Možnosti využití dynamického programování v bezsenzorovém řízení

Václav Šmídl

21. února 2011

Obsah

Motivace

- GACR projekt "Regulace a identifikace parametrů střídavých elektrických pohonů v kritických provozních stavech"

  FEL ZČU nejlepší existující používané metody pro řízení pohonů (PID+expertní znalost)

  ÚTIA pokročilé techniky řízení (Bayesovské odhadování, dynamické programování)

- Cílem je zlepšit stávající techniky ve dvou směrech:
  1. shora dolů - aplikovat pokročilé věci na pohony a získat zpětnou vazbu pro vylepšování,
  2. zdola nahoru - navrhnout pomocí pokročilých technik jednoduché vylepšení stávajících.
- Cíl přednášky je
  1. představit současný stav,
  2. vytipovat směr vývoje.

Dynamické programování

Máme systém popsaný diferenciální (diferenční) rovnicí

$$x_{t+1} = f(x_t, u_t),$$

a chceme dosáhnout požadovaného chování.

Řízení je optimální pokud minimalizuje ztrátu:

$$L = \sum_{t=1}^{T} l_t(x_t, u_t).$$

[1] Dynamic Programming

Dynamické programování

Máme systém popsaný diferenciální (diferenční) rovnicí

$$x_{t+1} = f(x_t, u_t),$$

a chceme dosáhnout požadovaného chování.

Řízení je optimální pokud minimalizuje ztrátu:

$$L = \sum_{t=1}^{T} l_t(x_t, u_t).$$

Nutná a postačující podmínka optima, (Bellman, 1957)[1]:

$$\begin{aligned} V(x_\tau) &= \min_{u(x_t)} \left\{ l_t\left(x_\tau, u_\tau\right) + V\left(f(x_\tau, u_\tau)\right)\right\}, \\ u(x_\tau) &= \arg\min_{u(x_\tau)} \left\{ l\left(x_\tau, u_\tau\right) + V\left(f(x_\tau, u_\tau)\right)\right\}. \end{aligned}$$

Rekurze $\tau = t + h, \ldots, t,\ V(x_{t+h}) = 0$.

[1] Dynamic Programming

Vlastnosti dynamického programování

- Analytické řešení pro
  - **LQ** lineární systém s kvadratickou ztrátou, (Peterka and Bláha, 1966)[2]
  - konečný počet stavů systému a řídících zásahů,
- Četná numerická řešení pro nízké dimenze (curse of dimensionality)
- Snadné rozšíření na **stochastické** problémy
  - **LQG** - LQ s Gausovskou distribucí neurčitosti,
  - **MDP** - Markov Decision Process
  - numerická řešení jsou **snáze** řešitelná (Belbas, personal communication).
- Alternativní formulace,
  - neurodynamic control (Bertsekas, 2001)[3],
  - reinforcement learning (Sutton and Barto, 1998)[4]
  - fully probabilistic approach (Kárný, 1995)[5]

---

[2] Synthesis of regulation loops according to quadratic criterion
[3] Dynamic Programming and Optimal Control
[4] Reinforcement Learning: An Introduction
[5] Towards fully probabilistic control design

Dynamické programování v řízení pohonů

- DP se prakticky nepoužívá (začíná se s LQ řízením)
- Hlavní směry
  1. **vektorové řízení**: řízení PID regulátory v d-q souřednicích.
  2. **přímé řízení momentu:** výběr ze 4 možných akcí.
- Souvislosti:
  - vektorové řízení a PID $\Longleftrightarrow$LQG a varianty
  - přímé řízení momentu $\Longleftrightarrow$ diskrétní (Markovské) modely

Vektorové řízení pohonu

Model

$$\begin{aligned}
\frac{di_{s\alpha}}{dt} &= -\frac{R_s}{L_s} i_{s\alpha} + \frac{\Psi_{PM}}{L_s}\omega_{me}\sin\vartheta + \frac{u_{s\alpha}}{L_s}, \\
\frac{di_{s\beta}}{dt} &= -\frac{R_s}{L_s} i_{s\beta} - \frac{\Psi_{PM}}{L_s}\omega_{me}\cos\vartheta + \frac{u_{s\beta}}{L_s}, \\
\frac{d\omega}{dt} &= \frac{k_p p_p^2 \Psi_{pm}}{J}\left(i_{s\beta}\cos\vartheta - i_{s\alpha}\sin\vartheta\right) - \frac{B}{J}\omega - \frac{p_p}{J} T_L, \\
\frac{d\vartheta}{dt} &= \omega.
\end{aligned} \tag{1}$$

Ztráta

$$L = \sum_{t=1}^{T} (\omega - \overline{\omega})^2$$

Složité řízení, s rostoucím ω hrají sin() a cos() významnější úlohu.

d-q souřadnice

Po transformaci

$$\begin{bmatrix} i_{sd} \\ i_{sq} \end{bmatrix} = \begin{bmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{bmatrix} \begin{bmatrix} i_{s\alpha} \\ i_{s\beta} \end{bmatrix}$$

získáme stavový prostor,

$$x_t = [i_{sd}, i_{sq}, \omega, \vartheta]$$

ve kterém lze konstantní žádanou hodnotu ω převést na **konstantní** žádané hodnoty$\bar{i}_{sd}, \bar{i}_{sq}$.

Problém se redukuje na

$$x_t = A(\omega)x_{t-1} + Bu_t.$$

se zanedbáním časového vývoje ω dostáváme lineární systém (LQG).

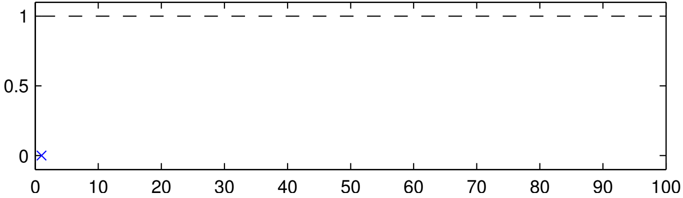
LQ řízení: příklad

Jednorozměrný systém

$$x_{t+1} = bx_t + cu_t$$

stabilní ($b < 1$).

- počáteční podmínka: $x_t = 0$
- požadovaná hodnota: $\overline{x}_t = 1$ dosažená vstupem s malou energií,



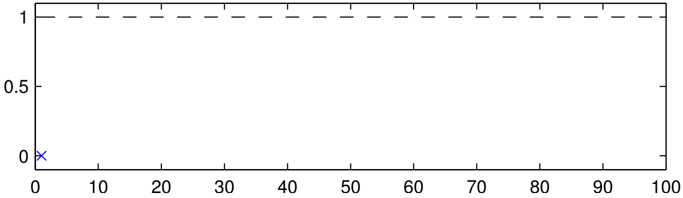
LQ řízení: příklad

Jednorozměrný systém

$$x_{t+1} = bx_t + cu_t$$

stabilní ($b < 1$).

- počáteční podmínka: $x_t = 0$
- požadovaná hodnota: $\overline{x}_t = 1$ dosažená vstupem s malou energií,



Kritérium:

$$L = \sum_{t=1}^{T} (x_t - \overline{x}_t)^2 + \beta u_t^2$$

kde β je zvolená konstanta.

LQ řízení: odvození

Čas $t = t + h$

$$u_{t+h} = \arg\min_{u(x_{t+h})} \left((x_{t+h} - \overline{x}_{t+h})^2 + \beta u_{t+h}^2 \right)$$

LQ řízení: odvození

Čas $t = t + h$

$$u_{t+h} = \arg\min_{u(x_{t+h})} \left((x_{t+h} - \overline{x}_{t+h})^2 + \beta u_{t+h}^2\right) = 0$$

LQ řízení: odvození

Čas $t = t + h$

$$
\begin{aligned}
u_{t+h} &= \arg\min_{u(x_{t+h})} \left((x_{t+h} - \overline{x}_{t+h})^2 + \beta u_{t+h}^2 \right) = 0 \\
V(x_{t+h}) &= (x_{t+h} - \overline{x}_{t+h})^2
\end{aligned}
$$

LQ řízení: odvození

Čas $t = t + h$

$$
\begin{aligned}
u_{t+h} &= \arg\min_{u(x_{t+h})} \left((x_{t+h} - \overline{x}_{t+h})^2 + \beta u_{t+h}^2\right) = 0 \\
V(x_{t+h}) &= (x_{t+h} - \overline{x}_{t+h})^2 = [x_{t+h}, 1] \begin{bmatrix} 1 & -\overline{x}_{t+h} \\ -\overline{x}_{t+h} & \overline{x}_{t+h}^2 \end{bmatrix} \begin{bmatrix} x_{t+h} \\ 1 \end{bmatrix}
\end{aligned}
$$

LQ řízení: odvození

Čas $t = t + h$

$$
\begin{aligned}
u_{t+h} &= \arg\min_{u(x_{t+h})} \left((x_{t+h} - \overline{x}_{t+h})^2 + \beta u_{t+h}^2\right) = 0 \\
V(x_{t+h}) &= (x_{t+h} - \overline{x}_{t+h})^2 = [x_{t+h}, 1] \begin{bmatrix} 1 & -\overline{x}_{t+h} \\ -\overline{x}_{t+h} & \overline{x}_{t+h}^2 \end{bmatrix} \begin{bmatrix} x_{t+h} \\ 1 \end{bmatrix}
\end{aligned}
$$

Čas $t = t + h - 1$

$$
u_{t+h-1} = \arg\min_{u(x_{t+h-1})} \left((x_{t+h-1} - \overline{x}_{t+h-1})^2 + \beta u_{t+h-1}^2 + V(bx_{t+h-1} + cu_{t\text{-}}\right.
$$

LQ řízení: odvození

Čas $t = t + h$

$$
\begin{aligned}
u_{t+h} &= \arg\min_{u(x_{t+h})} \left((x_{t+h} - \overline{x}_{t+h})^2 + \beta u_{t+h}^2\right) = 0 \\
V(x_{t+h}) &= (x_{t+h} - \overline{x}_{t+h})^2 = [x_{t+h}, 1] \begin{bmatrix} 1 & -\overline{x}_{t+h} \\ -\overline{x}_{t+h} & \overline{x}_{t+h}^2 \end{bmatrix} \begin{bmatrix} x_{t+h} \\ 1 \end{bmatrix}
\end{aligned}
$$

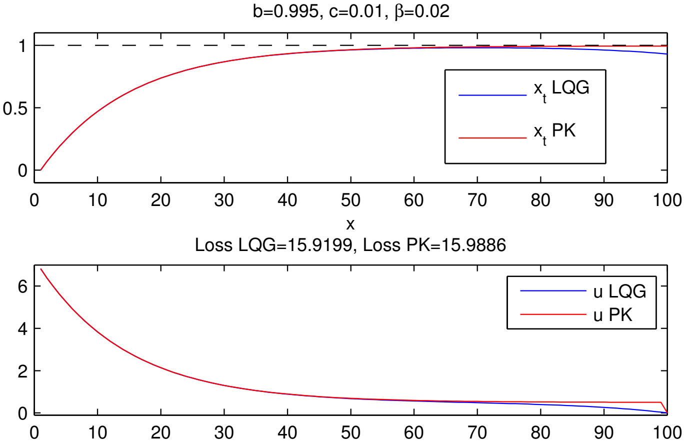
Čas $t = t + h - 1$

$$
\begin{aligned}
u_{t+h-1} &= \arg\min_{u(x_{t+h-1})} \left((x_{t+h-1} - \overline{x}_{t+h-1})^2 + \beta u_{t+h-1}^2 + V(bx_{t+h-1} + cu_{t-}\right. \\
&\quad \ldots = -l_1 x_{t+h-1} - l_2, \\
V(x_{t+h-1}) &= [x_{t+h-1}, 1] \; \Psi \begin{bmatrix} x_{t+h-1} \\ 1 \end{bmatrix}.
\end{aligned}
$$

Výsledkem je PK regulátor s proměnnými koeficienty.

LQ řízení: simulace

PK regulace jsou ustálené hodnoty LQ.



LQ s neznámými parametry

Jednorozměrný systém

$$x_{t+1} = x_t + c u_t + \sigma_e e_t.$$

Neznámé c**.** Kritérium:

$$L = \sum_{t=1}^{T} (x_t - \overline{x}_t)^2 + \beta u_t^2$$

Bellmanova rovnice je nyní:

$$\begin{aligned} V(x_\tau) &= \min_{u(x_\tau)} \mathsf{E}_c \left\{ l\left(x_\tau, u_\tau\right) + V\left(f(x_\tau, u_\tau)\right) \right\}, \\ u(x_\tau) &= \arg\min_{u(x_\tau)} \mathsf{E}_c \left\{ l\left(x_\tau, u_\tau\right) + V\left(f(x_\tau, u_\tau)\right) \right\}. \end{aligned}$$

kde $\mathsf{E}_c[]$ je očekávaná hodnota pro neurčitou hodnotu c daná $p(c|x_t, u(x_t))$.

Odhad parametru c

Předpokládejme, že počáteční odhad c je $p(c) = \mathcal{N}(c_0, \sigma_0)$. Pak vývoj odhadu parametru je:

$$\begin{aligned} p(c|x_{1:t+1}) &= \mathcal{N}(\hat{c}_{t+1}, \sigma_{t+1}), \\ \hat{c}_{t+1} &= \hat{c}_t + \frac{u_t \sigma_t}{u_t^2 \sigma_t + \sigma_e}(x_{t+1} - x_t - \hat{c}_t u_t), \\ \sigma_{t+1} &= \left(1 - \frac{u_t^2 \sigma_t}{u_t^2 \sigma_t + \sigma_e}\right) \sigma_t. \end{aligned}$$

Je možné interpretovat jako 3rozměrný systém se stavem $H_t = [x_t, \hat{c}_t, \sigma_t]$ a očekávanou ztrátou

$$\mathrm{E}((x_t - \overline{x}_t)^2 + \beta u_t^2) = (x_{t-1} + \hat{c}_t u_t - \overline{x}_t)^2 + \left(\beta + \sigma_t^2\right) u_t^2.$$

Analytické řešení neexistuje. Dobrý cvičný příklad, integrátor s neznámým ziskem (Åström and Helmersson, 1986)[6]

[6] Dual control of an integrator with unknown gain

Integrátor s neznámým ziskem

Simuluje bezsensorové řízení v nulových otáčkách:

- c_t může být jak kladné tak záporné se stejnou pravděpodobností,
- c_t může být 0 - brzda,
- počáteční neurčitost je veliká (v ořezaných gausovkách dokonce roste),
- lze rozšířit na časově proměnné c_t - simuluje průchody nulou.

[7] Dual control of an integrator with unknown gain
[8] Stochastic Iterative dynamic programming: a Monte Carlo approach to dual control
[9] Iterative local dynamic programming

Integrátor s neznámým ziskem

Simuluje bezsensorové řízení v nulových otáčkách:

- c_t může být jak kladné tak záporné se stejnou pravděpodobností,
- c_t může být 0 - brzda,
- počáteční neurčitost je veliká (v ořezaných gausovkách dokonce roste),
- lze rozšířit na časově proměnné c_t - simuluje průchody nulou.

Existující řešení:

- Numerické na gridu (Åström and Helmersson, 1986)[7],
- Monte Carlo (Thompson and Cluet, 2005)[8], ověřoval M. Zima,
- Pomocí lokální LQ aproximace (Todorov and Tassa, 2009)[9], ověřoval M. Vahala,

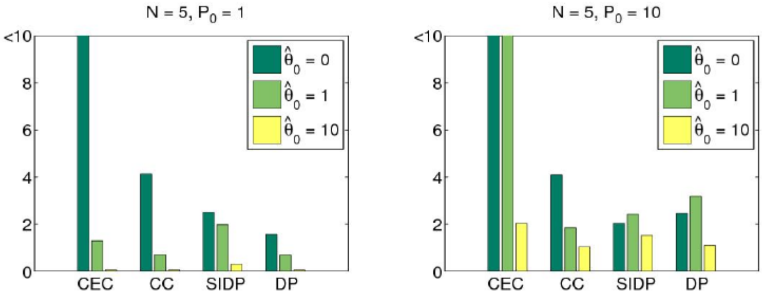
[7] Dual control of an integrator with unknown gain

[8] Stochastic Iterative dynamic programming: a Monte Carlo approach to dual control

[9] Iterative local dynamic programming

Opatrné LQ řízení

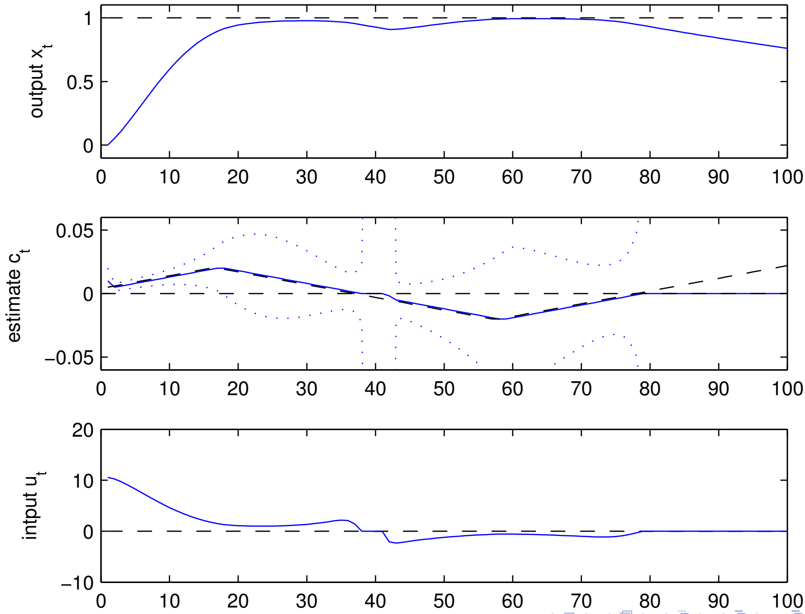
Výsledky pro Monte Carlo řešení metodou SIDP, (M. Zima, 2010)



Opatrné řízení:

- stejné jako standardní LQ pro penalizaci energie $(\beta + \sigma_t)u_t^2$ s konstatním σ_t
- není duální. Dualita je významná pouze pro $\hat{c}_0 = 0$.

Certainty equivalence: proměnné c



Dual control

Optimální řízení může být aproximativně rozděleno na dvě složky:

$$u_t = u_{\text{opatrné}}(x_t) + u_{\text{budící}}(x_t).$$

kde:

- opatrné řízení je optimální 1-krokové řízení.
- budící řízení udržuje systém (pokud možno) pozorovatelný.

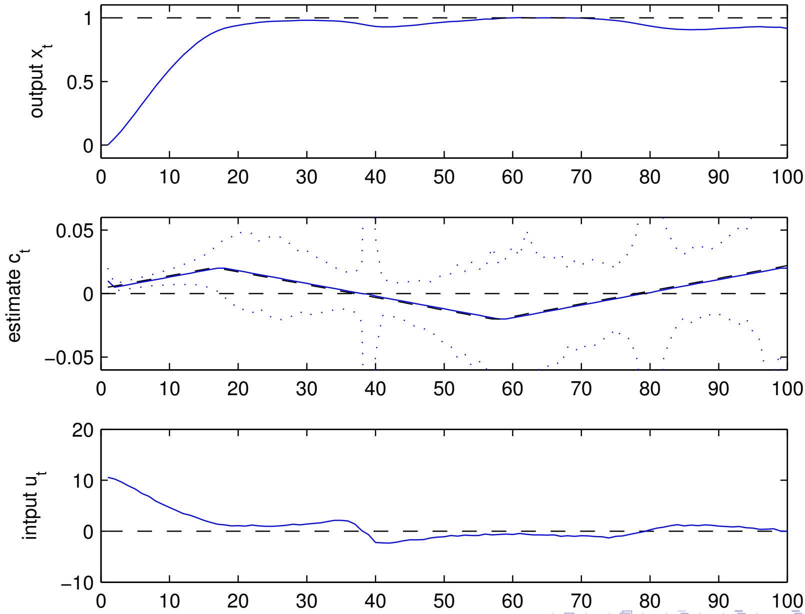
Obecně je těžké najít kompromis (Filatov and Unbehauen, 2000)[10], (Simpkins, Callafon, and Todorov, 2008)[11]. Známé "triky" (Peterka, Böhm, Hallousková, Kárný, and Maršík, 1982)[12].

Můžeme zkusit náhodný šum:

$$u_{\text{budící}}(x_t) = \mathcal{N}(0, \sigma_{\text{budící}}).$$

[10] Survey of adaptive dual control methods
[11] Optimal trade-off between exploration and exploitation
[12] Algorithms for microprocessor-based control of technologic processes

Duální řízení s konstantním buzením



Plně pravděpodonostní varianta LQ

LQ je speciálním případem FPD

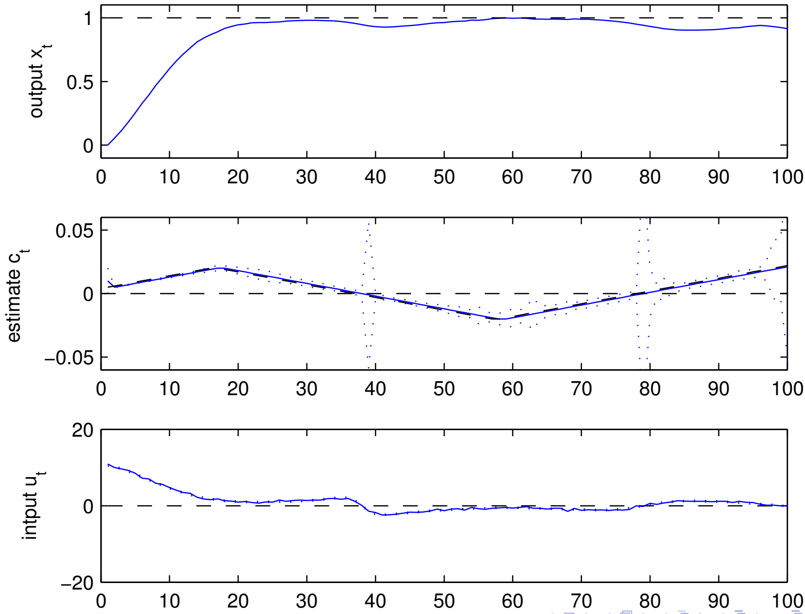
$$\begin{aligned} {}^I f(u_t) &= \mathcal{N}(0, \beta^{-1}) \\ {}^I f(x_t) &= \mathcal{N}(0, \sigma_e) \end{aligned}$$

$${}^o f(u_t|x_t) = \mathcal{N}(u_{LQ}(x_t), \sigma_{FPD,t}).$$

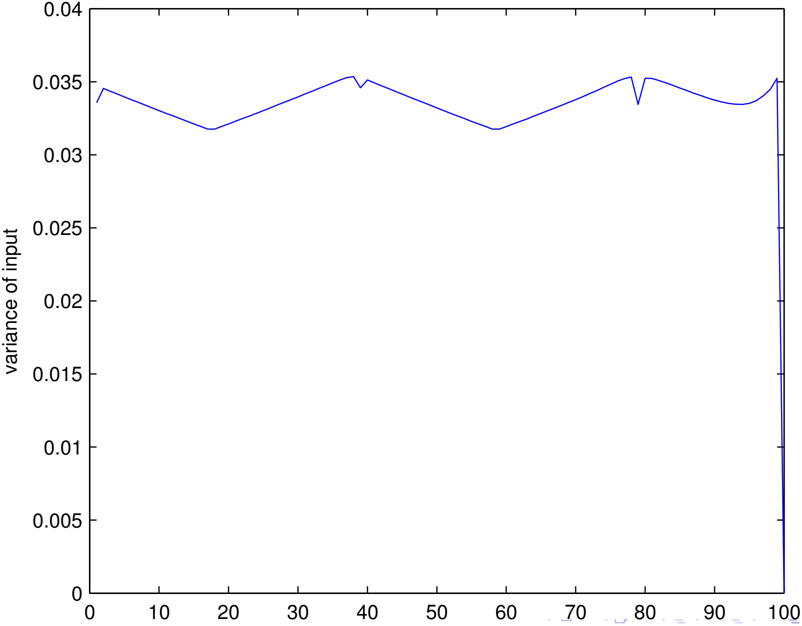
kde LQ vznikne zanedbáním $\sigma_{FPD,t}$.

Co když $\sigma_{FPD,t}$ použijeme?

Plně pravděpodonostní varianta LQ

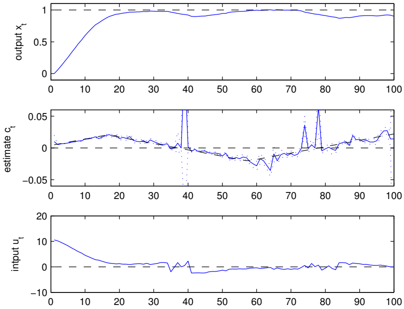


Plně pravděpodonostní varianta LQ II



Druhý pohled

Test s šumem v pozorování, σ_e:



Ztráty:

CE	CC	FPD
12.64	12.61	12.67

Limita pro neproměnné $c_t = c$, σ_{FPD} nekonverguje k 0.

Závěr

Současný stav:

- PI s naladěným konstantním ziskem + rozladěný Kalman filtr
- LQ varianty se zkoušejí jako nová (výpočetně náročná) technika

Vybrané směry:

- Opatrné řízení + správně naladěný Klaman filtr
- prozkoumání variant FPD (λ, jiná forma entropie?)
- diskrétní modelování

Literatura I

Åström, KJ and A. Helmersson (1986). "Dual control of an integrator with unknown gain". In: *Computers & Mathematics with Applications* 12.6, pp. 653–662.

Bellman, R. (1957). *Dynamic Programming*. New York: Princeton University Press.

Bertsekas, D.P. (2001). *Dynamic Programming and Optimal Control*. 2nd edition. Nashua, US: Athena Scientific.

Filatov, NM and H. Unbehauen (2000). "Survey of adaptive dual control methods". In: *IEE Proceedings-Control Theory and Applications* 147.1, pp. 118–128.

Kárný, M. (1995). "Towards fully probabilistic control design". In: *IFAC Symposium ACASP'95*. Ed. by Cs. Bányász. Budapest, pp. 261–264.

Peterka, V. and S. Bláha (1966). "Synthesis of regulation loops according to quadratic criterion". In: *Kybernetika* 1. In Czech, pp. 127–143.

Peterka, V., J. Böhm, A. Halousková, M. Kárný, and J. Maršík (1982). *Algorithms for microprocessor-based control of technologic processes*. Tech. rep. in Czech. Prague, Plzen: UTIA CSAV, Skoda.

Literatura II

Simpkins, A., R. de Callafon, and E. Todorov (2008). "Optimal trade-off between exploration and exploitation". In: *American Control Conference, 2008*, pp. 33–38.

Sutton, R.S. and A.G. Barto (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

Thompson, A.M. and W.R. Cluet (2005). "Stochastic Iterative dynamic programming: a Monte Carlo approach to dual control". In: *Automatica* 41, pp. 767–778.

Todorov, E. and Y. Tassa (2009). "Iterative local dynamic programming". In: *Proceedings of the 2nd IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning*, pp. 90 –95.