

Programmation dynamique

Modèles déterministes, temps continu

Fabian Bastin

IFT-6521 – Hiver 2011

Commande en temps continu

Système dynamique en temps continu, évoluant selon:

$$\dot{x}(t) = f(x(t), u(t)), \quad 0 \leq t \leq T, \quad (1)$$

$x(0)$ fixé,

$x(t) \in \mathbb{R}^n$ = état au temps t (vecteur colonne),

$\dot{x}(t) \in \mathbb{R}^n$ = vecteur des dérivées par rapport à t ,

$u(t) \in U \subset \mathbb{R}^m$ = valeur de la commande au temps t .

On dénote les composantes i de x , $\dot{x}$, u et f par x_i , $\dot{x}_i$, u_i et f_i , respectivement. L'équation (1) s'écrit alors

$$\dot{x}_i(t) = \frac{dx_i(t)}{dt} = f_i(x(t), u(t)), \quad 0 \leq t \leq T, \quad 1 \leq i \leq n.$$

On supposera que les f_i sont continûment différentiables par rapport à x et continues par rapport à u .

Une **commande admissible** est une fonction $\{u(t), t \geq 0\}$, continue par morceaux, et telle que $u(t) \in U$ pour $t \in [0, T]$.

On **suppose** ici qu'à une commande admissible donnée correspond une **trajectoire** $\{x^u(t), 0 \leq t \leq T\}$ qui est l'unique solution de (1).

On cherche une commande admissible qui minimise la **fonction de coût**:

$$h(x^u(T)) + \int_0^T g(x^u(t), u(t)) dt$$

où g et h sont continument différentiables par rapport à x et g est continue par rapport à u .

Équation de Hamilton-Jacobi-Bellman (HJB)

Dérivation très informelle.

Soit $J(t, x)$ le coût optimal pour l'intervalle $[t, T]$, si $x(t) = x$:

$$J(t, x) = \min_u \left[h(x^u(T)) + \int_t^T g(x^u(s), u(s)) ds \right].$$

Divisons l'horizon $[0, T]$ en N morceaux, suivant la discrétisation

$$\delta = \frac{T}{N}.$$

Soit

$$x_k = x(k\delta), \quad k = 0, 1, \dots, N$$

$$u_k = u(k\delta), \quad k = 0, 1, \dots, N,$$

et nous approximations le système en temps continu par

$$x_{k+1} = x_k + f(x_k, u_k)\delta.$$

Équation de Hamilton-Jacobi-Bellman (HJB) (suite)

De la même manière, nous discrétisons la fonction de coût par

$$h(x_N) + \sum_{k=0}^{N-1} g(x_k, u_k) \delta.$$

Appliquons l'algorithme de programmation dynamique au modèle discrétisé. Les équations de programmation dynamique sont

$$\begin{aligned}\tilde{J}(N\delta, x) &= h(x), \\ \tilde{J}(k\delta, x) &= \min_{u \in U} [g(x, u)\delta + \tilde{J}((k+1)\delta, x + f(x, u)\delta)], \\ k &= 0, \dots, N-1.\end{aligned}$$

En supposant que $\tilde{J}$ est suffisamment lisse, le développement de Taylor au premier ordre donne:

$$\begin{aligned}\tilde{J}((k+1)\delta, x + f(x, u)\delta) &= \tilde{J}(k\delta, x) + \nabla_t \tilde{J}(k\delta, x) \delta \\ &\quad + \nabla_x \tilde{J}(k\delta, x)^\top f(x, u) \delta + o(\delta).\end{aligned}$$

Équation de Hamilton-Jacobi-Bellman (HJB) (suite)

Supposons que la fonction en temps discrète converge vers sa contrepartie continue quand $\delta \rightarrow 0$:

$$\lim_{k \rightarrow \infty, \delta \rightarrow 0, k\delta = t} \tilde{J}(k\delta, x) = J(t, x),$$

On simplifie les $\tilde{J}(k\delta, x)$, on divise par δ , pour obtenir l'équation de **Hamilton-Jacobi-Bellman (HJB)**:

$$0 = \min_{u \in U} \left[g(x, u) + \nabla_t J(t, x) + \nabla_x J(t, x)^\top f(x, u) \right], \quad (2)$$

$$J(T, x) = h(x). \quad (3)$$

Conditions suffisantes d'optimalité

Proposition:

Supposons que $V(t, x)$ est continûment différentiable en t et en x , et est une solution de (2-3). Si $u = u^*(t) = \mu^*(t, x)$ fait atteindre le minimum dans (2) pour tout t , si $\{x^*(t), 0 \leq t \leq T\}$ est l'unique trajectoire qui correspond à cette commande (avec $x^*(0) = x(0)$), et si cette commande est continue par morceaux par rapport à t , alors $V(t, x) = J(t, x)$ pour tout t et tout x , et la politique μ^* est optimale.

Preuve: voir DPOC, Proposition 3.2.1.

Idée: on essaie de “deviner” une la forme de J , puis on **vérifie** si notre candidat satisfait HJB.

Mais comment trouve-t-on de bons candidats?

- A. Parfois directement, par une bonne compréhension du problème.
- B. En utilisant le principe du minimum (à venir).
- C. Discrétiser t et x , puis résoudre par PD.

Résolution numérique

Approche directe: discrétiser le problème original par rapport à t et x , puis utiliser les équations de récurrence habituelles en temps discret.

Techniques numériques pour les équations aux dérivées partielles (e.g., éléments finis, etc.).

HJB tient dans tout l'espace des valeurs de (t, x) . Mais résoudre numériquement dans tout l'espace peut devenir coûteux.

En fait, dans le cas déterministe, il suffit de résoudre l'équation **le long de la trajectoire optimale**.

Pas besoin de résoudre avec beaucoup de précision partout. On peut résoudre grossièrement pour avoir une bonne idée de la trajectoire optimale, puis construire un **tube** autour de la trajectoire optimale et raffiner l'approximation dans ce tube. On peut itérer cette approche.

Principe du minimum de Pontryagin

(Dérivation très informelle.)

L'idée est de trouver des conditions **nécessaires** d'optimalité en se basant uniquement sur ce qui se passe le long de la trajectoire optimale. Rappel de HJB:

$$\begin{aligned} 0 &= g(x^*(t), u^*(t)) + \nabla_t J(t, x^*(t)) + \nabla_x J(t, x^*(t))^T f(x^*(t), u^*(t)) \\ &= g(x^*(t), u^*(t)) + p_0(t) + p(t)^T f(x^*(t), u^*(t)) \\ &= p_0(t) + H(x^*(t), u^*(t), p(t)), \end{aligned}$$

où $p(t) = \nabla_x J(t, x^*(t))$, $p_0(t) = \nabla_t J(t, x^*(t))$, et

$$H(x(t), u(t), p(t)) = g(x(t), u(t)) + p(t)^T f(x(t), u(t))$$

est la **fonction Hamiltonienne** (fonction de t).

En **dérivant** HJB par rapport à x et par rapport à t , on obtient:

$$\begin{aligned} 0 &= \nabla_x g(x^*(t), u^*(t)) + \nabla_{xt}^2 J(t, x^*(t)) \\ &\quad + \nabla_{xx}^2 J(t, x^*(t)) f(x^*(t), u^*(t)) \\ &\quad + \nabla_x f(x^*(t), u^*(t)) \nabla_x J(t, x^*(t)), \end{aligned} \quad (4)$$

$$0 = \nabla_{tt}^2 J(t, x^*(t)) + \nabla_{xt}^2 J(t, x^*(t))^T f(x^*(t), u^*(t)), \quad (5)$$

qui se réécrit, en dérivant, puis en utilisant (4):

$$\dot{p}_0(t) = \frac{d}{dt} \nabla_t J(t, x^*(t)) = 0 \quad (\text{de (7), et lemme 3.3.1 de DOCP}),$$

et

$$\begin{aligned} \dot{p}(t) &= \frac{d}{dt} \nabla_x J(t, x^*(t)) \\ &= \nabla_{xt}^2 J(t, x^*(t)) + \nabla_{xx}^2 J(t, x^*(t)) f(x^*(t), u^*(t)) \\ &= -\nabla_x g(x^*(t), u^*(t)) - \nabla_x f(x^*(t), u^*(t)) p(t), \end{aligned}$$

où $\nabla_x f(x^*(t), u^*(t))$ est le Jacobien de f .

L'équation pour $\dot{p}(t)$, appelée **l'équation adjointe**, se réécrit

$$\dot{p}(t) = -\nabla_x H(x^*(t), u^*(t), p(t)) \quad (6)$$

avec la condition terminale

$$p(T) = \nabla_x h(x^*(T)). \quad (7)$$

En utilisant HJB et $\dot{p}_0(t) = 0$, on obtient aussi que

$$H(x^*(t), u^*(t), p(t)) = -\nabla_t J(t, x^*(t)) = -p_0(t) = C, \quad (8)$$

une constante. Finalement, pour tout $t \in [0, T]$, en vertu de (2),

$$u^*(t) = \arg \min_{u \in U} H(x^*(t), u, p(t)). \quad (9)$$

Proposition: **Principe du minimum de Pontryagin.**

Si u^* est une commande optimale, alors la trajectoire correspondante satisfait (8) et (9), où p est une solution de l'équation adjointe.

Donne des **conditions nécessaires** (mais pas suffisantes) d'optimalité. Permet de trouver de bons **candidats** de solutions, que l'on peut ensuite vérifier en vérifiant les conditions de HJB.

Note: Si f et/ou g dépendent explicitement de t , alors le Hamiltonien n'est plus constant le long de la trajectoire optimale, mais les autres conditions tiennent (voir Section 3.4.4.).

Résumé.

En pratique, on doit résoudre les **équation différentielles**:

$$\begin{aligned}\dot{x}^*(t) &= f(x^*(t), u^*(t)), \\ \dot{p}(t) &= -\nabla_x H(x^*(t), u^*(t), p(t)),\end{aligned}$$

avec les conditions initiales et terminales

$$x^*(0) \text{ fixé et } p(T) = \nabla_x h(x^*(T)).$$

Il existe une riche collection de **méthodes numériques** pour résoudre ces équations.

Exemple: problème de production

Un producteur dispose d'une capacité de production $x(t)$ au temps t , et réinvestit une proportion $u(t)$ de sa production et stocke (ou utilise, ou consomme) une proportion $1 - u(t)$, au temps t .

Le capacité de production évolue selon

$$\dot{x}(t) = \gamma u(t)x(t)$$

où $\gamma > 0$ est une constante.

Le producteur veut maximiser la quantité totale utilisée ou stockée,

$$\int_0^T (1 - u(t))x(t)dt,$$

sous les contraintes: $0 \leq u(t) \leq 1$ pour tout t , et $x(0) > 0$ fixé.

On a $f(x(t), u(t)) = \gamma u(t)x(t)$ et $g(x(t), u(t)) = (1 - u(t))x(t)$.

Hamiltonien: $H(x(t), u(t), p(t)) = (1 - u(t))x(t) + p(t)\gamma u(t)x(t)$.

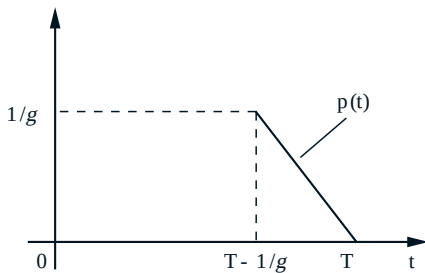
Équation adjointe: $p(T) = 0$ et

$$\dot{p}(t) = -\nabla_x H(x^*(t), u^*(t), p(t)) = u^*(t)(1 - \gamma p(t)) - 1.$$

On maximise le Hamiltonien par rapport à $u \in [0, 1]$:

$$u^*(t) = \begin{cases} 0 & \text{si } p(t) < 1/\gamma; \\ 1 & \text{si } p(t) \geq 1/\gamma. \end{cases}$$

Puisque $p(T) = 0$, pour t proche de T , on aura $p(t) < 1/\gamma$, et donc $u^*(t) = 0$ et $\dot{p}(t) = -1$.

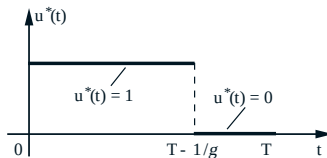
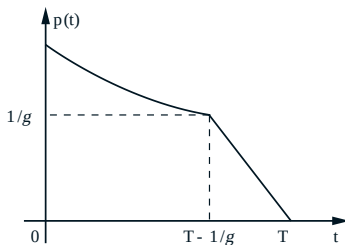


On voit ainsi que $p(t) = 1/\gamma$ lorsque $t = T - 1/\gamma$.

Pour $t < 1/\gamma$, on a alors $u^*(t) = 1$.

Cela donne $\dot{p}(t) = -\gamma p(t)$, ou $p(t) = \alpha e^{-\gamma t}$, avec $\alpha = \frac{1}{\gamma} e^{\gamma T - 1}$.

Puisque cette solution est la seule qui satisfait au principe du minimum, elle satisfait aussi nécessairement à HJB.



Exemple: construction de route

On veut construire une route de 0 à T sur un terrain dont la hauteur au point t est $z(t)$ (connu). La hauteur de la route au point t sera $x(t)$. On fera de l'excavation ou du remplissage au besoin. La pente de la route ne doit jamais dépasser $a > 0$. On veut minimiser

$$\frac{1}{2} \int_0^T (x(t) - z(t))^2 dt.$$

sous les contraintes

$$\dot{x}(t) = u(t) \quad \text{et} \quad |u(t)| \leq a \quad \text{pour tout } t.$$

Un défaut majeur de cette formulation:

La pente peut changer brusquement. On pourrait plutôt mettre une borne sur la dérivée seconde, par exemple. Mais on ne va pas le faire ici.

Le Hamiltonien:

$$H(x^*(t), u, p(t), t) = \frac{1}{2}(x^*(t) - z(t))^2 + p(t)u.$$

L'équation adjointe:

$$\dot{p}(t) = -x^*(t) + z(t), \quad p(T) = 0.$$

Au temps t , on veut donc choisir u qui minimise $p(t)u$.

La solution sera

$$u^*(t) = \begin{cases} -a & \text{si } p(t) > 0; \\ a & \text{si } p(t) < 0; \\ \dot{z}(t) & \text{si } p(t) = 0; \end{cases}$$

Quand $p(t) = 0$, on veut le garder à 0 si possible, i.e., garder la pente de la route égale à la pente du terrain: $u^*(t) = \dot{x}^*(t) = \dot{z}(t)$.

On a donc des portions où la pente est de a ou $-a$, et entre ces portions, on a $p(t) = 0$. Donc pour chaque portion $[t_1, t_2]$ de pente a ou $-a$, on a $p(t_1) = p(t_2) = 0$, et donc

$$\int_{t_1}^{t_2} [z(t) - x^*(t)] dt = 0.$$