I4C: First there was variance, then bias, what now?

Håkan Hjalmarsson
with Kévin Colin and Yue Ju
KTH Royal Institute of Technology, Stockholm

Four decades of data-driven modeling in systems and control, April 19, 2024

KTH VETENSKAP OCH KONST

WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

Swedish Research Council

First there was variance

First there was variance

K.J. Åström and T. Bohlin (1965). **"Numerical identification of linear dynamic systems from normal operating records".** In: *Proc. IFAC Symp. Self-Adaptive Systems*. Teddington, U.K., pp. 96–111

K.J. Åström and B. Wittenmark (1971). **"Problems of identification and control".** In: *J. Math. Analalysis and Applications* 34, pp. 90–113

K.J. Åström et al. (1977). **"Theory and Applications of Self-Tuning Regulators".** In: *Automatica* 13, pp. 457–476

M. Gevers and L. Ljung (1986). **"Optimal experiment designs with respect to the intended model application".** In: *Automatica* 22.5, pp. 543–554

K.J. Åström and B. Wittenmark (1989). ***Adaptive Control.*** Reading, Massachusetts: Addison-Wesley

- True system in model set
- Certainty equivalence principle used in control design
- MSE for control performance used as criterion in optimal experiment design

and then there was bias

and then there was bias

B. Wahlberg and L. Ljung (1986). **"Design variables for bias distribution in transfer function estimation".** In: *IEEE Trans. Automatic Control* 31.2, pp. 134–144

R. J. P. Schrama (1992). **"Accurate models for control design: the necessity of an iterative scheme".** In: *IEEE Transactions on Automatic Control* 37.7, pp. 991–994

R.J.P. Schrama and P.M.J. Van den Hof (1992). **"An Iterative Scheme for Identification and Control Design Based on Coprime Factorizations".** In: *Proc. ACC*. Chicago, pp. 2842–2846

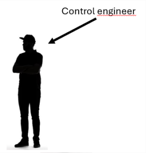
R.G. Hakvoort, R.J.P. Schrama, and P.M.J. Van den Hof (1992). **"Approximate Identification in view of LQG Feedback Design".** In: *Proc. ACC*. Chicago

R.J.P. Schrama and P.M.J. Van den Hof (1993). **"Iterative Identification and Control Design: A Three Step Procedure with Robustness Analysis".** In: *2nd European Control Conference*. Groningen, pp. 237–241

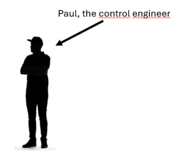
- Restricted complexity models
- Focus on the bias error
- Match model closed loop to true closed loop for the same controller
- Iterative procedures
- Closed loop identification under non-ideal conditions

and what now???

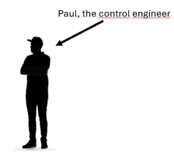
A typical day in the life of a control engineer

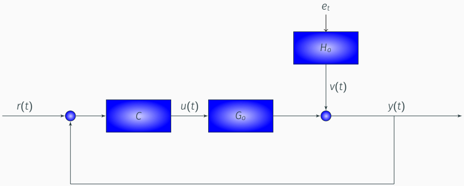


A typical day in the life of a control engineer

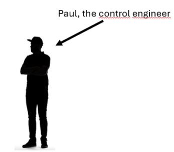


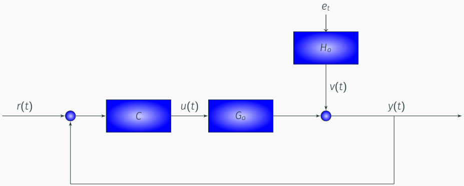
A typical day in the life of a control engineer





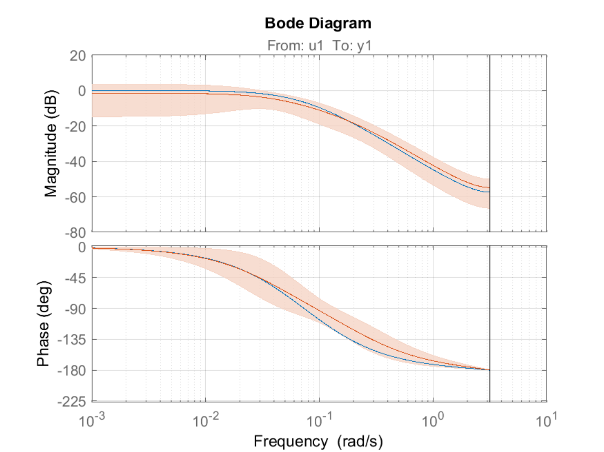
A typical day in the life of a control engineer



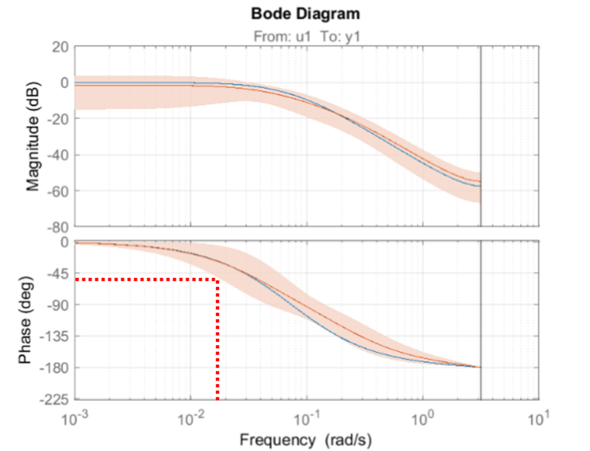


Task: Design a PI-controller $C(z) = K/(1 - z^{-1})$ using data so that step disturbances rejected as quickly as possible.

Given some data, an ARX-model seems like a good starting point:



Given some data, an ARX-model seems like a good starting point:



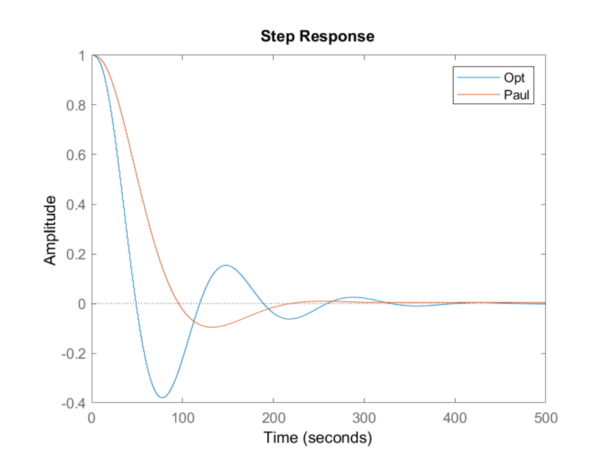
Choose cross-over frequency $\omega_c = 0.017$ [rad/s]

A typical day in the life of a control engineer

Optimal controller: $K(\theta) = \arg\min_\theta V_{\text{Step}}(K, \theta) = \arg\min_\theta \sum_t y^2_{\text{step}}(t, K, \theta)$

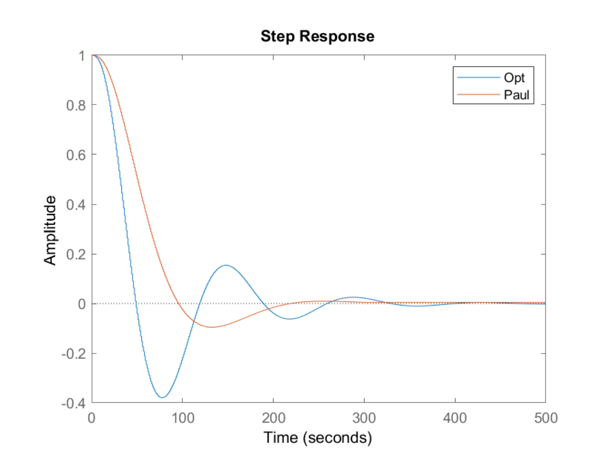
A typical day in the life of a control engineer

Optimal controller: $K(\theta) = \arg\min_\theta V_{\text{Step}}(K, \theta) = \arg\min_\theta \sum_t y^2_{\text{step}}(t, K, \theta)$



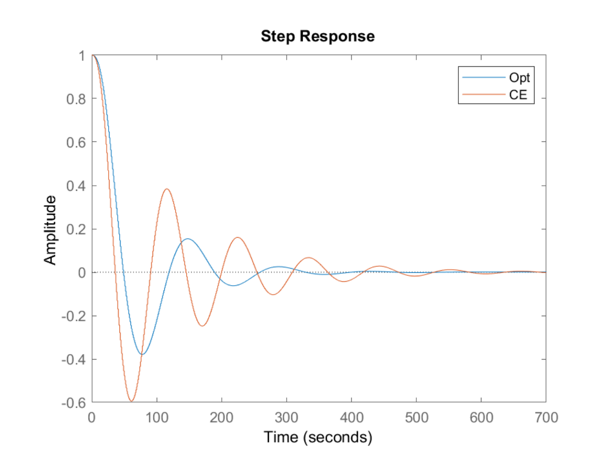
A typical day in the life of a control engineer

Optimal controller: $K(\theta) = \arg\min_\theta V_{\text{Step}}(K, \theta) = \arg\min_\theta \sum_t y^2_{\text{step}}(t, K, \theta)$

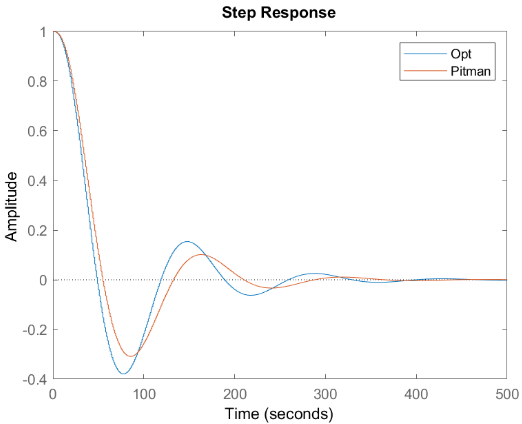


Well done lad!

Alternative 1: Certainty equivalence $K(\hat{\theta}_{\mathrm{ARX}})$



Alternative 2: Pitman type tuning



A typical day in the life of a control engineer

What's this?

Suppose we have the ARX-model $y(t) = \varphi^T(t)\theta + e(t)$

A typical day in the life of a control engineer

What's this?

Suppose we have the ARX-model $y(t) = \varphi^T(t)\theta + e(t)$

$$K_{\mathrm{Pitman}} := \arg\min_K \int V_{\mathrm{Step}}(K,\theta) \underbrace{e^{-\frac{1}{2\sigma^2}\sum_{t=1}^{N}(y(t)-\varphi^T(t)\theta)^2}} d\theta$$

What's this?

Suppose we have the ARX-model $y(t) = \varphi^T(t)\theta + e(t)$

$$K_{\text{Pitman}} := \arg\min_K \int V_{\text{Step}}(K, \theta) \underbrace{e^{-\frac{1}{2\sigma^2}\sum_{t=1}^{N}(y(t)-\varphi^T(t)\theta)^2}}_{\text{Weighting}} d\theta$$

- More weight to step-responses for models which are consistent with data

What's this?

Suppose we have the ARX-model $y(t) = \varphi^T(t)\theta + e(t)$

$$K_{\mathrm{Pitman}} := \arg\min_K \int V_{\mathrm{Step}}(K,\theta)\underbrace{e^{-\frac{1}{2\sigma^2}\sum_{t=1}^{N}(y(t)-\varphi^T(t)\theta)^2}}_{\text{Weighting}}d\theta$$

- More weight to step-responses for models which are consistent with data
- This is basically what's under the hood in DeePC

J. Coulson, J. Lygeros, and F. Dörfler (2019). **"Data-enabled predictive control: In the shallows of the DeePC".** In: *18th European Control Conference (ECC)*, pp. 307–312

A typical day in the life of a control engineer

What's this?

Suppose we have the ARX-model $y(t) = \varphi^T(t)\theta + e(t)$

$$K_{\mathrm{Pitman}} := \arg\min_K \int V_{\mathrm{Step}}(K,\theta)\underbrace{e^{-\frac{1}{2\sigma^2}\sum_{t=1}^{N}(y(t)-\varphi^T(t)\theta)^2}}_{\text{Weighting}} d\theta$$

- More weight to step-responses for models which are consistent with data
- This is basically what's under the hood in DeePC

J. Coulson, J. Lygeros, and F. Dörfler (2019). **"Data-enabled predictive control: In the shallows of the DeePC".** In: *18th European Control Conference (ECC)*, pp. 307–312

as shown in

Alessandro Chiuso et al. (2023). ***Harnessing the Final Control Error for Optimal Data-Driven Predictive Control.*** arXiv: `2312.14788 [eess.SY]`

A typical day in the life of a control engineer

What's this?

Suppose we have the ARX-model $y(t) = \varphi^T(t)\theta + e(t)$

$$K_{\text{Pitman}} := \arg\min_K \int V_{\text{Step}}(K,\theta) \underbrace{e^{-\frac{1}{2\sigma^2}\sum_{t=1}^{N}(y(t)-\varphi^T(t)\theta)^2}}_{\text{Weighting}} d\theta$$

- More weight to step-responses for models which are consistent with data
- This is basically what's under the hood in DeePC

  J. Coulson, J. Lygeros, and F. Dörfler (2019). **"Data-enabled predictive control: In the shallows of the DeePC".** In: *18th European Control Conference (ECC)*, pp. 307–312

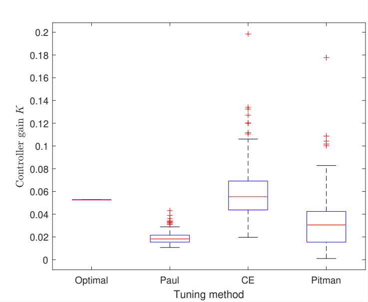
  as shown in

  Alessandro Chiuso et al. (2023). ***Harnessing the Final Control Error for Optimal Data-Driven Predictive Control.*** arXiv: *2312.14788 [eess.SY]*

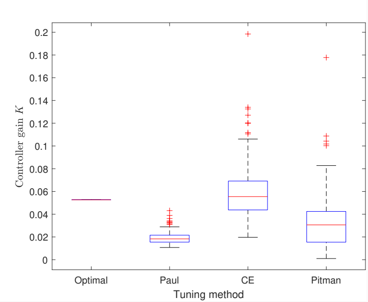
- Same principle used by Pitman for estimating the location parameter θ when y has distribution $f(y-\theta) = f(y(1)-\theta), \ldots, y(N)-\theta)$

  E. J. G. Pitman (1939). **"The Estimation of the Location and Scale Parameters of a Continuous Population of any Given Form".** In: *Biometrika* 30.3/4, pp. 391–421

Monte Carlo study: 100 simulations.



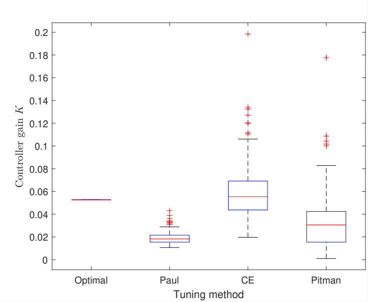
Monte Carlo study: 100 simulations.

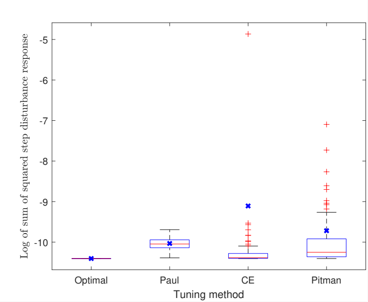


- Certainty equivalence takes too big risks

A typical day in the life of a control engineer

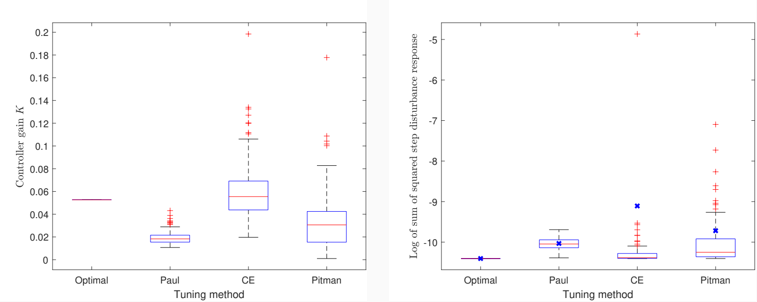
Monte Carlo study: 100 simulations.





- Certainty equivalence takes too big risks

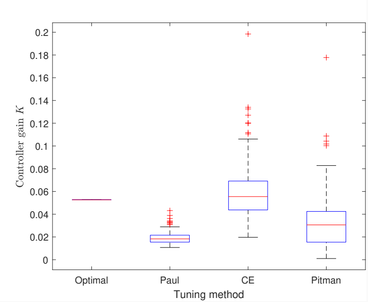
Monte Carlo study: 100 simulations.

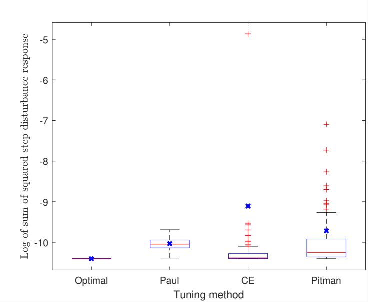


- Certainty equivalence takes too big risks
- Pitman type more cautious, but still makes a few chancy decisions

A typical day in the life of a control engineer

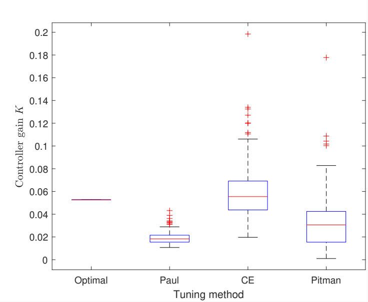
Monte Carlo study: 100 simulations.

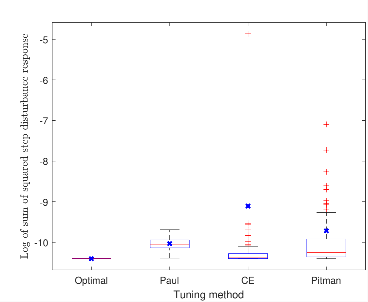




- Certainty equivalence takes too big risks
- Pitman type more cautious, but still makes a few chancy decisions
- Paul plays it safe

Monte Carlo study: 100 simulations.





- Certainty equivalence takes too big risks
- Pitman type more cautious, but still makes a few chancy decisions
- Paul plays it safe

Before proceeding, who's this gifted control engineer?

A typical day in the life of a control engineer

Autoprofit


Kroatien
Serbien
Italien
Rom
Marseille
Barcelona
Bulgarien
Svarta havet
Georgien
Azerbajdzjan
Uzbekistan
Kyrgyzstan
Almaty
Turkmenistan
Tadzjikistan
Portugal
Lissabon
Spanien
Alger
Tunis
Grekland
Istanbul
Aten
Teheran
XINJIANG
QINGHAI
Tunisien
Medelhavet
Syrien
Libanon
Irak
Israel
Jordanien
Marocko
Kuwait
Libyen
Egypten
Riyadh
Persian Gulf
Förenade Arabemiraten
Västsahara
Jidda
Oman
Mauretanien
Röda havet
Nepal
Bhutan
Bangladesh
TIBET
SICHUAN
YUNNAN
Ahmedabad
Bombay
Hyderabad
Senegal
Gambia
Guinea-Bissau
Bamako
Guinea
Sierra Leone
Liberia
Togo
Ghana
Abidjan
Lomé
Lagos
Guineabukten
Eritrea
Yemen
Adenviken
Djibouti
Arabiska havet
Bangalore
Chennai
Coimbatore
Bengaliska bukten
Thailand
Bangkok
Andamansjön
Thailandviken
Sri Lanka
Lackadivsjön
Malaysia
Kuala Lumpur
Etiopien
Sydsudan
Kamerun
Yaoundé
Ekvatorialguinea
Somalia
Uganda
Kampala
Nairobi
Mombasa
Gabon
Kongo
Demokratiska republiken Kongo
Tanzania
Luanda
Angola
Zambia
Malawi
Moçambique
Zimbabwe
Madagaskar
Port Louis
Namibia
Botswana
Indiska oceanen
Södra Atlanten
Lesotho
Sydafrika
Kapstaden
Port Elizabeth
Google

A typical day in the life of a control engineer

Autoprofit

Sasol's Synthetic Fuel Plant in Secunda

A typical day in the life of a control engineer

Autoprofit

Here's our FT-depropanizer unit:

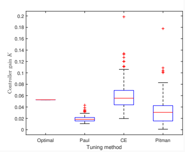
A typical day in the life of a control engineer

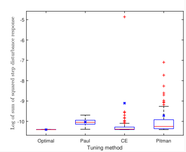
Autoprofit

Here's our FT-depropanizer unit:

An old Dutch proverb: What happened in Secunda stays in Secunda

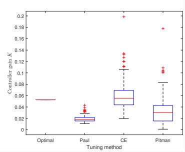
A typical day in the life of a control engineer

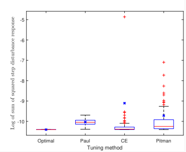




Some observations:

A typical day in the life of a control engineer

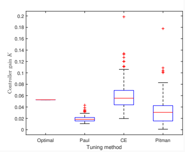


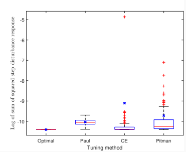


Some observations:

- Clearly the certainty equivalence controller is model based

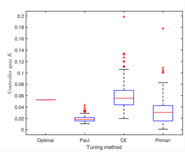
A typical day in the life of a control engineer

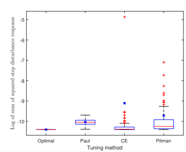




Some observations:

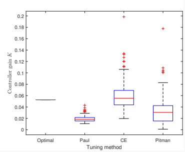
- Clearly the certainty equivalence controller is model based
- The Pitman type controller does not seem to involve an explicit model, although it makes use of a predictor. A direct data driven method?

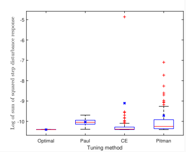




Some observations:

- Clearly the certainty equivalence controller is model based
- The Pitman type controller does not seem to involve an explicit model, although it makes use of a predictor. A direct data driven method?
- Paul's engineering approach can be interpreted as using the model to modify the control criterion. How should it be classified?





Some observations:

- Clearly the certainty equivalence controller is model based
- The Pitman type controller does not seem to involve an explicit model, although it makes use of a predictor. A direct data driven method?
- Paul's engineering approach can be interpreted as using the model to modify the control criterion. How should it be classified?

The essence of this talk:

- Highlight the degree of freedom in data-driven control offered by the tuning of the control criterion

A quick recap of decision theory

- Unknown θ

A quick recap of decision theory

- Unknown θ
- Optimal decision: $h(\theta)$

A quick recap of decision theory

- Unknown θ
- Optimal decision: $h(\theta)$
- Decision: $\hat{h}$

A quick recap of decision theory

- Unknown θ
- Optimal decision: $h(\theta)$
- Decision: $\hat{h}$
- Loss: $L_\theta(\hat{h}) = |\hat{h} - h(\theta)|^2_{W(\theta)}$

A quick recap of decision theory

- Unknown θ
- Optimal decision: $h(\theta)$
- Decision: $\hat{h}$
- Loss: $L_\theta(\hat{h}) = |\hat{h} - h(\theta)|^2_{W(\theta)}$
- Data: Z

A quick recap of decision theory

- Unknown θ
- Optimal decision: $h(\theta)$
- Decision: $\hat{h}$
- Loss: $L_\theta(\hat{h}) = |\hat{h} - h(\theta)|^2_{W(\theta)}$
- Data: Z
    - Provides indirect information of θ

A quick recap of decision theory

- Unknown θ
- Optimal decision: $h(\theta)$
- Decision: $\hat{h}$
- Loss: $L_\theta(\hat{h}) = |\hat{h} - h(\theta)|^2_{W(\theta)}$
- Data: Z
  - Provides indirect information of θ
  - Z modelled as a random variable drawn fom a distribution with pdf $p(z;\theta)$

A quick recap of decision theory

- Unknown θ
- Optimal decision: $h(\theta)$
- Decision: $\hat{h}$
- Loss: $L_\theta(\hat{h}) = |\hat{h} - h(\theta)|^2_{W(\theta)}$
- Data: Z
  - Provides indirect information of θ
  - Z modelled as a random variable drawn fom a distribution with pdf $p(z; \theta)$
- Decision rule: $\hat{h}(Z)$

A quick recap of decision theory

- Unknown θ
- Optimal decision: $h(\theta)$
- Decision: $\hat{h}$
- Loss: $L_\theta(\hat{h}) = |\hat{h} - h(\theta)|^2_{W(\theta)}$
- Data: Z
  - Provides indirect information of θ
  - Z modelled as a random variable drawn fom a distribution with pdf $p(z;\theta)$
- Decision rule: $\hat{h}(Z)$
- Risk: $R_\theta(\hat{h}) = \mathbb{E}_\theta\left[L_\theta(\hat{h}(Z))\right]$

A quick recap of decision theory

- Unknown θ
- Optimal decision: $h(\theta)$
- Decision: $\hat{h}$
- Loss: $L_\theta(\hat{h}) = |\hat{h} - h(\theta)|^2_{W(\theta)}$
- Data: Z
  - Provides indirect information of θ
  - Z modelled as a random variable drawn fom a distribution with pdf $p(z;\theta)$
- Decision rule: $\hat{h}(Z)$
- Risk: $R_\theta(\hat{h}) = \mathbb{E}_\theta\left[L_\theta(\hat{h}(Z))\right]$
- Bias-variance decomp: $R_\theta(\hat{h}) = \underbrace{\|b(\theta)\|^2_{W(\theta)}}_{\text{Bias increase } \geq 0} + \underbrace{\mathrm{Tr}\left\{W(\theta)\mathbb{V}\mathrm{ar}_\theta\left[\hat{h}(Z)\right]\right\}}_{\text{Variance increase } \geq 0}$

  where the bias is $b(\theta) = \mathbb{E}_\theta\left[\hat{h}(Z)\right] - h(\theta)$

A running example: Feedforward control

- System: $y(t) = G(q, \theta)u(t)$

A running example: Feedforward control

- System: $y(t) = G(q, \theta)u(t)$
- Criterion: $J_\theta(\hat{F}) := \mathbb{E}\left[|r(t) - G(q,\theta)u(t)|^2\right] + \lambda\,\mathbb{E}\left[|u(t)|^2\right]$ where $\{r(t)\}$ is stationary and expectation is over r.

A running example: Feedforward control

- System: $y(t) = G(q, \theta)u(t)$
- Criterion: $J_\theta(\hat{F}) := \mathbb{E}\left[|r(t) - G(q,\theta)u(t)|^2\right] + \lambda\,\mathbb{E}\left[|u(t)|^2\right]$ where $\{r(t)\}$ is stationary and expectation is over r.
- Input: $u(t) = \hat{F}(q)r(t)$

A running example: Feedforward control

- System: $y(t) = G(q,\theta)u(t)$
- Criterion: $J_\theta(\hat{F}) := \mathbb{E}\left[|r(t) - G(q,\theta)u(t)|^2\right] + \lambda\,\mathbb{E}\left[|u(t)|^2\right]$ where $\{r(t)\}$ is stationary and expectation is over r.
- Input: $u(t) = \hat{F}(q)r(t)$

A running example: Feedforward control

- System: $y(t) = G(q, \theta)u(t)$
- Criterion: $J_\theta(\hat{F}) := \mathbb{E}\left[|r(t) - G(q,\theta)u(t)|^2\right] + \lambda\,\mathbb{E}\left[|u(t)|^2\right]$ where $\{r(t)\}$ is stationary and expectation is over r.
- Input: $u(t) = \hat{F}(q)r(t)$
- Frequency domain expression: $J_\theta(\hat{F}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} Q_\theta(e^{i\omega}, \hat{F})\Phi_{rr}(e^{i\omega})d\omega$ where $Q_\theta(e^{i\omega}, \hat{F}) := |1 - G(e^{i\omega})\hat{F}(e^{i\omega})|^2 + \lambda|\hat{F}(e^{i\omega})|^2$

A running example: Feedforward control

- System: $y(t) = G(q,\theta)u(t)$
- Criterion: $J_\theta(\hat{F}) := \mathbb{E}\left[|r(t) - G(q,\theta)u(t)|^2\right] + \lambda\,\mathbb{E}\left[|u(t)|^2\right]$ where $\{r(t)\}$ is stationary and expectation is over r.
- Input: $u(t) = \hat{F}(q)r(t)$
- Frequency domain expression: $J_\theta(\hat{F}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} Q_\theta(e^{i\omega},\hat{F})\Phi_{rr}(e^{i\omega})d\omega$ where $Q_\theta(e^{i\omega},\hat{F}) := |1 - G(e^{i\omega})\hat{F}(e^{i\omega})|^2 + \lambda|\hat{F}(e^{i\omega})|^2$
- Optimal controller: $F_\theta(e^{i\omega}) := \frac{\overline{G(e^{i\omega},\theta)}}{|G(e^{i\omega},\theta)|^2+\lambda} \Leftrightarrow h(\theta)$

A running example: Feedforward control

- System: $y(t) = G(q,\theta)u(t)$
- Criterion: $J_\theta(\hat{F}) := \mathbb{E}\left[|r(t) - G(q,\theta)u(t)|^2\right] + \lambda\,\mathbb{E}\left[|u(t)|^2\right]$ where $\{r(t)\}$ is stationary and expectation is over r.
- Input: $u(t) = \hat{F}(q)r(t)$
- Frequency domain expression: $J_\theta(\hat{F}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} Q_\theta(e^{i\omega},\hat{F})\Phi_{rr}(e^{i\omega})d\omega$ where $Q_\theta(e^{i\omega},\hat{F}) := |1 - G(e^{i\omega})\hat{F}(e^{i\omega})|^2 + \lambda|\hat{F}(e^{i\omega})|^2$
- Optimal controller: $F_\theta(e^{i\omega}) := \frac{\overline{G(e^{i\omega},\theta)}}{|G(e^{i\omega},\theta)|^2+\lambda} \Leftrightarrow h(\theta)$
- Minimum cost: $Q_\theta(e^{i\omega},F_\theta) = \frac{\lambda}{|G(e^{i\omega},\theta)|^2+\lambda}$

A running example: Feedforward control

- System: $y(t) = G(q, \theta)u(t)$
- Criterion: $J_\theta(\hat{F}) := \mathbb{E}\left[|r(t) - G(q, \theta)u(t)|^2\right] + \lambda\, \mathbb{E}\left[|u(t)|^2\right]$ where $\{r(t)\}$ is stationary and expectation is over r.
- Input: $u(t) = \hat{F}(q)r(t)$
- Frequency domain expression: $J_\theta(\hat{F}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} Q_\theta(e^{i\omega}, \hat{F})\Phi_{rr}(e^{i\omega})d\omega$ where $Q_\theta(e^{i\omega}, \hat{F}) := |1 - G(e^{i\omega})\hat{F}(e^{i\omega})|^2 + \lambda|\hat{F}(e^{i\omega})|^2$
- Optimal controller: $F_\theta(e^{i\omega}) := \frac{\overline{G(e^{i\omega}, \theta)}}{|G(e^{i\omega}, \theta)|^2 + \lambda} \Leftrightarrow h(\theta)$
- Minimum cost: $Q_\theta(e^{i\omega}, F_\theta) = \frac{\lambda}{|G(e^{i\omega}, \theta)|^2 + \lambda}$
- Decision $\hat{F}(e^{i\omega}) \Leftrightarrow \hat{h}$

A running example: Feedforward control

- System: $y(t) = G(q,\theta)u(t)$
- Criterion: $J_\theta(\hat{F}) := \mathbb{E}\left[|r(t) - G(q,\theta)u(t)|^2\right] + \lambda\,\mathbb{E}\left[|u(t)|^2\right]$ where $\{r(t)\}$ is stationary and expectation is over r.
- Input: $u(t) = \hat{F}(q)r(t)$
- Frequency domain expression: $J_\theta(\hat{F}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} Q_\theta(e^{i\omega},\hat{F})\Phi_{rr}(e^{i\omega})d\omega$ where $Q_\theta(e^{i\omega},\hat{F}) := |1 - G(e^{i\omega})\hat{F}(e^{i\omega})|^2 + \lambda|\hat{F}(e^{i\omega})|^2$
- Optimal controller: $F_\theta(e^{i\omega}) := \frac{\overline{G(e^{i\omega},\theta)}}{|G(e^{i\omega},\theta)|^2+\lambda} \Leftrightarrow h(\theta)$
- Minimum cost: $Q_\theta(e^{i\omega},F_\theta) = \frac{\lambda}{|G(e^{i\omega},\theta)|^2+\lambda}$
- Decision $\hat{F}(e^{i\omega}) \Leftrightarrow \hat{h}$
- Per frequency loss: $L_\theta(\hat{F}(e^{i\omega})) := \underbrace{(|G(e^{i\omega},\theta)|^2 + \lambda)}_{W(\theta)}|\hat{F}(e^{i\omega}) - F_\theta(e^{i\omega})|^2$

A running example: Feedforward control

- System: $y(t) = G(q,\theta)u(t)$
- Criterion: $J_\theta(\hat{F}) := \mathbb{E}\left[|r(t) - G(q,\theta)u(t)|^2\right] + \lambda\,\mathbb{E}\left[|u(t)|^2\right]$ where $\{r(t)\}$ is stationary and expectation is over r.
- Input: $u(t) = \hat{F}(q)r(t)$
- Frequency domain expression: $J_\theta(\hat{F}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} Q_\theta(e^{i\omega},\hat{F})\Phi_{rr}(e^{i\omega})d\omega$ where $Q_\theta(e^{i\omega},\hat{F}) := |1 - G(e^{i\omega})\hat{F}(e^{i\omega})|^2 + \lambda|\hat{F}(e^{i\omega})|^2$
- Optimal controller: $F_\theta(e^{i\omega}) := \frac{\overline{G(e^{i\omega},\theta)}}{|G(e^{i\omega},\theta)|^2+\lambda} \Leftrightarrow h(\theta)$
- Minimum cost: $Q_\theta(e^{i\omega},F_\theta) = \frac{\lambda}{|G(e^{i\omega},\theta)|^2+\lambda}$
- Decision $\hat{F}(e^{i\omega}) \Leftrightarrow \hat{h}$
- Per frequency loss: $L_\theta(\hat{F}(e^{i\omega})) := \underbrace{(|G(e^{i\omega},\theta)|^2+\lambda)}_{W(\theta)}|\hat{F}(e^{i\omega}) - F_\theta(e^{i\omega})|^2$
- Frequency domain criterion: $Q_\theta(e^{i\omega},\hat{F}) = L_\theta(\hat{F}(e^{i\omega})) + Q_\theta(e^{i\omega},F_\theta)$

A running example: Feedforward control

- System: $y(t) = G(q,\theta)u(t)$
- Criterion: $J_\theta(\hat{F}) := \mathbb{E}\left[|r(t) - G(q,\theta)u(t)|^2\right] + \lambda\,\mathbb{E}\left[|u(t)|^2\right]$ where $\{r(t)\}$ is stationary and expectation is over r.
- Input: $u(t) = \hat{F}(q)r(t)$
- Frequency domain expression: $J_\theta(\hat{F}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} Q_\theta(e^{i\omega},\hat{F})\Phi_{rr}(e^{i\omega})d\omega$ where $Q_\theta(e^{i\omega},\hat{F}) := |1 - G(e^{i\omega})\hat{F}(e^{i\omega})|^2 + \lambda|\hat{F}(e^{i\omega})|^2$
- Optimal controller: $F_\theta(e^{i\omega}) := \frac{\overline{G(e^{i\omega},\theta)}}{|G(e^{i\omega},\theta)|^2+\lambda} \Leftrightarrow h(\theta)$
- Minimum cost: $Q_\theta(e^{i\omega},F_\theta) = \frac{\lambda}{|G(e^{i\omega},\theta)|^2+\lambda}$
- Decision $\hat{F}(e^{i\omega}) \Leftrightarrow \hat{h}$
- Per frequency loss: $L_\theta(\hat{F}(e^{i\omega})) := \underbrace{(|G(e^{i\omega},\theta)|^2+\lambda)}_{W(\theta)}|\hat{F}(e^{i\omega}) - F_\theta(e^{i\omega})|^2$
- Frequency domain criterion: $Q_\theta(e^{i\omega},\hat{F}) = L_\theta(\hat{F}(e^{i\omega})) + Q_\theta(e^{i\omega},F_\theta)$
- Per frequency risk: $R_\theta(\hat{F}(e^{i\omega},Z)) := \mathbb{E}_\theta\left[L_\theta(\hat{F}(e^{i\omega},Z))\right]$

A running example: Feedforward control

- System: $y(t) = G(q,\theta)u(t)$
- Criterion: $J_\theta(\hat{F}) := \mathbb{E}\left[|r(t) - G(q,\theta)u(t)|^2\right] + \lambda\,\mathbb{E}\left[|u(t)|^2\right]$ where $\{r(t)\}$ is stationary and expectation is over r.
- Input: $u(t) = \hat{F}(q)r(t)$
- Frequency domain expression: $J_\theta(\hat{F}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} Q_\theta(e^{i\omega},\hat{F})\Phi_{rr}(e^{i\omega})d\omega$ where $Q_\theta(e^{i\omega},\hat{F}) := |1 - G(e^{i\omega})\hat{F}(e^{i\omega})|^2 + \lambda|\hat{F}(e^{i\omega})|^2$
- Optimal controller: $F_\theta(e^{i\omega}) := \frac{\overline{G(e^{i\omega},\theta)}}{|G(e^{i\omega},\theta)|^2+\lambda} \Leftrightarrow h(\theta)$
- Minimum cost: $Q_\theta(e^{i\omega}, F_\theta) = \frac{\lambda}{|G(e^{i\omega},\theta)|^2+\lambda}$
- Decision $\hat{F}(e^{i\omega}) \Leftrightarrow \hat{h}$
- Per frequency loss: $L_\theta(\hat{F}(e^{i\omega})) := \underbrace{(|G(e^{i\omega},\theta)|^2 + \lambda)}_{W(\theta)}|\hat{F}(e^{i\omega}) - F_\theta(e^{i\omega})|^2$
- Frequency domain criterion: $Q_\theta(e^{i\omega},\hat{F}) = L_\theta(\hat{F}(e^{i\omega})) + Q_\theta(e^{i\omega},F_\theta)$
- Per frequency risk: $R_\theta(\hat{F}(e^{i\omega},Z)) := \mathbb{E}_\theta\left[L_\theta(\hat{F}(e^{i\omega},Z))\right]$
- Total loss: $L_\theta(\hat{F}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} L_\theta(\hat{F}(e^{i\omega}))d\omega$

A running example: Feedforward control

- System: $y(t) = G(q, \theta)u(t)$
- Criterion: $J_\theta(\hat{F}) := \mathbb{E}\left[|r(t) - G(q,\theta)u(t)|^2\right] + \lambda\,\mathbb{E}\left[|u(t)|^2\right]$ where $\{r(t)\}$ is stationary and expectation is over r.
- Input: $u(t) = \hat{F}(q)r(t)$
- Frequency domain expression: $J_\theta(\hat{F}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} Q_\theta(e^{i\omega}, \hat{F})\Phi_{rr}(e^{i\omega})d\omega$ where $Q_\theta(e^{i\omega}, \hat{F}) := |1 - G(e^{i\omega})\hat{F}(e^{i\omega})|^2 + \lambda|\hat{F}(e^{i\omega})|^2$
- Optimal controller: $F_\theta(e^{i\omega}) := \frac{\overline{G(e^{i\omega},\theta)}}{|G(e^{i\omega},\theta)|^2+\lambda} \Leftrightarrow h(\theta)$
- Minimum cost: $Q_\theta(e^{i\omega}, F_\theta) = \frac{\lambda}{|G(e^{i\omega},\theta)|^2+\lambda}$
- Decision $\hat{F}(e^{i\omega}) \Leftrightarrow \hat{h}$
- Per frequency loss: $L_\theta(\hat{F}(e^{i\omega})) := \underbrace{(|G(e^{i\omega},\theta)|^2 + \lambda)}_{W(\theta)}|\hat{F}(e^{i\omega}) - F_\theta(e^{i\omega})|^2$
- Frequency domain criterion: $Q_\theta(e^{i\omega}, \hat{F}) = L_\theta(\hat{F}(e^{i\omega})) + Q_\theta(e^{i\omega}, F_\theta)$
- Per frequency risk: $R_\theta(\hat{F}(e^{i\omega}, Z)) := \mathbb{E}_\theta\left[L_\theta(\hat{F}(e^{i\omega}, Z))\right]$
- Total loss: $L_\theta(\hat{F}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} L_\theta(\hat{F}(e^{i\omega}))d\omega$
- Total risk: $R_\theta(\hat{F}(Z)) = \frac{1}{2\pi}\int_{-\pi}^{\pi} R_\theta(\hat{F}(e^{i\omega}, Z))d\omega = \mathbb{E}_\theta\left[L_\theta(\hat{F}(Z))\right]$

Performance bounds

- No uniformly best decision rule. Optimal controller for a given system $G_\circ$ has zero risk for that system and no truly data-dependent controller can achieve this

Design principles

- Unbiased decision rules: $\mathbb{E}_\theta\left[\hat{h}(Z)\right] = h(\theta)$.
  Uniformly minimum risk unbiased decision rule: UMRU

Design principles

- Unbiased decision rules: $\mathbb{E}_\theta\left[\hat{h}(Z)\right] = h(\theta)$.
  Uniformly minimum risk unbiased decision rule: UMRU
- Loss tuning: $\hat{h}_{\mathrm{CE}(\hat{\theta})}(Z) = \operatorname{arg\,min}_{\hat{h}} L_{\hat{\theta}(Z)}(\hat{h}) = \operatorname{arg\,min}_{\hat{h}} W(\hat{\theta}(Z))|\hat{h} - h(\hat{\theta}(Z)|^2$
  $\Rightarrow \quad \hat{h}_{\mathrm{CE}(\hat{\theta})}(Z) = h(\hat{\theta}(Z))$

Design principles

- Unbiased decision rules: $\mathbb{E}_\theta\left[\hat{h}(Z)\right] = h(\theta)$.
  Uniformly minimum risk unbiased decision rule: UMRU
- Loss tuning: $\hat{h}_{\mathrm{CE}(\hat{\theta})}(Z) = \arg\min_{\hat{h}} L_{\hat{\theta}(Z)}(\hat{h}) = \arg\min_{\hat{h}} W(\hat{\theta}(Z))|\hat{h} - h(\hat{\theta}(Z)|^2$
  $\Rightarrow \quad \hat{h}_{\mathrm{CE}(\hat{\theta})}(Z) = h(\hat{\theta}(Z))$

Certainty equivalence decision rule

Design principles

- Unbiased decision rules: $\mathbb{E}_\theta\left[\hat{h}(Z)\right] = h(\theta)$.
  Uniformly minimum risk unbiased decision rule: UMRU
- Loss tuning: $\hat{h}_{\mathrm{CE}(\hat{\theta})}(Z) = \arg\min_{\hat{h}} L_{\hat{\theta}(Z)}(\hat{h}) = \arg\min_{\hat{h}} W(\hat{\theta}(Z))|\hat{h} - h(\hat{\theta}(Z)|^2$
  $\Rightarrow \quad \hat{h}_{\mathrm{CE}(\hat{\theta})}(Z) = h(\hat{\theta}(Z))$

Certainty equivalence decision rule

- Average risk tuning: Weighting losses for different models more robust:

$$\hat{h}_{\mathrm{A}(\pi)}(z) := \arg\min_{\hat{h}} \int L_\theta(\hat{h})\, w(\theta; z)\, d\theta$$

Design principles

- Unbiased decision rules: $\mathbb{E}_\theta \left[\hat{h}(Z)\right] = h(\theta)$.
  Uniformly minimum risk unbiased decision rule: UMRU
- Loss tuning: $\hat{h}_{\mathrm{CE}(\hat{\theta})}(Z) = \arg\min_{\hat{h}} L_{\hat{\theta}(Z)}(\hat{h}) = \arg\min_{\hat{h}} W(\hat{\theta}(Z))|\hat{h} - h(\hat{\theta}(Z)|^2$
  $\Rightarrow \quad \hat{h}_{\mathrm{CE}(\hat{\theta})}(Z) = h(\hat{\theta}(Z))$

Certainty equivalence decision rule

- Average risk tuning: Weighting losses for different models more robust:

$$\hat{h}_{\mathrm{A}(\pi)}(z) := \arg\min_{\hat{h}} \int L_\theta(\hat{h})\, w(\theta; z)\, d\theta$$

  How to choose $w(z;\theta)$?

Design principles

- Unbiased decision rules: $\mathbb{E}_\theta\left[\hat{h}(Z)\right] = h(\theta)$.
  Uniformly minimum risk unbiased decision rule: UMRU
- Loss tuning: $\hat{h}_{\mathrm{CE}(\hat{\theta})}(Z) = \arg\min_{\hat{h}} L_{\hat{\theta}(Z)}(\hat{h}) = \arg\min_{\hat{h}} W(\hat{\theta}(Z))|\hat{h} - h(\hat{\theta}(Z)|^2$
  $\Rightarrow \quad \hat{h}_{\mathrm{CE}(\hat{\theta})}(Z) = h(\hat{\theta}(Z))$

  Certainty equivalence decision rule

- Average risk tuning: Weighting losses for different models more robust:

$$\hat{h}_{\mathrm{A}(\pi)}(z) := \arg\min_{\hat{h}} \int L_\theta(\hat{h})\, w(\theta; z)\, d\theta$$

  How to choose $w(z;\theta)$? Need to link data to θ.

Design principles

- Unbiased decision rules: $\mathbb{E}_\theta\left[\hat{h}(Z)\right] = h(\theta)$.
  Uniformly minimum risk unbiased decision rule: UMRU
- Loss tuning: $\hat{h}_{\mathrm{CE}(\hat{\theta})}(Z) = \operatorname{arg\,min}_{\hat{h}} L_{\hat{\theta}(Z)}(\hat{h}) = \operatorname{arg\,min}_{\hat{h}} W(\hat{\theta}(Z))|\hat{h} - h(\hat{\theta}(Z)|^2$
  $\Rightarrow \quad \hat{h}_{\mathrm{CE}(\hat{\theta})}(Z) = h(\hat{\theta}(Z))$

Certainty equivalence decision rule

- Average risk tuning: Weighting losses for different models more robust:

$$\hat{h}_{\mathrm{A}(\pi)}(z) := \operatorname*{arg\,min}_{\hat{h}} \int L_\theta(\hat{h})\, w(\theta; z)\, d\theta$$

How to choose $w(z;\theta)$? Need to link data to θ. $p(z;\theta)$ only link

Design principles

- Unbiased decision rules: $\mathbb{E}_\theta\left[\hat{h}(Z)\right] = h(\theta)$.
  Uniformly minimum risk unbiased decision rule: UMRU
- Loss tuning: $\hat{h}_{\mathrm{CE}(\hat{\theta})}(Z) = \arg\min_{\hat{h}} L_{\hat{\theta}(Z)}(\hat{h}) = \arg\min_{\hat{h}} W(\hat{\theta}(Z))|\hat{h} - h(\hat{\theta}(Z)|^2$
  $\Rightarrow \quad \hat{h}_{\mathrm{CE}(\hat{\theta})}(Z) = h(\hat{\theta}(Z))$

Certainty equivalence decision rule

- Average risk tuning: Weighting losses for different models more robust:

$$\hat{h}_{\mathrm{A}(\pi)}(z) := \arg\min_{\hat{h}} \int L_\theta(\hat{h})\, w(\theta; z)\, d\theta$$

  How to choose $w(z;\theta)$? Need to link data to θ. $p(z;\theta)$ only link

  - Pitman type: $w(z;\theta) = p(z;\theta)$

Design principles

- Unbiased decision rules: $\mathbb{E}_\theta\left[\hat{h}(Z)\right] = h(\theta)$.
  Uniformly minimum risk unbiased decision rule: UMRU
- Loss tuning: $\hat{h}_{\mathrm{CE}(\hat{\theta})}(Z) = \arg\min_{\hat{h}} L_{\hat{\theta}(Z)}(\hat{h}) = \arg\min_{\hat{h}} W(\hat{\theta}(Z))|\hat{h} - h(\hat{\theta}(Z)|^2$
  $\Rightarrow \quad \hat{h}_{\mathrm{CE}(\hat{\theta})}(Z) = h(\hat{\theta}(Z))$

Certainty equivalence decision rule

- Average risk tuning: Weighting losses for different models more robust:

$$\hat{h}_{\mathrm{A}(\pi)}(z) := \arg\min_{\hat{h}} \int L_\theta(\hat{h})\, w(\theta; z)\, d\theta$$

How to choose $w(z;\theta)$? Need to link data to θ. $p(z;\theta)$ only link

  - Pitman type: $w(z;\theta) = p(z;\theta)$
  - Weighting: $w(z;\theta) = p(z;\theta)\pi(\theta)$ (Bayesian interpretation)

Design principles

- Unbiased decision rules: $\mathbb{E}_\theta\left[\hat{h}(Z)\right] = h(\theta)$.
  Uniformly minimum risk unbiased decision rule: UMRU
- Loss tuning: $\hat{h}_{\mathrm{CE}(\hat{\theta})}(Z) = \arg\min_{\hat{h}} L_{\hat{\theta}(Z)}(\hat{h}) = \arg\min_{\hat{h}} W(\hat{\theta}(Z))|\hat{h} - h(\hat{\theta}(Z)|^2$
  $\Rightarrow \quad \hat{h}_{\mathrm{CE}(\hat{\theta})}(Z) = h(\hat{\theta}(Z))$

Certainty equivalence decision rule

- Average risk tuning: Weighting losses for different models more robust:

$$\hat{h}_{\mathrm{A}(\pi)}(z) := \arg\min_{\hat{h}} \int L_\theta(\hat{h})\, w(\theta; z)\, d\theta$$

How to choose $w(z;\theta)$? Need to link data to θ. $p(z;\theta)$ only link

  - Pitman type: $w(z;\theta) = p(z;\theta)$
  - Weighting: $w(z;\theta) = p(z;\theta)\pi(\theta)$ (Bayesian interpretation)
  - Note π does not have to be integrable (generalized average risk decision rules, cf. Pitman type)

Design principles

- Minimax decision rules: $\hat{h}_{\mathrm{wc}} = \arg\min_{\hat{h}} \sup_{\theta \in D_\theta} R_\theta(\hat{h})$

Design principles

- Minimax decision rules: $\hat{h}_{\mathrm{wc}} = \arg\min_{\hat{h}} \sup_{\theta \in D_\theta} R_\theta(\hat{h})$
- Risk tuning:
  - Estimate the risk R_θ and minimize the estimate $\hat{R}(\hat{h}, Z)$ wrt $\hat{h}$

Design principles

- Minimax decision rules: $\hat{h}_{\mathrm{wc}} = \arg\min_{\hat{h}} \sup_{\theta \in D_\theta} R_\theta(\hat{h})$
- Risk tuning:
  - Estimate the risk R_θ and minimize the estimate $\hat{R}(\hat{h}, Z)$ wrt $\hat{h}$
  - but this is also a decision problem ($h(\theta) = R_\theta(\hat{h})$)

Design principles

- Minimax decision rules: $\hat{h}_{\mathrm{wc}} = \arg\min_{\hat{h}} \sup_{\theta \in D_\theta} R_\theta(\hat{h})$
- Risk tuning:
  - Estimate the risk R_θ and minimize the estimate $\hat{R}(\hat{h}, Z)$ wrt $\hat{h}$
  - but this is also a decision problem $(h(\theta) = R_\theta(\hat{h}))$
  - All previous methods can be used!

Design principles

- Minimax decision rules: $\hat{h}_{\mathrm{wc}} = \arg\min_{\hat{h}} \sup_{\theta \in D_\theta} R_\theta(\hat{h})$
- Risk tuning:
    - Estimate the risk R_θ and minimize the estimate $\hat{R}(\hat{h}, Z)$ wrt $\hat{h}$
    - but this is also a decision problem ($h(\theta) = R_\theta(\hat{h})$)
    - All previous methods can be used!
        - Certainty equivalence tuning: $\hat{h}_{\mathrm{R}(\hat{\theta}(Z))} := \arg\min_{\hat{h} \in D_h} R_{\hat{\theta}(Z)}(\hat{h})$

Design principles

- Minimax decision rules: $\hat{h}_{\mathrm{wc}} = \operatorname{arg\,min}_{\hat{h}} \sup_{\theta \in D_\theta} R_\theta(\hat{h})$
- Risk tuning:
  - Estimate the risk R_θ and minimize the estimate $\hat{R}(\hat{h}, Z)$ wrt $\hat{h}$
  - but this is also a decision problem $(h(\theta) = R_\theta(\hat{h}))$
  - All previous methods can be used!
    - Certainty equivalence tuning: $\hat{h}_{\mathrm{R}(\hat{\theta}(Z))} := \operatorname{arg\,min}_{\hat{h} \in D_h} R_{\hat{\theta}(Z)}(\hat{h})$
    - Unbiased risk tuning: Example Stein's unbiased risk estimate (SURE)

Design principles

- Minimax decision rules: $\hat{h}_{\mathrm{wc}} = \arg\min_{\hat{h}} \sup_{\theta \in D_\theta} R_\theta(\hat{h})$
- Risk tuning:
  - Estimate the risk R_θ and minimize the estimate $\hat{R}(\hat{h}, Z)$ wrt $\hat{h}$
  - but this is also a decision problem $(h(\theta) = R_\theta(\hat{h}))$
  - All previous methods can be used!
    - Certainty equivalence tuning: $\hat{h}_{\mathrm{R}(\hat{\theta}(Z))} := \arg\min_{\hat{h} \in D_h} R_{\hat{\theta}(Z)}(\hat{h})$
    - Unbiased risk tuning: Example Stein's unbiased risk estimate (SURE)
    - Average risk tuning: Take $L_\theta(\hat{R}) = W_{\mathrm{R}}(\theta)|\hat{R} - R_\theta(\hat{h})|^2$ as loss

Design principles

- Minimax decision rules: $\hat{h}_{\mathrm{wc}} = \arg\min_{\hat{h}} \sup_{\theta \in D_\theta} R_\theta(\hat{h})$
- Risk tuning:
  - Estimate the risk R_θ and minimize the estimate $\hat{R}(\hat{h}, Z)$ wrt $\hat{h}$
  - but this is also a decision problem $(h(\theta) = R_\theta(\hat{h}))$
  - All previous methods can be used!
    - Certainty equivalence tuning: $\hat{h}_{\mathrm{R}(\hat{\theta}(Z))} := \arg\min_{\hat{h} \in D_h} R_{\hat{\theta}(Z)}(\hat{h})$
    - Unbiased risk tuning: Example Stein's unbiased risk estimate (SURE)
    - Average risk tuning: Take $L_\theta(\hat{R}) = W_{\mathrm{R}}(\theta)|\hat{R} - R_\theta(\hat{h})|^2$ as loss
- Composite methods: Combine all the above!

Design principles

- Minimax decision rules: $\hat{h}_{\mathrm{wc}} = \operatorname{arg\,min}_{\hat{h}} \sup_{\theta \in D_\theta} R_\theta(\hat{h})$
- Risk tuning:
  - Estimate the risk R_θ and minimize the estimate $\hat{R}(\hat{h}, Z)$ wrt $\hat{h}$
  - but this is also a decision problem $(h(\theta) = R_\theta(\hat{h}))$
  - All previous methods can be used!
    - Certainty equivalence tuning: $\hat{h}_{\mathrm{R}(\hat{\theta}(Z))} := \operatorname{arg\,min}_{\hat{h} \in D_h} R_{\hat{\theta}(Z)}(\hat{h})$
    - Unbiased risk tuning: Example Stein's unbiased risk estimate (SURE)
    - Average risk tuning: Take $L_\theta(\hat{R}) = W_{\mathrm{R}}(\theta)|\hat{R} - R_\theta(\hat{h})|^2$ as loss
- Composite methods: Combine all the above!
- HELP!

Design guidelines

- A decision rule $\hat{h}$ is said to be *admissible* if there is no other decision rule that improves on it (has at least as low risk for all systems, and lower for some systems)

Design guidelines

- A decision rule $\hat{h}$ is said to be *admissible* if there is no other decision rule that improves on it (has at least as low risk for all systems, and lower for some systems)
- A class $\mathcal{A}$ of decision rules is *complete* if for any decision rule $\hat{h}$ not in $\mathcal{A}$ there exists a decision rule $\hat{\hbar}$ in $\mathcal{A}$ that improves on it

Design guidelines

- A decision rule $\hat{h}$ is said to be *admissible* if there is no other decision rule that improves on it (has at least as low risk for all systems, and lower for some systems)
- A class $\mathcal{A}$ of decision rules is *complete* if for any decision rule $\hat{h}$ not in $\mathcal{A}$ there exists a decision rule $\hat{\hbar}$ in $\mathcal{A}$ that improves on it
- Sufficient statistic $T(Z)$: $p(Z|T(Z))$ does not depend on θ

Design guidelines

- A decision rule $\hat{h}$ is said to be *admissible* if there is no other decision rule that improves on it (has at least as low risk for all systems, and lower for some systems)
- A class $\mathcal{A}$ of decision rules is *complete* if for any decision rule $\hat{h}$ not in $\mathcal{A}$ there exists a decision rule $\hat{\hbar}$ in $\mathcal{A}$ that improves on it
- Sufficient statistic $T(Z)$: $p(Z|T(Z))$ does not depend on θ
  - Compresses data without information loss

Design guidelines

- A decision rule $\hat{h}$ is said to be *admissible* if there is no other decision rule that improves on it (has at least as low risk for all systems, and lower for some systems)
- A class $\mathcal{A}$ of decision rules is *complete* if for any decision rule $\hat{h}$ not in $\mathcal{A}$ there exists a decision rule $\hat{\hbar}$ in $\mathcal{A}$ that improves on it
- Sufficient statistic $T(Z)$: $p(Z|T(Z))$ does not depend on θ
    - Compresses data without information loss
    - Minimal sufficient statistics compresses data the most

Design guidelines

- A decision rule $\hat{h}$ is said to be *admissible* if there is no other decision rule that improves on it (has at least as low risk for all systems, and lower for some systems)
- A class $\mathcal{A}$ of decision rules is *complete* if for any decision rule $\hat{h}$ not in $\mathcal{A}$ there exists a decision rule $\hat{\hbar}$ in $\mathcal{A}$ that improves on it
- Sufficient statistic $T(Z)$: $p(Z|T(Z))$ does not depend on θ
    - Compresses data without information loss
    - Minimal sufficient statistics compresses data the most
    - For any decision rule $\hat{h}(Z)$ there exists $\hat{\hbar}(T)$ which has the same risk as $\hat{h}(Z)$.

Design guidelines

- A decision rule $\hat{h}$ is said to be *admissible* if there is no other decision rule that improves on it (has at least as low risk for all systems, and lower for some systems)
- A class $\mathcal{A}$ of decision rules is *complete* if for any decision rule $\hat{h}$ not in $\mathcal{A}$ there exists a decision rule $\hat{\hbar}$ in $\mathcal{A}$ that improves on it
- Sufficient statistic $T(Z)$: $p(Z|T(Z))$ does not depend on θ
    - Compresses data without information loss
    - Minimal sufficient statistics compresses data the most
    - For any decision rule $\hat{h}(Z)$ there exists $\hat{\hbar}(T)$ which has the same risk as $\hat{h}(Z)$.
    - Let $\hat{h}(Z)$ be an arbitrary decision rule and T a suff. stat. Then $\hat{h}_T(Z) := \mathbb{E}\left[\hat{h}(Z)|T(Z)\right]$ improves on $\hat{h}$ unless $\hat{h}(Z) = \hat{h}_T(T(Z))$ with probability 1.

Design guidelines

- A decision rule $\hat{h}$ is said to be *admissible* if there is no other decision rule that improves on it (has at least as low risk for all systems, and lower for some systems)
- A class $\mathcal{A}$ of decision rules is *complete* if for any decision rule $\hat{h}$ not in $\mathcal{A}$ there exists a decision rule $\hat{\hbar}$ in $\mathcal{A}$ that improves on it
- Sufficient statistic $T(Z)$: $p(Z|T(Z))$ does not depend on θ
  - Compresses data without information loss
  - Minimal sufficient statistics compresses data the most
  - For any decision rule $\hat{h}(Z)$ there exists $\hat{\hbar}(T)$ which has the same risk as $\hat{h}(Z)$.
  - Let $\hat{h}(Z)$ be an arbitrary decision rule and T a suff. stat. Then $\hat{h}_T(Z) := \mathbb{E}\left[\hat{h}(Z)|T(Z)\right]$ improves on $\hat{h}$ unless $\hat{h}(Z) = \hat{h}_T(T(Z))$ with probability 1.
  - Largest improvement when using a minimal sufficient statistic

Design guidelines

- A decision rule $\hat{h}$ is said to be *admissible* if there is no other decision rule that improves on it (has at least as low risk for all systems, and lower for some systems)
- A class $\mathcal{A}$ of decision rules is *complete* if for any decision rule $\hat{h}$ not in $\mathcal{A}$ there exists a decision rule $\hat{\hbar}$ in $\mathcal{A}$ that improves on it
- Sufficient statistic $T(Z)$: $p(Z|T(Z))$ does not depend on θ
  - Compresses data without information loss
  - Minimal sufficient statistics compresses data the most
  - For any decision rule $\hat{h}(Z)$ there exists $\hat{\hbar}(T)$ which has the same risk as $\hat{h}(Z)$.
  - Let $\hat{h}(Z)$ be an arbitrary decision rule and T a suff. stat. Then $\hat{h}_T(Z) := \mathbb{E}\left[\hat{h}(Z)|T(Z)\right]$ improves on $\hat{h}$ unless $\hat{h}(Z) = \hat{h}_T(T(Z))$ with probability 1.
  - Largest improvement when using a minimal sufficient statistic
  - Example: FIR model with Gaussian noise and known noise variance: $\hat{\theta}_{\mathrm{ML}}$ minimal sufficient

Design guidelines

- Average risk decision rules
    - Admissible if $p(z; \theta)$ continuous on D_θ for all z, and π has support D_θ.

Design guidelines

- Average risk decision rules
  - Admissible if $p(z;\theta)$ continuous on D_θ for all z, and π has support D_θ.
  - For exponential families, every admissible decision rule is a generalized average risk decision rule and they form a complete class.

Design guidelines

- Average risk decision rules
  - Admissible if $p(z;\theta)$ continuous on D_θ for all z, and π has support D_θ.
  - For exponential families, every admissible decision rule is a generalized average risk decision rule and they form a complete class.
  - Asymptotically efficient (under regularity conditions)

Design guidelines

Summary: Seems like a good idea to

- Restrict attention to a complete class

Summary: Seems like a good idea to

- Restrict attention to a complete class
- Use minimal sufficient statistics
  (unfortunately, sometimes only one to one transformations of Z may be the only available sufficient statistics)

Summary: Seems like a good idea to

- Restrict attention to a complete class
- Use minimal sufficient statistics
  (unfortunately, sometimes only one to one transformations of Z may be the only available sufficient statistics)
- (Generalized) average risk decision rules attractive class (admissible, asymptotically efficient, complete class for exponential families)

Feedforward control

Back to our running example:

- System: $y(t) = G(q,\theta)u(t)$, controller $u(t) = \hat{F}(q)r(t)$
- Frequency domain expression: $J_\theta(\hat{F}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} Q_\theta(e^{i\omega},\hat{F})\Phi_{rr}(e^{i\omega})d\omega$ where $Q_\theta(e^{i\omega},\hat{F}) := |1 - G(e^{i\omega},\theta)\hat{F}(e^{i\omega})|^2 + \lambda|\hat{F}(e^{i\omega})|^2$
- Optimal controller: $F_\theta(e^{i\omega}) := \frac{\overline{G(e^{i\omega},\theta)}}{|G(e^{i\omega},\theta)|^2+\lambda} \Leftrightarrow h(\theta)$
- Minimum cost: $Q_\theta(e^{i\omega}, F_\theta) = \frac{\lambda}{|G(e^{i\omega},\theta)|^2+\lambda}$
- Decision $\hat{F}(e^{i\omega}) \Leftrightarrow \hat{h}$
- Per frequency loss: $L_\theta(\hat{F}(e^{i\omega})) := \underbrace{(|G(e^{i\omega},\theta)|^2 + \lambda)}_{W(\theta)}|\hat{F}(e^{i\omega}) - F_\theta(e^{i\omega})|^2$
- Per frequency risk: $R_\theta(\hat{F}(e^{i\omega},Z)) := \mathbb{E}_\theta\left[L_\theta(\hat{F}(e^{i\omega},Z))\right]$

Suppose the data is generated from the FIR-system

$$y(t) = \varphi^T(t)\theta + e(t)$$

where $e(t)$ is white Gaussian with known variance σ^2.

Feedforward control

Suppose the data is generated from the FIR-system

$$y(t) = \varphi^T(t)\theta + e(t)$$

where $e(t)$ is white Gaussian with known variance σ^2.

What have we learnt?

- $\hat{\theta}_{\mathrm{ML}}(Z)$ is a minimal sufficient statistic

Suppose the data is generated from the FIR-system

$$y(t) = \varphi^T(t)\theta + e(t)$$

where $e(t)$ is white Gaussian with known variance σ^2.

What have we learnt?

- $\hat{\theta}_{\mathrm{ML}}(Z)$ is a minimal sufficient statistic

$\Rightarrow$ Can consider $\hat{\theta}_{\mathrm{ML}}(Z)$ as our data

Suppose the data is generated from the FIR-system

$$y(t) = \varphi^T(t)\theta + e(t)$$

where $e(t)$ is white Gaussian with known variance σ^2.

What have we learnt?

- $\hat{\theta}_{\mathrm{ML}}(Z)$ is a minimal sufficient statistic

$\Rightarrow$ Can consider $\hat{\theta}_{\mathrm{ML}}(Z)$ as our data

$\Rightarrow$ Need only consider controllers of the form $\hat{F}(q, Z) = \hat{F}(z, \hat{\theta}_{\mathrm{ML}})$

Suppose the data is generated from the FIR-system

$$y(t) = \varphi^T(t)\theta + e(t)$$

where $e(t)$ is white Gaussian with known variance σ^2.

What have we learnt?

- $\hat{\theta}_{\mathrm{ML}}(Z)$ is a minimal sufficient statistic

$\Rightarrow$ Can consider $\hat{\theta}_{\mathrm{ML}}(Z)$ as our data

$\Rightarrow$ Need only consider controllers of the form $\hat{F}(q, Z) = \hat{F}(z, \hat{\theta}_{\mathrm{ML}})$

- $\hat{\theta}_{\mathrm{ML}}(Z) \sim \mathcal{N}(\theta, C)$, $C = \sigma^2(\sum_{t=1}^N \varphi(t)\varphi^T(t))^{-1}$

Suppose the data is generated from the FIR-system

$$y(t) = \varphi^T(t)\theta + e(t)$$

where $e(t)$ is white Gaussian with known variance σ^2.

What have we learnt?

- $\hat{\theta}_{\mathrm{ML}}(Z)$ is a minimal sufficient statistic

$\Rightarrow$ Can consider $\hat{\theta}_{\mathrm{ML}}(Z)$ as our data

$\Rightarrow$ Need only consider controllers of the form $\hat{F}(q, Z) = \hat{F}(z, \hat{\theta}_{\mathrm{ML}})$

- $\hat{\theta}_{\mathrm{ML}}(Z) \sim \mathcal{N}(\theta, C)$, $C = \sigma^2(\sum_{t=1}^N \varphi(t)\varphi^T(t))^{-1}$

$\Rightarrow$ All admissible controllers are (generalized) average risk controllers which form a complete class

Average risk tuning:

$$\arg\min_{\hat{h}} \int L_\theta(\hat{h})\, w(\theta; z)\, d\theta, \quad L_\theta(\hat{h}) = W(\theta)|\hat{h} - h(\theta)|^2$$

Average risk tuning:

$$\arg\min_{\hat{h}} \int L_\theta(\hat{h})\, w(\theta; z)\, d\theta, \quad L_\theta(\hat{h}) = W(\theta)|\hat{h} - h(\theta)|^2$$

Feedforward control:

- $W(\theta) = |G(e^{i\omega}, \theta)|^2 + \lambda, \quad h(\theta) = \frac{\overline{G(e^{i\omega}, \theta)}}{|G(e^{i\omega}, \theta)|^2 + \lambda}$

The Pitman type controller

Average risk tuning:

$$\arg\min_{\hat{h}} \int L_\theta(\hat{h})\, w(\theta; z)\, d\theta, \quad L_\theta(\hat{h}) = W(\theta)|\hat{h} - h(\theta)|^2$$

Feedforward control:

- $W(\theta) = |G(e^{i\omega}, \theta)|^2 + \lambda, \quad h(\theta) = \frac{\overline{G(e^{i\omega}, \theta)}}{|G(e^{i\omega}, \theta)|^2 + \lambda}$
- The Pitman type controller

$$\hat{F}_{\mathrm{Pitman}}(e^{i\omega}) = \frac{\overline{G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}})}}{|G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}})|^2 + \lambda + \mathbb{V}\mathrm{ar}_\theta\left[G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}})\right]}$$

The Pitman type controller

Average risk tuning:

$$\arg\min_{\hat{h}} \int L_\theta(\hat{h})\, w(\theta; z)\, d\theta, \quad L_\theta(\hat{h}) = W(\theta)|\hat{h} - h(\theta)|^2$$

Feedforward control:

- $W(\theta) = |G(e^{i\omega}, \theta)|^2 + \lambda, \quad h(\theta) = \frac{\overline{G(e^{i\omega},\theta)}}{|G(e^{i\omega},\theta)|^2+\lambda}$
- The Pitman type controller

$$\hat{F}_{\mathrm{Pitman}}(e^{i\omega}) = \frac{\overline{G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}})}}{|G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}})|^2 + \lambda + \mathbb{V}\mathrm{ar}_\theta \left[G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}})\right]}$$

- Recall that criterion $Q_\theta(e^{i\omega}, \hat{F}) := |1 - G(e^{i\omega}, \theta)\hat{F}(e^{i\omega})|^2 + \lambda|\hat{F}(e^{i\omega})|^2$ gives optimal controller $F_\theta(e^{i\omega}) := \frac{\overline{G(e^{i\omega},\theta)}}{|G(e^{i\omega},\theta)|^2+\lambda}$

The Pitman type controller

Average risk tuning:

$$\arg\min_{\hat{h}} \int L_\theta(\hat{h})\, w(\theta; z)\, d\theta, \quad L_\theta(\hat{h}) = W(\theta)|\hat{h} - h(\theta)|^2$$

Feedforward control:

- $W(\theta) = |G(e^{i\omega}, \theta)|^2 + \lambda, \quad h(\theta) = \frac{\overline{G(e^{i\omega}, \theta)}}{|G(e^{i\omega}, \theta)|^2 + \lambda}$
- The Pitman type controller

$$\hat{F}_{\mathrm{Pitman}}(e^{i\omega}) = \frac{\overline{G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}})}}{|G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}})|^2 + \lambda + \mathbb{V}\mathrm{ar}_\theta \left[G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}}) \right]}$$

- Recall that criterion $Q_\theta(e^{i\omega}, \hat{F}) := |1 - G(e^{i\omega}, \theta)\hat{F}(e^{i\omega})|^2 + \lambda|\hat{F}(e^{i\omega})|^2$ gives optimal controller $F_\theta(e^{i\omega}) := \frac{\overline{G(e^{i\omega}, \theta)}}{|G(e^{i\omega}, \theta)|^2 + \lambda}$
- Certainty equivalence controller with a modified control criterion

$$|1 - G(e^{i\omega})F(e^{i\omega})|^2 + \left(\lambda + \mathbb{V}\mathrm{ar}_\theta \left[G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}}) \right] \right) |F(e^{i\omega})|^2$$

The Pitman type controller

Average risk tuning:

$$\arg\min_{\hat{h}} \int L_\theta(\hat{h})\, w(\theta; z)\, d\theta, \quad L_\theta(\hat{h}) = W(\theta)|\hat{h} - h(\theta)|^2$$

Feedforward control:

- $W(\theta) = |G(e^{i\omega}, \theta)|^2 + \lambda, \quad h(\theta) = \frac{\overline{G(e^{i\omega},\theta)}}{|G(e^{i\omega},\theta)|^2+\lambda}$
- The Pitman type controller

$$\hat{F}_{\mathrm{Pitman}}(e^{i\omega}) = \frac{\overline{G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}})}}{|G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}})|^2 + \lambda + \mathbb{V}\mathrm{ar}_\theta \left[G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}}) \right]}$$

- Recall that criterion $Q_\theta(e^{i\omega}, \hat{F}) := |1 - G(e^{i\omega}, \theta)\hat{F}(e^{i\omega})|^2 + \lambda|\hat{F}(e^{i\omega})|^2$ gives optimal controller $F_\theta(e^{i\omega}) := \frac{\overline{G(e^{i\omega},\theta)}}{|G(e^{i\omega},\theta)|^2+\lambda}$
- Certainty equivalence controller with a modified control criterion

$$|1 - G(e^{i\omega})F(e^{i\omega})|^2 + \left(\lambda + \mathbb{V}\mathrm{ar}_\theta \left[G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}}) \right]\right) |F(e^{i\omega})|^2$$

- Pauls approach but frequency dependent input penalty!

The Pitman type controller

Average risk tuning:

$$\arg\min_{\hat{h}} \int L_\theta(\hat{h})\, w(\theta; z)\, d\theta, \quad L_\theta(\hat{h}) = W(\theta)|\hat{h} - h(\theta)|^2$$

Feedforward control:

- $W(\theta) = |G(e^{i\omega}, \theta)|^2 + \lambda, \quad h(\theta) = \frac{\overline{G(e^{i\omega},\theta)}}{|G(e^{i\omega},\theta)|^2+\lambda}$
- The Pitman type controller

$$\hat{F}_{\mathrm{Pitman}}(e^{i\omega}) = \frac{\overline{G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}})}}{|G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}})|^2 + \lambda + \mathbb{V}\mathrm{ar}_\theta \left[G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}})\right]}$$

- Recall that criterion $Q_\theta(e^{i\omega}, \hat{F}) := |1 - G(e^{i\omega}, \theta)\hat{F}(e^{i\omega})|^2 + \lambda|\hat{F}(e^{i\omega})|^2$ gives optimal controller $F_\theta(e^{i\omega}) := \frac{\overline{G(e^{i\omega},\theta)}}{|G(e^{i\omega},\theta)|^2+\lambda}$
- Certainty equivalence controller with a modified control criterion

$$|1 - G(e^{i\omega})F(e^{i\omega})|^2 + \left(\lambda + \mathbb{V}\mathrm{ar}_\theta \left[G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}})\right]\right) |F(e^{i\omega})|^2$$

- Pauls approach but frequency dependent input penalty!
- Control criterion detuning instead of modifying the controller (or model)

Hmm, this rings a bell.....

Hmm, this rings a bell.....

A long time ago at the dear green place in the heart of Scotland

The Pitman type controller

Hmm, this rings a bell.....

A long time ago at the dear green place in the heart of Scotland

IEE PROCEEDINGS-D, Vol. 139, No. 1, JANUARY 1992

LQG optimal control design for uncertain systems

M.J. Grimble

Abstract: The modelling and control of systems with uncertain parameters which can be represented by random variables are considered. The model structure is chosen so that high-frequency lag terms may be introduced whose parameters have zero expected value. Similarly uncertain zeros may be considered. An LQG controller is obtained which stabilises the nominal plant model and also provides a degree of robustness to the uncertain poles or zeros. The LQG cost function to be minimised includes an expectation over the parameters, as well as on the stochastic signals. The design procedure introduced enables the cost function weighting terms to be parameterised in terms of the variances of the unknown parameters and a design scalar.

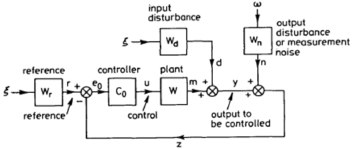


Fig. 1 *Feedback system with input and output disturbances*

$$J_0 = \frac{1}{2\pi j}\oint_D E_p\left\{\frac{1}{\delta W_c^*}\,[\delta W_d^*(q_c\,\bar{S}^*\bar{S} + r_c\,\bar{M}^*\bar{M})\delta W_d]\,\frac{1}{\delta W_c}\,\Phi_{cc}\right\} ds \quad (24)$$

Average risk tuning:

$$\arg\min_{\hat{h}} \int L_\theta(\hat{h})\, w(\theta; z)\, d\theta$$

- The Pitman type controller

$$\hat{F}_{\mathrm{Pitman}}(e^{i\omega}) = \frac{\overline{G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}})}}{|G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}})|^2 + \lambda + \mathbb{V}\mathrm{ar}_\theta \left[G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}}) \right]}$$

The Pitman type controller

Average risk tuning:

$$\arg\min_{\hat{h}} \int L_\theta(\hat{h})\, w(\theta; z)\, d\theta$$

- The Pitman type controller

$$\hat{F}_{\text{Pitman}}(e^{i\omega}) = \frac{\overline{G(e^{i\omega}, \hat{\theta}_{\text{ML}})}}{|G(e^{i\omega}, \hat{\theta}_{\text{ML}})|^2 + \lambda + \mathbb{V}\text{ar}_\theta\left[G(e^{i\omega}, \hat{\theta}_{\text{ML}})\right]}$$

- Gaussian weighting: $\pi(\theta) = \mathcal{N}(\theta; 0, P) \Rightarrow$

The Pitman type controller

Average risk tuning:

$$\arg\min_{\hat{h}} \int L_\theta(\hat{h})\, w(\theta; z)\, d\theta$$

- The Pitman type controller

$$\hat{F}_{\mathrm{Pitman}}(e^{i\omega}) = \frac{\overline{G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}})}}{|G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}})|^2 + \lambda + \mathbb{V}\mathrm{ar}_\theta\left[G(e^{i\omega}, \hat{\theta}_{\mathrm{ML}})\right]}$$

- Gaussian weighting: $\pi(\theta) = \mathcal{N}(\theta; 0, P) \Rightarrow$

$$\hat{F}_P(e^{i\omega}) = \frac{\overline{\hat{G}_P(e^{i\omega})}}{\left|\hat{G}_P(e^{i\omega})\right|^2 + \lambda + \mathbb{V}\mathrm{ar}\left[G(e^{i\omega}, \theta) \mid \hat{\theta}_{\mathrm{ML}}\right]}$$

where

$$\hat{G}_P(e^{i\omega}) = \mathbb{E}\left[G(e^{i\omega}, \theta) \mid \hat{\theta}_{\mathrm{ML}}\right]$$

A palette of feedforward controllers

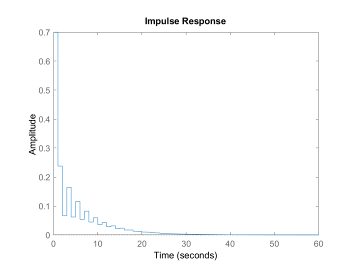
Name	Acronym	Controller
Certainty equivalence	CE	$\dfrac{\overline{G(e^{i\omega},\hat{\theta}_{\mathrm{ML}})}}{\lvert G(e^{i\omega},\hat{\theta}_{\mathrm{ML}})\rvert^2+\lambda}$
Kernel-based CE	KE	$\dfrac{\overline{\hat{G}_{P(\eta)}(e^{j\omega},\hat{\theta}_{\mathrm{ML}})}}{\lvert \hat{G}_{P(\eta)}(e^{j\omega},\hat{\theta}_{\mathrm{ML}})\rvert^2+\lambda}$
Shrinked CE	SH	$\eta\ \hat{F}_{\hat{\theta}_{\mathrm{ML}}}(e^{i\omega})$, $0<\eta<1$
Pitman type	P	$\dfrac{\overline{G(e^{i\omega},\hat{\theta}_{\mathrm{ML}})}}{\lvert G(e^{i\omega},\hat{\theta}_{\mathrm{ML}})\rvert^2+\lambda+\mathbb{V}\mathrm{ar}_{\theta}\left[G(e^{i\omega},\hat{\theta}_{\mathrm{ML}})\right]}$
Kernel based average risk	KE A	$\dfrac{\overline{\hat{G}_{P(\eta)}(e^{i\omega})}}{\left\lvert \hat{G}_{P(\eta)}(e^{i\omega})\right\rvert^2+\lambda+\mathbb{V}\mathrm{ar}\left[G(e^{i\omega},\theta)\mid\hat{\theta}_{\mathrm{ML}}\right]}$
Input penalized CE	IP	$\dfrac{\overline{G(e^{i\omega},\hat{\theta}_{\mathrm{ML}})}}{\lvert G(e^{i\omega},\hat{\theta}_{\mathrm{ML}})\rvert^2+\eta(\omega)}$

A palette of feedforward controllers

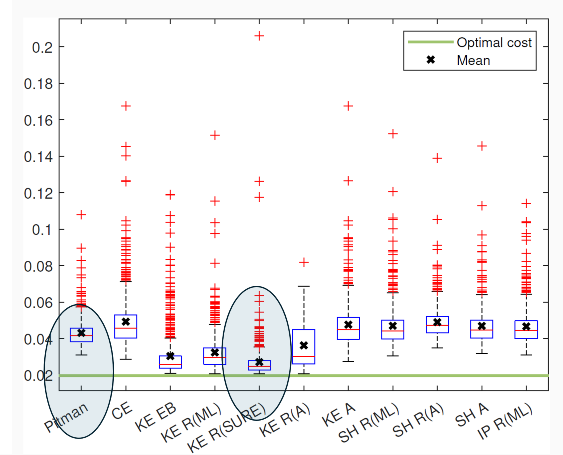
Name	Acronym	Controller
Certainty equivalence	CE	$\dfrac{\overline{G(e^{i\omega},\hat{\theta}_{\mathrm{ML}})}}{\lvert G(e^{i\omega},\hat{\theta}_{\mathrm{ML}})\rvert^2+\lambda}$
Kernel-based CE	KE	$\dfrac{\overline{\hat{G}_{P(\eta)}(e^{j\omega},\hat{\theta}_{\mathrm{ML}})}}{\lvert \hat{G}_{P(\eta)}(e^{j\omega},\hat{\theta}_{\mathrm{ML}})\rvert^2+\lambda}$
Shrinked CE	SH	$\eta\, \hat{F}_{\hat{\theta}_{\mathrm{ML}}}(e^{i\omega})$, $0<\eta<1$
Pitman type	P	$\dfrac{\overline{G(e^{i\omega},\hat{\theta}_{\mathrm{ML}})}}{\lvert G(e^{i\omega},\hat{\theta}_{\mathrm{ML}})\rvert^2+\lambda+\mathbb{V}\mathrm{ar}_\theta\left[G(e^{i\omega},\hat{\theta}_{\mathrm{ML}})\right]}$
Kernel based average risk	KE A	$\dfrac{\overline{\hat{G}_{P(\eta)}(e^{i\omega})}}{\left\lvert \hat{G}_{P(\eta)}(e^{i\omega})\right\rvert^2+\lambda+\mathbb{V}\mathrm{ar}\left[G(e^{i\omega},\theta)\mid \hat{\theta}_{\mathrm{ML}}\right]}$
Input penalized CE	IP	$\dfrac{\overline{G(e^{i\omega},\hat{\theta}_{\mathrm{ML}})}}{\lvert G(e^{i\omega},\hat{\theta}_{\mathrm{ML}})\rvert^2+\eta(\omega)}$

Remember that for any of these controllers and any tuning method for η, there is an average risk controller that improves on it.

Numerical illustrations



Var1	Var2
"Order FIR"	57
"# Monte Carlo"	500
"# samples for risk approx"	200
"# data for identification"	400
"SNR"	-4.1759
"lambda"	0.1



Kernel based method with unbiased risk estimation outperforms Pitman

Conclusions

- Direct data driven or indirect model based not a concern from a statistical perspective
- Instead minimal sufficient statistics central
- Variance due to the complexity of this statistic must be handled in the usual way, i.e. by tailoring the controller structure to the class of systems in question
- Average risk controllers possess several attractive properties
- Despite this, a rich palette of ways to construct data-driven controllers
- But remember that for any controller you can come up with, there is a better average risk controller (for exponential families at least)
- The importance of using the degree of freedom offered by the control criterion has been highlighted highlighted (well known by practitioners)

Related work (1/2)

- Main inspiration for this talk comes from

A. Chiuso (2023). ***ERNSI Workshop 2023: Optimal Data Driven Predictive Control for linear stochastic systems.*** URL: *https://www.kth.se/polopoly_fs/1.1284895.1696833266!/ERNSI_2023_Alessandro_Chiuso.pdf*

Alessandro Chiuso et al. (2023). ***Harnessing the Final Control Error for Optimal Data-Driven Predictive Control.*** arXiv: *2312.14788 [eess.SY]*

- Statistical decision theory

E. L. Lehmann and G. Casella (1998). ***Theory of Point Estimation.*** Second. New York: John Wiley & Sons

Related work (2/2)

- Average risk control (a.k.a. Bayes control, Bayesian control)

A. Scampicchio et al. (2019). **"Bayesian Kernel-Based Linear Control Design".** In: *2019 IEEE 58th Conference on Decision and Control (CDC)*, pp. 822–827

S. Formentin and A. Chiuso (2021). **"Control-oriented regularization for linear system identification".** In: *Automatica* 127, p. 109539

- Recent literature

F. Dörfler (2023a). **"Data-Driven Control: Part One of Two: A Special Issue Sampling from a Vast and Dynamic Landscape".** In: *IEEE Control Systems Magazine* 43.5, pp. 24–27. DOI: *10.1109/MCS.2023.3291624*

F. Dörfler (2023b). **"Data-Driven Control: Part Two of Two: Hot Take: Why not go with Models?"** In: *IEEE Control Systems Magazine* 43.6, pp. 27–31. DOI: *10.1109/MCS.2023.3310302*

A typical day in the life of a control engineer

Good luck with your future projects Paul!

PS: I hear there's an opening in Secunda

References

Åström, K.J. and T. Bohlin (1965). **"Numerical identification of linear dynamic systems from normal operating records".** In: *Proc. IFAC Symp. Self-Adaptive Systems*. Teddington, U.K., pp. 96–111.

Åström, K.J. and B. Wittenmark (1971). **"Problems of identification and control".** In: *J. Math. Analalysis and Applications* 34, pp. 90–113.

Åström, K.J. and B. Wittenmark (1989). ***Adaptive Control.*** Reading, Massachusetts: Addison-Wesley.

Åström, K.J. et al. (1977). **"Theory and Applications of Self-Tuning Regulators".** In: *Automatica* 13, pp. 457–476.

Chiuso, A. (2023). ***ERNSI Workshop 2023: Optimal Data Driven Predictive Control for linear stochastic systems.*** URL: *https://www.kth.se/polopoly_fs/1.1284895.1696833266!/ERNSI_2023_Alessandro_Chiuso.pdf*.

Chiuso, Alessandro et al. (2023). ***Harnessing the Final Control Error for Optimal Data-Driven Predictive Control.*** arXiv: *2312.14788 [eess.SY]*.

Colin, K. et al. (2024). **"A bias-variance perspective of data-driven control".** In: *20th IFAC Symposium on System Identification*. To appear. Boston, USA.

Coulson, J., J. Lygeros, and F. Dörfler (2019). **"Data-enabled predictive control: In the shallows of the DeePC".** In: *18th European Control Conference (ECC)*, pp. 307–312.

Dörfler, F. (2023a). **"Data-Driven Control: Part One of Two: A Special Issue Sampling from a Vast and Dynamic Landscape".** In: *IEEE Control Systems Magazine* 43.5, pp. 24–27. DOI: *10.1109/MCS.2023.3291624*.

— (2023b). **"Data-Driven Control: Part Two of Two: Hot Take: Why not go with Models?"** In: *IEEE Control Systems Magazine* 43.6, pp. 27–31. DOI: *10.1109/MCS.2023.3310302*.

Formentin, S. and A. Chiuso (2021). **"Control-oriented regularization for linear system identification".** In: *Automatica* 127, p. 109539.

Gevers, M. and L. Ljung (1986). **"Optimal experiment designs with respect to the intended model application".** In: *Automatica* 22.5, pp. 543–554.

Grimble, M.J. (1992). **"LOG optimal control design for uncertain systems".** In: *IEE Proceedings D (Control Theory and Applications)* 139 (1), pp. 21–30.

Hakvoort, R.G., R.J.P. Schrama, and P.M.J. Van den Hof (1992). **"Approximate Identification in view of LQG Feedback Design".** In: *Proc. ACC*. Chicago.

Lehmann, E. L. and G. Casella (1998). ***Theory of Point Estimation.*** Second. New York: John Wiley & Sons.

Pitman, E. J. G. (1939). **"The Estimation of the Location and Scale Parameters of a Continuous Population of any Given Form".** In: *Biometrika* 30.3/4, pp. 391–421.

Scampicchio, A. et al. (2019). **"Bayesian Kernel-Based Linear Control Design".** In: *2019 IEEE 58th Conference on Decision and Control (CDC)*, pp. 822–827.

Schrama, R. J. P. (1992). **"Accurate models for control design: the necessity of an iterative scheme".** In: *IEEE Transactions on Automatic Control* 37.7, pp. 991–994.

Schrama, R.J.P. and P.M.J. Van den Hof (1992). **"An Iterative Scheme for Identification and Control Design Based on Coprime Factorizations".** In: *Proc. ACC*. Chicago, pp. 2842–2846.

— (1993). **"Iterative Identification and Control Design: A Three Step Procedure with Robustness Analysis".** In: *2nd European Control Conference*. Groningen, pp. 237–241.

Wahlberg, B. and L. Ljung (1986). **"Design variables for bias distribution in transfer function estimation".** In: *IEEE Trans. Automatic Control* 31.2, pp. 134–144.