System pro simulaci

December 10, 2009

Puvodni zadani:

Vychazime ze zadani [1]:

$$\begin{aligned} SYSTEM: \; y_{t+1} &= y_t + bu_t + \sigma e_t, \;\; e_t \sim \mathcal{N}(0,1), && (1) \\ ZTRATA: \; L_t &= (y_{t+1} - r_{t+1})^2 && (2) \end{aligned}$$

Reseni schematicky:

$$V_t = \min_{u_t} \mathsf{E}_{e_t,b} \left\{ L_t + V_{t+1} | y_t, u_{t-1}, y_{t-1}, u_{t-2}, \ldots \right\} \tag{3}$$

kde stredni hodnota se pocita pres neurcitost v e_t a pres neurcitost v b.

Pro linearni a Gausovsky system (1) je k dispozici konjugovana hustota ve forme Normalniho rozlozeni pravdepodobnosti $f(b_t) = \mathcal{N}(\hat{b}_t, P_t)$, jejiz parametry se vyvijeji rekurzivne, rovnice (28) v [1]. Tim padem je mozne vycislit ocekavanou hodnotu pres b v (3) analyticky:

$$\begin{aligned} V_t &= \min_{u_t} \mathsf{E}_{e_t,b} \left\{ (y_t + bu_t + \sigma e_t - r_{t+1})^2 + V_{t+1} | y_t, u_{t-1}, y_{t-1}, u_{t-2}, \ldots \right\} \\ &= \min_{u_t} \mathsf{E}_{e_t} \left\{ (y_t + \hat{b} u_t + \sigma e_t - r_{t+1})^2 + P_t u_t^2 | y_t, u_{t-1}, y_{t-1}, u_{t-2}, \ldots \right\} + \\ &\quad + \mathsf{E}_{e_t,b} \left\{ V_{t+1} | y_t, u_{t-1}, y_{t-1}, u_{t-2}, \ldots \right\} \end{aligned}$$

Muzeme provest preznaceni

$$V_{t+1}(H_t) = \mathsf{E}_{e_t,b} \left\{ V_{t+1} | y_t, u_{t-1}, y_{t-1}, u_{t-2}, \ldots \right\}$$

kde $H_t = [y_t, \hat{b}_t, P_t]$. Vysledna uloha je ekvivalentni tomu, kdyby zadani bylo:

$$\begin{aligned} SYSTEM: \; H_{t+1} &= \begin{bmatrix} y_{t+1} \\ \hat{b}_{t+1} \\ P_{t+1} \end{bmatrix} = \begin{bmatrix} y_t + \hat{b}_t u_t \\ (28) \\ (28) \end{bmatrix} + \begin{bmatrix} \sigma e_t \\ 0 \\ 0 \end{bmatrix} && (4) \\ ZTRATA: \; L_t &= (y_{t+1} - r_{t+1})^2 + P_t u_t^2. && (5) \end{aligned}$$

(28) je opet rovnice (28) z [1]. Takto upravenou ulohu muzeme resit pomoci algoritmu [2].

LQ rizeni

Algoritmus LQ rizeni je aplikovatelny v pripade, ze b v (1) je zname. V pripade, ze b nezmame je mozne optimalni rizeni aproximovat tzv. receding horizon strategii. Tato strategie spociva v nahrazeni $b \equiv \hat{b}_t$, spocteni optimalniho zasahu, provedeni u_t , oprava b_t a opetovne prepocteni strategie.

Tomuto postupu se rika certainty equivalence. Nevyhodou tohoto pristupu je, ze chyba rizeni pro chybny odhad $\hat{b}$ je znacna.

Druhou moznosti aproximace je pouziti systemu (4) s nahradou $\hat{b}_{t+1} = \hat{b}_t$, $P_{t+1} = P_t$. Vysledek je velmi podobny jako u CE strategie, avsak do ztratove funkce pribyl penalizacni clen $P_t u_t^2$, ktery penalizuje velke hodnoty u_t. Pro velke hodnoty P_t tak vznika preference pro male hodnoty u_t. Vysledne strategii rizeni se proto rika cautious, tedy opatrna. Nevyhodou teto strategie je prilisna "opatrnost", ktera vychazi z predpokladu konstantnosti P_t, tedy velke penalizace u_t na celem horizontu. Kvuli aproximaci neni ve strategii zohlednen vliv u_t na P_t, a tim i fakt, ze vhodne zvolene u_t muze hodnoty P_t snizit.

Tento efekt se da kompenzovat tim, ze predpokladame, ze P_t bude s casem klesat, napr:

$$P_{t+1} = \frac{1}{2} P_t.$$

pripadne az do krajnosti:

$$P_{t+1} = 0.$$

References

[1] A.M. Thompson and W.R. Cluet. Stochastic iterative dynamic programming: a monte carlo approach to dual control. *Automatica*, 41:767–778, 2005.

[2] E. Todorov and Tassa. Y. Iterative local dynamic programming. In *Proceedings of the 2nd IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning*, pages 90 – 95, 2009.