Zasada optymalności Bellmana.
Uogólniony optymalny regulator stanu.

Podstawowy problem sterowania optymalnego dla układów nieliniowych z czasem ciągłym

W podstawowym problemie sterowania optymalnego z czasem ciągłym minimalizacji podlega całkowy wskaźnik jakości

$$G(x,u) \doteq \int_{t_0}^{t_1} g(x(t),u(t),t)dt$$

z uwzględnieniem ograniczeń w postaci równania stanu procesu

$$\dot{x}(t) = f(x(t),u(t),t), \quad t \in [t_0,t_1]$$

z zadanym warunkiem początkowym

$$x(t_0) = x_0$$

oraz ograniczeń w postaci dopuszczalnego chwilowego zakresu wartości sterowania

$$u(t) \in U, \quad t \in [t_0,t_1],$$

gdzie funkcje

$$g : R^n \times R^m \times R \to R, \quad f : R^n \times R^m \times R \to R^n$$

są, ogólnie biorąc, nieliniowe.

Klasa sterowań dopuszczalnych w przedziale $[t_0,t_1]$ oznaczana jako $\mathcal{U}[t_0,t_1]$ jest zbiorem funkcji przedziałami ciągłych o skończonej liczbie nieciągłości pierwszego rodzaju, tj. $\mathcal{U}[t_0,t_1] \doteq PC([t_0,t_1],R^m)$.

Niech trajektoria stanu $ABCD$ będzie trajektorią optymalną związaną ze stanem początkowym $x(0) = x_0$ i niech $x(\tilde{t}) = \tilde{x}$ będzie stanem pośrednim związanym z punktem B trajektorii optymalnej.

Zasada optymalności Bellmana: Każdy końcowy odcinek BCD optymalnej trajektorii stanu $ABCD$ związanej ze stanem początkowym $x(0)$ jest optymalną trajektorią stanu związaną ze stanem pośrednim $x(\tilde{t})$.

Dowód: Przypuśćmy, że końcowy odcinek trajektorii BCD nie jest trajektorią optymalną dla pośredniego stanu początkowego $x(\tilde{t})$. Wobec tego między punktami B i D istnieje inna trajektoria $BC'D$, która zapewnia mniejszą wartość wskaźnika jakości niż trajektoria BCD. Wówczas łączna trajektoria stanu $ABC'D$ zapewnia lepszą wartość wskaźnika jakości niż trajektoria stanu $ABCD$, co jest sprzeczne z założeniem o optymalności trajektorii $ABCD$.

Równoważne sformułowania **zasady optymalności**:

• Każdy końcowy odcinek trajektorii optymalnej jest sam dla siebie trajektorią optymalną.

• Strategia optymalna nie zależy od historii procesu i może być określona wyłącznie na podstawie stanu procesu w danej chwili t. Trajektoria optymalna wychodząca z punktu B jest całkowicie określona przez jej stan początkowy $x(\tilde{t})$ i nie zależy od sterowania $u(t)$ dla $t < \tilde{t}$.

• Jeśli dana jest trajektoria optymalna x^o zapewniająca minimum funkcjonału $G(x, u)$ w przedziale $[t_0, t_1]$ i dane są chwile $\tilde{t}, \tilde{\tilde{t}} \in [t_0, t_1]$, to trajektoria pośrednia $x^o(t)$ $(t \in [\tilde{t}, \tilde{\tilde{t}}])$ jest również trajektorią optymalną dla problemu minimalizacji funkcjonału $G(x, u)$ w przedziale $[\tilde{t}, \tilde{\tilde{t}}]$.

Funkcja Bellmana i warunek optymalności procesu sterowania w postaci równania Hamiltona-Jacobiego-Bellmana

Definiujemy funkcję Bellmana (zwaną także funkcją jakości optymalnej) jako optymalną wartość wskaźnika jakości na końcowym odcinku trajektorii procesu $[t, t_1]$

$$S(x(t), t) \doteq \min_{u \in \mathcal{U}[t, t_1]} \int_t^{t_1} g(x(\tau), u(\tau), \tau) d\tau.$$

Z zasady optymalności wynika zależność

$$S(x(t),t) \doteq \min_{u\in\mathcal{U}[t,t_1]} \big(\int_t^{t+\epsilon\Delta t} g(x(\tau),u(\tau),\tau)d\tau + S(x(t+\epsilon\Delta t), t+\epsilon\Delta t)\big) \quad (*).$$

• Założenie: funkcja jakości optymalnej S jest różniczkowalna względem stanu x i czasu t.

Rozpatrzymy przyrost czasu $\epsilon\Delta t$ oraz odpowiadający mu przyrost stanu $\epsilon\Delta x$ i rozwiniemy w szereg Taylora wyrażenie $S(x(t)+\epsilon\Delta x, t+\epsilon\Delta t)$ (Δx i Δt są ustalone, a ϵ jest zmiennym małym parametrem):

$$S(x(t)+\epsilon\Delta x, t+\epsilon\Delta t) = S(x(t),t) + S_x(x(t),t)\epsilon\Delta x + S_t(x(t),t)\epsilon\Delta t + o(\epsilon).$$

Dla małych wartości ϵ uzyskujemy

$$\epsilon\Delta x = \epsilon f(x(t),u(t),t)\Delta t + o(\epsilon),$$

$$\int_t^{t+\epsilon\Delta t} g(x(\tau),u(\tau),\tau)d\tau = g(x(t),u(t),t)\epsilon\Delta t + o(\epsilon).$$

Z zależności (*) wynika, że

$$S(x(t),t) = \min_{u\in\mathcal{U}[t,t+\epsilon\Delta t]} \big(g(x(t),u(t),t)\epsilon\Delta t + S(x(t),t) +$$

$$S_x(x(t),t)f(x(t),u(t),t)\epsilon\Delta t + S_t(x(t),t)\epsilon\Delta t + o(\epsilon)\big).$$

Ponieważ $S(x(t),t)$ i $S_t(x(t),t)$ nie zależą od sterowania (wyrażenia te są już zminimalizowane wzgłedem sterowania w całym przedziale $[t,t_1]$), więc możemy skrócić składnik $S(x(t),t)$ po lewej i prawej stronie równania i przesunąć na lewą stronę składnik $S_t(x(t),t)$

$$-S_t(x(t),t)\epsilon\Delta t = \min_{u\in\mathcal{U}[t,t+\epsilon\Delta t]} \big(g(x(t),u(t),t)\epsilon\Delta t$$

$$+S_x(x(t),t)f(x(t),u(t),t)\epsilon\Delta t + o(\epsilon)\big).$$

Dzielimy obydwie strony przez $\epsilon\Delta t$ i przechodzimy z $\epsilon \to 0$ uzyskując

$$-S_t(x(t),t) = \min_{u(t)\in U} \big(g(x(t),u(t),t) + S_x(x(t),t)f(x(t),u(t),t)\big),$$

przy czym

$$S(x(t_1),t_1) = 0.$$

Jest to tzw. równanie Hamiltona-Jacobiego-Bellmana (równanie HJB), które stanowi warunek konieczny optymalności procesu sterowania.

Równanie to stanowi podstawę, na której określana jest

Metoda programowania dynamicznego dla problemów optymalnego sterowania z czasem ciągłym:

1. Z minimalizacji prawej strony równania HJB wyznacza się sterowanie w funkcji stanu x, pochodnej funkcji Bellmana $S_x(x,t)$ i czasu t

   $$u^o(x, S_x(x,t), t),$$

   przy czym wielkości $x,\ S_x(x,t),\ t$ traktowane są jako parametry; realizacja tego punktu wymaga rozwiązania zadania optymalizacji funkcji wielu zmiennych z ograniczeniami i z parametrami,

2. Funkcję u^o podstawia się do równania HJB redukując je do, ogólnie biorąc, nieliniowego równania różniczkowego o pochodnych cząstkowych rzędu pierwszego

   $$-S_t(x,t) = g(x, u^o(x, S_x(x,t), t), t) + S_x(x,t) f(x, u^o(x, S_x(x,t), t), t)$$

   z warunkiem granicznym
   $$S(x, t_1) = 0.$$

3. Rozwiązuje się zredukowane równanie HJB (analitycznie lub numerycznie) określając w ten sposób sterowanie optymalne w układzie ze sprzężeniem zwrotnym
   $$u^o(x,t) \doteq u^o(x, S_x(x,t), t).$$

Przykład: Wyznaczyć proces sterowania minimalizujący wskaźnik jakości

$$G(x,u) = \int_0^1 (x^2(t) + u^2(t))dt$$

z uwzględnieniem modelu procesu

$$\dot{x}(t) = ax(t) + bu(t), \quad t \in [0,1],$$

$$x(0) = x_0,$$

$$u(t) \in R.$$

Zapisujemy równanie HJB

$$-S_t(x,t) = \min_{u(t)\in R}(x^2 + u^2 + S_x(x,t)(ax+bu)), \;\; S(x,1) = 0.$$

Wyznaczamy sterowanie w funkcji pochodnej S_x

$$u^o = -0.5bS_x(x,t)$$

i uzyskujemy zredukowane równanie HJB

$$-S_t(x,t) = x^2 - 0.25b^2(S_x(x,t))^2 + S_x(x,t)ax, \;\; S(x,1) = 0.$$

Rozwiązanie tego równania można określić metodą rozdzielania zmiennych. Przewidujemy rozwiązanie w postaci

$$S(x,t) = \alpha(t)x^2.$$

Wstawiając przewidywaną postać rozwiązania do równania HJB uzyskujemy

$$-\dot{\alpha}(t)x^2 = x^2 - 0.25b^2(2\alpha(t)x)^2 + 2\alpha(t)xx, \;\; \alpha(1)x^2 = 0.$$

Prowadzi to do nieliniowego równania różniczkowego o pochodnych zwyczajnych dla $\alpha(t)$

$$-\frac{d\alpha(t)}{dt} = 1 - b^2\alpha^2(t) + 2a\alpha(t), t \in [0,1], \;\; \alpha(1) = 0,$$

którego rozwiązanie daje się określić po rozdzieleniu zmiennych α i t jako

$$\alpha(t) = tgh\big((1-t)\sqrt{a^2+b^2}\big)/\big(\sqrt{a^2+b^2} - a\; tgh\;(1-t)\sqrt{a^2+b^2}\big).$$

Tak więc funkcja Bellmana wyrazi sięwzorem

$$S(x,t) = tgh\;((1-t)\sqrt{a^2+b^2})x^2/(\sqrt{a^2+b^2} - tgh(\sqrt{a^2+b^2}(1-t)),$$

a sterowanie optymalne w układzie zamkniętym przybierze postać

$$u^o(x,t) = -tgh\;((1-t)\sqrt{a^2+b^2})x/(\sqrt{2} - tgh(\sqrt{a^2+b^2}(1-t)).$$

Sterowanie optymalne uzyskane zostało dla rozważanego przykładu w funkcji nie tylko stanu i czasu, lecz także w funkcji parametrów a i b układu. Taka postać sterowania optymalnego jest użyteczna dla jego adaptacji przy zmianie wartości parametrów.

Dla trudniejszych przykładów zastosowanie równania HJB do określenia sterowania optymalnego możliwe jest tylko przy użyciu metod numerycznych, które wyznaczają aproksymację tego sterowania z dokładnością zależną od wymiarowości problemu i od charakteru jego nieliniowości.

Sterowanie optymalne dla układów nieliniowych. Zasada optymalności Bellmana. Problemy z czasem dyskretnym.

Metoda programowania dynamicznego znajduje zastosowanie także dla problemów sterowania optymalnego z czasem dyskretnym. Nie wymaga ona w tym przypadku przyjmowania krępujących założeń o różniczkowalności funkcji Bellmana ani rozwiązywania równań o pochodnych cząstkowych. Również w tym przypadku obowiązuje zasada Bellmana - każdy końcowy odcinek dyskretnej trajektorii optymalnej jest sam dla siebie dyskretną trajektorią optymalną.

Podstawowy problem sterowania optymalnego dla układów nieliniowych z czasem dyskretnym

W podstawowym problemie sterowania optymalnego z czasem dyskretnym minimalizacji podlega sumaryczny wskaźnik jakości

$$G(x,u) \doteq \sum_{k=0}^{K-1} g(x(k),u(k),k)$$

z uwzględnieniem ograniczeń w postaci dyskretnego równania stanu procesu

$$x(k+1) = f(x(k),u(k),k), \quad k = 0,1,...,K-1,$$

z zadanym warunkiem początkowym

$$x(0) = x_0$$

oraz ograniczeń w postaci dopuszczalnego chwilowego zakresu wartości sterowania rozszerzonego

$$u(k) \in U, \quad k = 0,1,...,K-1,$$

gdzie funkcje

$$g : R^n \times R^m \times R \to R, \quad f : R^n \times R^m \times R \to R^n$$

są, ogólnie biorąc, nieliniowe.

Definiujemy funkcję Bellmana dla końcowego odcinka dyskretnej trajektorii stanu

$$S(x(k), k) \doteq \min_{u(\kappa)\in U,\ \kappa=k,k+1,\dots,K-1} \sum_{\kappa=k}^{K-1} g(x(\kappa), u(\kappa), \kappa).$$

Na podstawie dyskretnej zasady optymalności zapisujemy zależność

$$S(x(k), k) = \min_{u(k)\in U} \big(g(x(k), u(k), k) + S(x(k+1), k+1)\big)$$

i podstawiamy do niej równanie stanu dla k-tego przedziału

$$x(k+1) = f(x(k), u(k), k)$$

uzyskując **równanie rekurencyjne Bellmana dla dyskretnych procesów sterowania**

$$S(x(k), k) = \min_{u(k)\in U} \big(g(x(k), u(k), k) + S(f(x(k), u(k), k), k+1)\big),$$

$$k = K-1, K-2, \dots, 0, \quad S(x(K), K) = 0.$$

Równanie to stanowi podstawę dyskretnej metody programowania dynamicznego, która redukuje wyznaczanie całego ciągu sterowań optymalnych

$$\{u^o(k),\ k = 0, 1, \dots, K-1\}$$

do kolejnego wyznaczania poszczególnych sterowań $u^o(k)$ z równania rekurencyjnego Bellmana. Taka dekompozycja wyznaczania sterowania optymalnego jest zaletą tej metody. Jednak związana jest ona z koniecznością rozwiązywania parametrycznych zadań optymalizacji tj. wyznaczania sterowania optymalnego w funkcji stanu $u^o(k) = u^o(x(k), k)$. W ogólnym przypadku można to zrobić tylko numerycznie np. w postaci tablic wartości.

Metoda programowania dynamicznego dla problemów optymalnego sterowania z czasem dyskretnym:

1. Podstawia się $k = K-1$ i rozwiązuje się zadanie optymalizacji ostatniego etapu procesu

$$S(x(K-1), K-1) = \min_{u(K-1) \in U} g(x(K-1), u(K-1), K-1)$$

wyznaczając sterowanie optymalne ostatniego etapu procesu $u^o(K-1) = u^o(x(K-1), K-1)$ w funkcji stanu początkowego tego etapu $x(K-1)$.

2. Na k-tym etapie procesu, korzystając ze znajomości funkcji Bellmana $S(x(k+1), k+1)$ rozwiązuje się zadanie optymalizacji k-tego etapu wynikające z równania rekurencyjnego Bellmana

$$S(x(k), k) = \min_{u(k) \in U} \big(g(x(k), u(k), k) + S(f(x(k), u(k), k), k+1)\big)$$

wyznaczając sterowanie optymalne k-tego etapu procesu $u^o(k) = u^o(x(k), k)$ w funkcji stanu początkowego tego etapu $x(k)$.

3. Po osiągnięciu etapu początkowego $k = 0$ wyznaczamy konkretną wartość sterowania optymalnego dla tego etapu $u^o(0) = u^o(x_0, 0)$ korzystając ze znajomości stanu początkowego $x(0) = x_0$. Następnie wyznaczamy kolejno konkretne wartości sterowań optymalnych na podstawie zależności

$$x^o(k) = f(x^o(k-1), u^o(k-1), k-1),\ u^o(k) = u^o(x^o(k), k),\ k = 1, 2, ..., K-1.$$

Algorytm ten nazywany jest **algorytmem podstawowym dyskretnego programowania dynamicznego**(algorytm z cofaniem się, the backward algorithm).

Przykład: zminimalizować wskaźnik jakości dyskretnego procesu sterowania $(K = 3)$

$$G(x, u) = \sum_{k=0}^{2} (x^2(k) + u^2(k))$$

przy ograniczeniach

$$x(k+1) = ax(k) + bu(k),\ k = 0, 1, 2,$$

$$x(0) = x_0,$$

$$u(k) \in R, \; k = 0, 1, 2.$$

Rozwiązujemy zadanie ostatniego etapu

$$S(x(2), 2) = \min_{u(2) \in R} (x^2 + u^2) \Rightarrow u^o(2) = 0, \; S(x(2), 2) = x^2(2).$$

Zapisujemy równanie Bellmana dla przedostatniego etapu

$$S(x(1), 1) = \min_{u(1) \in R} \big(x^2(1) + u^2(1) + x^2(2) \big)$$

i redukujemy je do równania optymalizacji tego etapu

$$S(x(1), 1) = \min_{u(1) \in R} \big(x^2(1) + u^2(1) + (ax(1) + bu(1))^2 \big).$$

Celem rozwiązania zadania optymalizacji tego etapu przyrównujemy do zera pochodną wyrażenia w nawiasie względem sterowania $u(1)$

$$2u(1) + 2(ax(1) + bu(1))b = 0$$

i określamy sterowanie optymalne w funkcji stanu początkowego tego etapu

$$u^o(1) = -\frac{ab}{1+b^2} x(1)$$

oraz funkcję Bellmana tego etapu

$$S(x(1), 1) = x^2(1) + \frac{a^2 b^2}{(1+b^2)^2} x^2(1) + \big(ax(1) - \frac{ab^2}{1+b^2} x(1) \big)^2 = cx^2(1),$$

gdzie $c = \frac{a^2+1+b^2}{1+b^2}$. Zapisujemy równanie Bellmana dla początkowego etapu

$$S(x(0), 0) = \min_{u(0) \in R} \big(x^2(0) + u^2(0) + cx^2(1) \big)$$

i jego postać zredukowaną

$$S(x(0), 0) = \min_{u(0) \in R} \big(x^2(0) + u^2(0) + c(ax(0) + bu(0))^2 \big).$$

Celem wyznaczenia sterowania optymalnego etapu początkowego obliczamy pochodną wyrażenia w nawiasie względem sterowania $u(0)$

$$2u(0) + 2c(ax_0 + bu(0))b = 0.$$

Tak więc sterowanie optymalne etapu początkowego ma postać

$$u^o(0) = -\frac{abc}{1+cb^2}x_0.$$

Określamy teraz kolejne optymalne stany i optymalne sterowania

$$x^o(1) = ax_0 + bu^o(0) = ax_0 - \frac{ab^2c}{1+cb^2}x_0 = \frac{a}{1+cb^2}x_0,$$

$$u^o(1) = -\frac{ab}{1+b^2}x^o(1) = -\frac{a^2b}{(1+b^2)^2}x_0,$$

$$x^o(2) = ax^o(1) + bu^o(1) = \frac{a^2}{1+cb^2}x_0 - \frac{a^2b^2}{(1+b^2)^2}x_0,$$

$$u^o(2) = 0.$$

Istotną zaletą metody dyskretnego programowania dynamicznego jest możliwość jej zastosowania do wyznaczania sterowania optymalnego dla procesów, na które oddziaływują zakłócenia przypadkowe. Minimalizacji podlega wtedy wartość oczekiwana wskaźnika jakości

$$G(x,u) = E\{\sum_{k=0}^{K-1} g(x(k),u(k),\zeta(k),k)\}$$

przy ograniczeniach

$$x(k+1) = f(x(k),u(k),\zeta(k),k), \quad k = 0,1,...,K-1,$$

$$x(0) = x_0,$$

$$u(k) \in U, \quad k = 0,1,...,K-1,$$

gdzie $\zeta(k)$ jest zakłóceniem przypadkowym. Dlatego stany $x(k)$ i funkcja kryterialna g stają się zmiennymi przypadkowymi. Założymy, że

$$\zeta(0), \zeta(1), ..., \zeta(K-1)$$

są niezależnymi zmiennymi przypadkowymi o znanych rozkładach prawdopodobieństwa

$$P(\zeta(0)), P(\zeta(1)), ..., P(\zeta(K-1)).$$

W niektórych przypadkach wystarczająca moż być znajomość wartości średniej i wariancji zakłócenia przypadkowego bez znajomości zakłócenia przypadkowego. Przy tych założeniach można sformułować następujące uogólnienie metody dyskretnego programowania dynamicznego na przypadek zakłóceń przypadkowych w równaniach stanu procesu.

Metoda stochastycznego programowania dynamicznego dla problemów optymalnego sterowania z czasem dyskretnym:

1. Podstawia się $k = K-1$ i rozwiązuje się zadanie optymalizacji ostatniego etapu procesu

   $$S(x(K-1), K-1) = \min_{u(K-1)\in U} E\{g(x(K-1), u(K-1), \zeta(K-1), K-1)\}$$

   $$= \min_{u(K-1)\in U} \int_{-\infty}^{+\infty} P(\zeta(K-1)) g(x(K-1), u(K-1), \zeta(K-1), K-1) d\zeta(K-1)$$

   wyznaczając sterowanie optymalne ostatniego etapu procesu $u^o(K-1) = u^o(x(K-1), K-1)$ w funkcji stanu początkowego tego etapu $x(K-1)$.

2. Na k-tym etapie procesu, korzystając ze znajomości funkcji Bellmana $S(x(k+1), k+1)$ rozwiązuje się zadanie optymalizacji k-tego etapu wynikające z równania rekurencyjnego Bellmana

   $$S(x(k), k) = \min_{u(k)\in U} E\{\big(g(x(k), u(k), \zeta(k), k) + S(f(x(k), u(k), \zeta(k), k+1)\big)\}$$

   $$= \min_{u(k)\in U} \int_{-\infty}^{+\infty} P(\zeta(k)) \big(g(x(k), u(k), \zeta(k), k) + S(f(x(k), u(k), \zeta(k), k+1)\big) d\zeta(k)$$

   wyznaczając sterowanie optymalne k-tego etapu procesu $u^o(k) = u^o(x(k), k)$ w funkcji stanu początkowego tego etapu $x(k)$.

3. Po osiągnięciu etapu początkowego $k = 0$ wyznaczamy konkretną wartość sterowania optymalnego dla tego etapu $u^o(0) = u^o(x_0, 0)$ korzystając ze znajomości stanu początkowego $x(0) = x_0$. Następnie wyznaczamy kolejno konkretne wartości sterowań optymalnych na podstawie zależności

   $$x^o(k) = f(x^o(k-1), u^o(k-1), \zeta(k-1), k-1),$$

   $$u^o(k) = u^o(x^o(k), k), \; k = 1, 2, ..., K-1.$$

Przykład: zminimalizować wskaźnik jakości procesu

$$G(x,u) = E\{\sum_{k=1}^{K} x^2(k)\}$$

przy ograniczeniach

$$x(k+1) = ax(k) + u(k)\zeta(k), \quad k = 0,1,...,K-1,$$

$$x(0) = x_0,$$

$$u(k) \in R, \quad k = 0,1,...,K-1.$$

Zakładamy, że $\zeta(k)$ jest procesem przypadkowym o wartości średniej równej $\bar{z}$ i wariancji σ^2.

Przekształcimy wskaźnik jakości do postaci standardowej korzystając z równania obiektu

$$G = E\{\sum_{k=0}^{K-1} x^2(k+1)\} = E\{\sum_{k=0}^{K-1} (ax(k) + u(k)\zeta(k))^2$$

$$= E\{\sum_{k=0}^{K-1} \big(a^2x^2(k) + u^2(k)\zeta^2(k) + 2ax(k)u(k)\zeta(k)\big).$$

Wyznaczamy funkcję Bellmana ostatniego etapu procesu

$$S(x(K-1), K-1) = \min_{u(K-1)\in U} E\{\big(a^2x^2(K-1)$$

$$+u^2(K-1)\zeta^2(K-1) + 2ax(K-1)u(K-1)\zeta(K-1)\big)\}$$

$$= \min_{u(K-1)} \big(a^2x^2(K-1)+u^2(K-1)E\{\zeta^2(K-1)\}+2ax(K-1)u(K-1)E\{\zeta(K-1)\}\big)$$

$$= \min_{u(K-1)} \big(a^2x^2(K-1) + u^2(K-1)(\bar{z}^2 + \sigma^2) + 2ax(K-1)u(K-1)\bar{z}\big).$$

Obliczamy minimum ostatniego wyrażenia przyrównując do zera jego pochodną względem $u(K-1)$

$$2u(K-1)(\bar{z}^2 + \sigma^2) + 2ax(K-1)\bar{z} = 0,$$

skąd dostajemy sterowanie optymalne ostatniego etapu

$$u^o(K-1) = -ax(K-1)\bar{z}/(\bar{z}^2+\sigma^2)$$

i funkcję Bellmana tego etapu

$$S(x(K-1), K-1) = a^2x^2(K-1) + \frac{a^2x^2(K-1)\bar{z}^2}{\bar{z}^2+\sigma^2}$$

$$-\frac{2a^2x^2(K-1)\bar{z}^2}{\bar{z}^2+\sigma^2}$$

$$= cx^2(K-1), \quad c \doteq a^2\big(1-\bar{z}^2/(\bar{z}^2+\sigma^2)\big).$$

Zapisujemy równanie Bellmana na przedostatnim etapie procesu

$$S(x(K-2), K-2) = \min_{u(K-2)} E\{\big(a^2x^2(K-2)$$

$$+u^2(K-2)\zeta^2(K-2) + 2ax(K-2)u(K-2)\zeta(K-2) + S(x(K-1), K-1)\big)\}$$

$$= \min_{u(K-2)} E\{\big(a^2x^2(K-2)$$

$$+u^2(K-2)\zeta^2(K-2) + 2ax(K-2)u(K-2)\zeta(K-2) + cx^2(K-1)\big)\}$$

$$= \min_{u(K-2)} E\{\big(a^2x^2(K-2) + u^2(K-2)\zeta^2(K-2) + 2ax(K-2)u(K-2)\zeta(K-2)$$

$$+c(a^2x^2(K-2) + u^2(K-2)\zeta^2(K-2) + 2ax(K-2)u(K-2)\zeta(K-2))\big)\}$$

$$= (1+c) \min_{u(K-2)} E\{\big(a^2x^2(K-2)+u^2(K-2)\zeta^2(K-2)+2ax(K-2)u(K-2)\zeta(K-2)\big)\}.$$

Tak więc na przedostatnim etapie procesu uzyskaliśmy dla określenia sterowania optymalnego tego etapu identyczne zadanie optymalizacji jak poprzednio, a zatem

$$u^o(K-2) = -ax(K-2)\bar{z}/(\bar{z}^2+\sigma^2).$$

Sytuacja powtarza się na następnych etapach metody, co prowadzi do ogólnego algorytmu sterowania optymalnego procesu z zakłóceniem przypadkowym

$$u^o(k) = -ax(k)\bar{z}/(\bar{z}^2+\sigma^2), \quad k = 0, 1, ..., K-1.$$

Sterowanie optymalne jest w tym przypadku liniową funkcją stanu i zależy od charakterystyk probabilistycznych zakłócenia - od jego wartości średniej i wariancji.

Metoda dyskretnego programowania dynamicznego posiada wiele wariantów związanych z optymalizacją wielostopniowych procesów technologicznych realizowanych w kaskadach aparatów, a także w aparatach typu półkowego. Indeks k nie oznacza wtedy czasu dyskretnego lecz numer kaskady lub numer półki. Dla niektórych systemów tego typu wygodnie jest stosować odwrotną numerację stopni procesu - w kierunku zmniejszających się wartości indeksów $K, K-1, ..., 1$, a równanie stanu k-tego stopnia zapisywać w postaci

$$x(k-1) = f(x(k), u(k), k), \quad k = K, K-1, ..., 1,$$

przy czym

$$x(K) = x_K$$

jest w tym przypadku zadanym stanem początkowym pierwszego stopnia, $x(k-1)$ jest stanem wyjściowym k-tego stopnia, a $x(k)$ jego stanem wejściowym. Wskaźnik jakości zapisujemy jako

$$G(x, u) = \sum_{k=1}^{K} g(x(k), u(k), k).$$

Funkcję Bellmana definiujemy w postaci

$$S(x(k), k) \doteq \min_{u(\kappa) \in U, \kappa=1,...,k} \sum_{\kappa=1}^{\kappa=k} g(x(\kappa), u(\kappa), \kappa).$$

Uzyskujemy wtedy odwrotne równanie Bellmana

$$S(x(k), k) = \min_{u(k) \in U} \big(g(x(k), u(k), k) + S(x(k-1), k-1)\big)$$

i wynikający stąd **algorytm odwrotny dyskretnego programowania dynamicznego** bazujący na odwrotnym równaniu rekurencyjnym

$$S(x(k), k) = \min_{u(k) \in U} \big(g(x(k), u(k), k) + S(f(x(k), u(k), k), k-1)\big), \quad k = 1, 2, ..., K,$$

przy czym $S(x(0), 0) = 0$.

Przykład: Optymalizacja kaskady ekstraktorów. Do każdej kaskady doprowadzany jest rozpuszczalnik z natężeniem $u(k)$, za pomocą którego ze strumienia wejściowego ekstrahowany jest składnik użyteczny. Stężenie strumienia wejściowego kaskady zmniejsza się z kaskady na kaskadę wskutek ekstrahowania składnika użytecznego zgodnie z równaniem

$$x(k-1) = x(k)/(1+au(k)), \quad k = K, K-1, ..., 1.$$

Zysk z procesu k-tej kaskady można określić jako różnicę między ilością ekstrahowanego składnika użytecznego $x(k) - x(k-1)$ i kosztami rozpuszczalnika $cu(k)$. Tak więc maksymalizacja zysku z procesu w całej kaskadzie jest równoważna minimalizacji wyrażenia

$$G(x,u) = \sum_{k=1}^{K} (cu(k) - (x(k) - x(k-1)).$$

Należy określić dyskretne sterowanie kaskady $u(k), \; k = 1, 2, ..., K$ (tj. natężenia dopływu rozpuszczalnika do poszególnych stopni kaskady) tak, aby zmaksymalizować zysk z procesu w całej kaskadzie.

Stosując algorytm odwrotny dyskretnego programowania dynamicznego wyznaczamy

$$S(x(1), 1) = \min_{u(1)} (cu(1) - x(1) + x(0)) = \min_{u(1)} (cu(1) - x(1) + x(1)/(1 + cu(1))),$$

skąd wyznaczamy sterowanie optymalne kaskady o numerze 1 (przyrównując do zera pochodną wyrażenia w nawiasie względem $u(1)$)

$$u^o(1) = \big(\sqrt{ax(1)/c} - 1\big)/a,$$

oraz funkcję Bellmana dla tej kaskady

$$S(x(1), 1) = \big(\sqrt{x(1)} - \sqrt{c/a}\big)^2.$$

Korzystamy z odwrotnego równania rekurencyjnego Bellmana

$$S(x(2), 2) = \min_{u(2)} \big(cu(2) - x(2) + x(1) + S(x(1), 1)\big)$$

tj. z równania

$$S(x(2), 2) = \min_{u(2)} \big(cu(2) - x(2) + x(1) + \sqrt{x(1)} - \sqrt{c/a}\big)$$

$$= \min_{u(2)} \big(cu(2) - x(2) + x(2)/(1 + cu(2)) + \sqrt{x(2)/(1 + cu(2))} - \sqrt{c/a}\big),$$

skąd wyznaczamy sterowanie optymalne kaskady o numerze 2 (przyrównując do zera pochodną wyrażenia w nawiasie względem $u(2)$)

$$u^o(2) = \big((ax(2)/c)^{1/3} - 1\big)/a$$

oraz funkcję Bellmana dla tej kaskady

$$S(x(2), 2) = x(2) - 3(c/a)^{2/3} x(2)^{1/3} + 2c/a.$$

Kontynuując ten proces uzyskuje się ogólny wzór na sterowanie optymalne w układzie zamkniętym dla kaskady ekstraktorów

$$u^o(k) = \big((ax(k)/c)^{1/(k+1)} - 1\big)/a, \quad k = K, K-1, ..., 1.$$

Uogólnione medele dyskretnych procesów sterowania

Duże znaczenie praktyczne mają modele procesów z czasem dyskretnym interpretowane jako aproksymacje modeli procesów z czasem ciągłym. Zastosowanie prawostronnego ilorazu różnicowego do aproksymacji pochodnej prowadzi do modelu chronologicznej dynamiki procesu z czasem dyskretnym

$$x(k+1) = x(k) + \tau(k) f(x(k), u(k), k), \quad k = 0, 1, ..., K-1,$$

czyli

$$x(k+1) = \tilde{f}(x(k), \tilde{u}(k), k), \quad k = 0, 1, ..., K-1,$$
$$\tilde{f}(x(k), \tilde{u}(k), k) \doteq x(k) + \tau(k) f(x(k), u(k), k),$$

przy czym $\tau(k)$ jest długością k-tego przedziału dyskretyzacji traktowaną jako dodatkowa zmienna optymalizacyjna k-tego etapu procesu oprócz sterowania dyskretnego $u(k)$ tego etapu procesu. Para $(\tau(k), u(k))$ nazywana jest sterowaniem rozszerzonym procesu dyskretnego.

Zastosowanie lewostronnego ilorazu różnicowego do aproksymacji pochodnej prowadzi do modelu antychronologicznej dynamiki procesu z czasem dyskretnym

$$\dot{x}(t) = f(x(t), u(t), t) \Rightarrow x(k-1) = x(k) + \tau(k) f(x(k), u(k), k), \quad k = K, K-1, ..., 1,$$

czyli

$$x(k-1) = \tilde{f}(x(k), \tilde{u}(k), k), \quad k = K, K-1, ..., 1,$$

$$\tilde{f}(x(k), \tilde{u}(k), k) \doteq x(k) + \tau(k) f(x(k), u(k), k),$$

przy czym $\tau(k)$ jest długością k-tego przedziału dyskretyzacji traktowaną jako dodatkowa zmienna optymalizacyjna k-tego etapu procesu oprócz sterowania dyskretnego $u(k)$ tego etapu procesu. Para $(\tau(k), u(k)$ nazywana jest sterowaniem rozszerzonym procesu dyskretnego z antychronologiczną dynamiką. Takie modele dynamiki procesów dyskretnych posiadają również ważne interpretacje dla optymalizacji wielostopniowych procesów technologicznych realizowanych w kaskadach aparatów, a także w aparatach typu półkowego. Dla takich procesów $\tau(k)$ może określać tzw. średni czas przebywania substancji w k-tym stopniu kaskady lub mieć interpretację nie związaną z czasem - może to być parametr k-tej kaskady odgrywający szczególną rolę w jej modelu np. objętość kaskady.

W podstawowym problemie sterowania optymalnego z czasem dyskretnym minimalizacji podlega sumaryczny wskaźnik jakości

$$G(x, u) \doteq \sum_{k=0}^{K-1} \tilde{g}(x(k), u\tilde{(k)}, k)$$

$$(\tilde{g}(x(k), u\tilde{(k)}, k) \doteq \tau(k) g(x(k), u(k), k))$$

z uwzględnieniem ograniczeń w postaci dyskretnego równania stanu procesu

$$x(k) = x(k-1) + \tau(k) f(x(k), u(k), k), \quad k = 1, 2, ..., K,$$

z zadanym warunkiem początkowym

$$x(0) = x_0$$

oraz ograniczeń w postaci dopuszczalnego chwilowego zakresu wartości sterowania rozszerzonego

$$(\tau(k), u(k)) \in \Omega(k), \quad k = 1, 2, ..., K,$$

gdzie funkcje

$$g : R^n \times R^m \times R \to R, \quad f : R^n \times R^m \times R \to R^n$$

są, ogólnie biorąc, nieliniowe. Dla wielu przykładów można przyjąć $\Omega(k) = R_0 \times U$, gdzie R_0 jest zbiorem nieujemnych liczb rzeczywistych.

Wprowadzając dyskretną funkcję Bellmana

$$S(x(k),k) \doteq \min_{(\tau(\kappa),u(\kappa))\in\Omega(\kappa),\ \kappa=k,k+1,\ldots,K} \sum_{\kappa=k}^{K} \tau(\kappa) g(x(\kappa),u(\kappa),\kappa)$$

i stosując dyskretną zasadę optymalności uzyskuje się równanie rekurencyjne z uwzględnieniem sterowania rozszerzonego

$$S(x(k),k) = \min_{(\tau(\kappa),u(\kappa))\in\Omega(\kappa)} (\tau(k) g(x(k),u(k),k)) + S(x(k+1),k+1)).$$

Mimo wielu udanych zastosowań metoda programowania dynamicznego posiada wady znacznie ograniczające jej praktyczne wykorzystanie. Jedną z wad jest konieczność rozwiązywania parametrycznego problemu optymalizacji tj. wyznaczania sterowania w funkcji stanu na każdym etapie procesu. Ogólnie biorąc można dokonać tego tylko metodami numerycznymi, przy czym liczba niezbędnych obliczeń lawinowo narasta wraz ze wzrostem wymiaru stanu problemu (jest tzw. *"przekleństwo wielowymiarowości"* związane z metodą programowania dynamicznego). Z drugiej strony duża moc obliczeniowa współczesnych komputerów poszerza zakres zastosowań metody programowania dynamicznego.