data with forgetting factor $\lambda \in [0, 1]$.

Controller map $\mathbf{K}$ gives K_t. $\mathbf{E}$ provides direction for excitation/exploration.

The Controller and Excitation Maps

Let $Q_t \succeq I$ be a solution to the "data driven Riccati equation"

$$\hat{\Sigma}_t^\top \min_K \left(\begin{bmatrix} I \\ K \end{bmatrix}^\top Q_t \begin{bmatrix} I \\ K \end{bmatrix} \right) \hat{\Sigma}_t = \Sigma_t (Q_t - I) \Sigma_t$$

and let $\mathbf{K}(\Sigma_t, \hat{\Sigma}_t)$ be a minimizing value of K.

Let $\Sigma_t^{uu} - \Sigma_t^{ux} (\Sigma_t^{xx})^{-1} \Sigma_t^{xu}$ have eigenvalues $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m \geq 0$ with corresponding eigenvectors $e_1, \ldots, e_m$. For $v \in \mathbb{R}^m$ with $v = v_1 e_1 + \cdots + v_m e_m$, define the *excitation map*

$$\mathbf{E}(\Sigma, v) := \text{sign}(v_m) e_m$$

Output Feedback

The input-output model

$$y_t + a_1 y_{t-1} + \cdots + a_n y_{t-n} = b_1 u_{t-1} + \cdots + b_n u_{t-n} + v_t$$

can be written with $x_t = \begin{bmatrix} y_t & \ldots & y_{t-n+1} & u_{t-1} & \ldots & u_{t-n+1} \end{bmatrix}^\top$ as

$$x_{t+1} = \underbrace{\begin{bmatrix} -a_1 & \ldots & -a_{n-1} & -a_n & b_2 & \ldots & b_{n-1} & b_n \\ 1 & & & 0 & 0 & & & 0 \\ & \ddots & & \vdots & & \ddots & & \vdots \\ & & 1 & 0 & & & & \\ 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & \ldots & 0 & 0 & 1 & & 0 & 0 \\ & \ddots & & \vdots & & \ddots & & \vdots \\ & & 0 & 0 & & & 1 & 0 \end{bmatrix}}_{A} x_t + \underbrace{\begin{bmatrix} b_1 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{B} u_t + \underbrace{\begin{bmatrix} v_t \\ 0 \\ \vdots \\ \vdots \\ \vdots \\ 0 \end{bmatrix}}_{w_t}$$

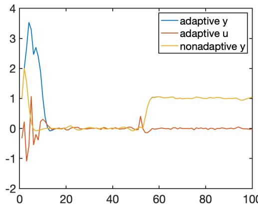
In practice: Other filters of past states and inputs give better conditioning!

Example 1: A Double Integrator

Simulate an input-output model with transfer function $(z-1)^{-2}$:

$$y_{t+1} = 2y_t - y_{t-1} + u_{t-1} + \text{white noise}$$

with unit initial value and step reference change at $t = 50$:



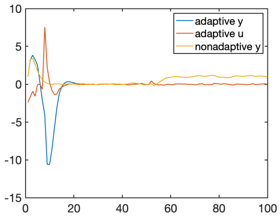
After inital adaptation, the adaptive controller follows the optimal perfectly.

Example 2: Add a Zero at $z = 2$

Simulate an input-output model with transfer function $(1 - z/2)(z-1)^{-2}$.

$$y_{t+1} = 2y_t - y_{t-1} - 0.5u_{t-1}u_t + u_{t-1} + \text{white noise}$$

Transients are bigger. That's all:



But can anything be proved rigorously?

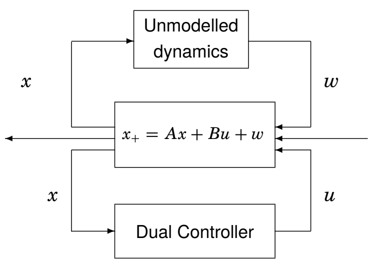
Outline

- Background
- A Data Driven Riccati Equation
- A Linear Quadratic Dual Controller
- **Quantitative analysis**

Main Result

(Robustness degree) $\geq$ (Excitation level) $\times$ (Degree of stabilizability)

Robustness degree



Consider $x_{t+1} = Ax_t + Bu_t + w_t$ with the following bounds:

$$\Sigma_t^{ww} \preceq \gamma^{-2} \Sigma_t^{xx}, \qquad 0 \preceq \begin{bmatrix} (\Sigma_t^{xx})^2 & \Sigma_t^{xw} \\ \Sigma_t^{wx} & \gamma^{-2} I \end{bmatrix}.$$

"Robustness degree" is the maximal γ^{-1} for which stability is guaranteed.

Excitation level

The system is said to have excitation level $\delta \in (0, 1)$ if

$$\begin{bmatrix} \Sigma_t^{xx} & \Sigma_t^{xu} \\ \Sigma_t^{ux} & \Sigma_t^{uu} \end{bmatrix} \succeq \delta \begin{bmatrix} \Sigma_t^{xx} & 0 \\ 0 & \|\Sigma_t\| I \end{bmatrix}.$$

for all $t \geq t_0$, where $t_0 \geq n + m$.

Remark:
This is a *quantative* notion of excitation, rather than then traditional *qualitative* one.

Main Theorem

Suppose that $(A, B) \in \mathcal{M}_\beta$ with Q and K being the corresponding solutions to the Riccati equation. If the excitation level is δ, then the linear quadratic dual controller connected to $x_+ = Ax + Bu + w$ gives exponential stability provided that

$$\gamma^{-1} \leq \underbrace{(\text{Excitation level})}_{\delta} \times \underbrace{(\text{Degree of stabilizability})}_{\left[2\sqrt{2}\beta(\beta^2+1)^2\right]^{-1}}$$

Moreover, for $t \geq t_0$, the closed loop system satisfies

$$\begin{bmatrix} x_{t+1} - w_{t+1} \\ K_{t+1}x_{t+1} \end{bmatrix}^\top Q \begin{bmatrix} x_{t+1} - w_{t+1} \\ K_{t+1}x_{t+1} \end{bmatrix} \leq \alpha\left(|x_t|^2 + |K_t x_t|^2|\right) + \begin{bmatrix} x_t \\ K_t x_t \end{bmatrix}^\top Q \begin{bmatrix} x_t \\ K_t x_t \end{bmatrix}.$$

where $\alpha := 1 - 2\sqrt{2}\beta(\beta^2 + 1)^2\gamma^{-1}\delta^{-1}$.

Proof ideas

- Use w-bounds and degree of excitation to verify that Q_t is close to Q (in controllable state directions).
- This gives that K_t is close to the optimal K.
- Uncontrollable state directions are fine by degree of stabilizability.
- Stability and dissipativity follows from corresponding properties of the optimal controller for known A and B.

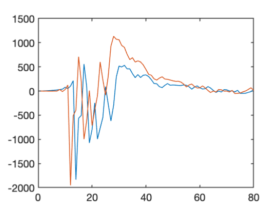
Outline

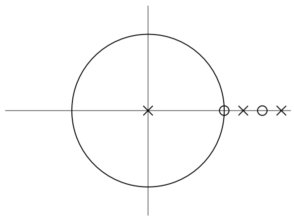
- Background
- A Data Driven Riccati Equation
- A Linear Quadratic Dual Controller
- Quantitative analysis
- **More examples and conclusions**

Example 3: A System Hard to Control

Even systems which require unstable controllers seem to work fine:

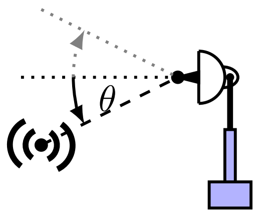
$$P(z) = \frac{(z-1)(z-1.1)}{z^2(z-1.05)(z-1.15)}$$





But other examples call for other solutions...

Example 4: Control based on absolute value



Objective: Direct antenna towards target.

Measurement: Signal strength gives absolute value of θ.

Dual control:
Move antenna to learn sign of θ at the expense of short term performance.

[Olle Kjellqvist arxiv.org/abs/2312.05156]

Conclusions

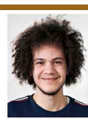
- Adaptive and dual control should be revisited. Parameterization using Q-matrix avoids many past difficulties. Åström/Wittenmark's missing theory is within sight!
- Natural step from data driven Riccati equation to MPC.
- Conservative bounds should be improved. Use statistical methods to further reduce conservatism.

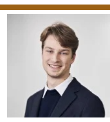
See personal web page and [arxiv.org/abs/2312.06014]

Thanks


Pauline Kergus


Felix Agner


Fethi Bencherki


David Ohlin


Richard Pates

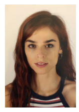

Taouba Jouini


Emil Vladu


Olle Kjellqvist


Carolina Bergeling


Johan Grönqvist


Alba Gurpegui


Venkatraman Renganathan

erc European Research Council
VETENSKAPSRÅDET
ELLIIT
WASP WALLENBERG AI AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM