MAP434 - Contrôle de modèles dynamiques

Équation de Hamilton–Jacobi–Bellman (HJB)

Alexandre Ern

ENPC - INRIA - X

amphi du mardi 28 mai 2019

Plan de la séance

- Fonction valeur
- Principe d'optimalité de Bellman et équation HJB
- Application au système LQ
- Bilan : PMP vs. HJB
- Poly, chapitre 8

Système de contrôle non-linéaire

- On considère le système de contrôle non-linéaire

$$\dot{x}_u(t) = f(t, x_u(t), u(t)), \quad \forall t \in [0,T], \qquad x_u(0) = x_0$$

  - $x_u : [0,T] \to \mathbb{R}^d$, $T > 0$, CI $x_0 \in \mathbb{R}^d$
  - $u : [0,T] \to U \subset \mathbb{R}^k$, **sous-ensemble fermé non-vide**
  - $f : [0,T] \times \mathbb{R}^d \times U \to \mathbb{R}^d$

- On cherche un contrôle optimal $\overline{u} \in \mathcal{U} = L^1([0,T]; U)$ qui minimise le critère

$$J(0, x_0; u) = \int_0^T g(t, x_u(t), u(t))\,\mathrm{d}t + h(x_u(T))$$

avec $g : [0,T] \times \mathbb{R}^d \times U \to \mathbb{R}$ et $h : \mathbb{R}^d \to \mathbb{R}$

Fonction valeur

- On plonge ce problème dans une **famille** de problèmes de contrôle optimal posés sur $I_s = [s, T]$ avec la CI $x(s) = \xi$; la famille est ainsi paramétrée par $s \in [0, T]$ et $\xi \in \mathbb{R}^d$

$$\dot{x}_u(t) = f(t, x_u(t), u(t)),\ t \in I_s, \quad x_u(s) = \xi$$
$$J(s, \xi; u) = \int_{I_s} g(t, x_u(t), u(t))\,\mathrm{d}t + h(x_u(T)), \quad \forall u \in \mathcal{U}_s = L^1(I_s; U)$$

- La **fonction valeur** $V : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ est t.q.

$$V(s, \xi) = \inf_{u \in \mathcal{U}_s} J(s, \xi; u)$$

- En $s = T$, on a $V(T, \xi) = h(\xi)$, $\forall \xi \in \mathbb{R}^d$

Non-différentiabilité de la fonction valeur

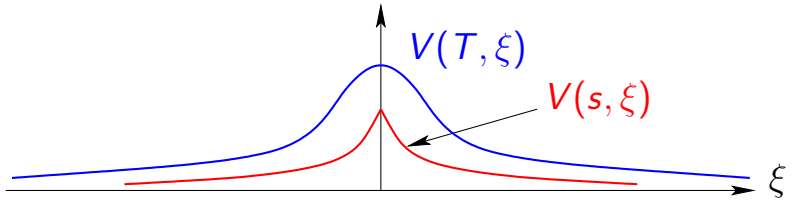
- La fonction valeur n'est pas toujours différentiable en tout point ...
- On considère le système de contrôle et le critère

$$\dot{x}_u(t) = u(t) \in U = [-1, 1],\ \forall t \in I_s, \quad x_u(s) = \xi, \quad J(s, \xi; u) = h(x_u(T))$$

où h est paire, régulière, t.q. $h'(x) < 0$ si $x > 0$ (e.g., $h(x) = e^{-x^2}$)

  - si $\xi > 0$, le contrôle optimal est $\overline{u} \equiv 1$ et $V(s, \xi) = h(\xi + T - s)$
  - si $\xi < 0$, le contrôle optimal est $\overline{u} \equiv -1$ et $V(s, \xi) = h(\xi - T + s)$
  - si $x = 0$, il y a deux contrôles optimaux $\overline{u} \equiv \pm 1$ et $V(s, 0) = h(\pm(T - s))$

- En conclusion, on a $V(s, \xi) = h(T - s + |\xi|)$, qui est régulière sauf en $\xi = 0$



- Cet exemple illustre un phénomène important et général : la perte de régularité de V en les points où **existent plusieurs contrôles optimaux**

Principe d'optimalité de Bellman

- **(Thm. 8.3)** Soit $(s,\xi) \in [0,T[\times \mathbb{R}^d$; pour tout $s' \in [s,T]$, on a

$$V(s,\xi) = \inf_{u \in \mathcal{U}_s^{s'}} \left\{ \int_{I_s^{s'}} g(t, x_u(t), u(t))\,\mathrm{d}t + V(s', x_u(s')) \right\}$$

avec $I_s^{s'} = [s,s']$ et $\mathcal{U}_s^{s'} = L^1(I_s^{s'}; U)$

- Preuve : tout $u \in \mathcal{U}_s$ peut être identifié à un couple $(u_1, u_2) \in \mathcal{U}_s^{s'} \times \mathcal{U}_{s'}$ en posant $u_1 = u|_{I_s^{s'}}$ et $u_2 = u|_{I_{s'}}$
  - u_1 produit la trajectoire $x_1 \equiv x_{u_1}$ sur $I_s^{s'}$ t.q. $x_1(s) = \xi$
  - u_2 produit la trajectoire $x_2 \equiv x_{u_2}$ sur $I_{s'}$ t.q. $x_2(s') = x_u(s') = x_1(s')$
  - on conclut grâce à l'**additivité du critère le long de la trajectoire**

$$\begin{aligned}
V(s,\xi) &= \inf_{u \in \mathcal{U}_s} \left\{ \int_{I_s} g(t, x_u(t), u(t))\,\mathrm{d}t + h(x_u(T)) \right\} \\
&= \inf_{(u_1,u_2) \in \mathcal{U}_s^{s'} \times \mathcal{U}_{s'}} \left\{ \int_{I_s^{s'}} g(t, x_1(t), u_1(t))\,\mathrm{d}t + \left\{ \int_{I_{s'}} g(t, x_2(t), u_2(t))\,\mathrm{d}t + h(x_2(T)) \right\} \right\} \\
&= \inf_{u_1 \in \mathcal{U}_s^{s'}} \left\{ \int_{I_s^{s'}} g(t, x_1(t), u_1(t))\,\mathrm{d}t + \inf_{u_2 \in \mathcal{U}_{s'}} \left\{ \int_{I_{s'}} g(t, x_2(t), u_2(t))\,\mathrm{d}t + h(x_2(T)) \right\} \right\} \\
&= \inf_{u_1 \in \mathcal{U}_s^{s'}} \left\{ \int_{I_s^{s'}} g(t, x_1(t), u_1(t))\,\mathrm{d}t + V(s', x_1(s')) \right\}
\end{aligned}$$

Hypothèses

- Les hypothèses sur la **dynamique** sont les suivantes
  - f continue sur $[0,T] \times \mathbb{R}^d \times U$, uniformément en $u \in U$
  - f dérivable p.r. à x et $\frac{\partial f}{\partial x}$ continue bornée sur $[0,T] \times \mathbb{R}^d \times U$
  - $\exists C \geq 0$ t.q. $|f(t,x,u)| \leq C(1+|x|+|u|)$ sur $[0,T] \times \mathbb{R}^d \times U$
- Les hypothèses sur le **critère** sont les suivantes
  - g continue sur $[0,T] \times \mathbb{R}^d$, uniformément en $u \in U$
  - $\exists \nu > 0$, $\exists C \geq 0$ t.q. $g(t,x,u) \geq \nu |u|^2 - C$ sur $[0,T] \times \mathbb{R}^d \times U$
  - h continue et minorée sur $\mathbb{R}^d$
- Ces hypothèses sont des CS pour que
  - $\forall u \in \mathcal{U} = L^1([0,T];U)$, $\exists!$ trajectoire $x_u \in AC([0,T];\mathbb{R}^d)$
  - J a bien un sens et $V(s,\xi) > -\infty$, $\forall (s,\xi) \in [0,T] \times \mathbb{R}^d$

Équation HJB (1)

- On rappelle que le Hamiltonien associé au système de contrôle non-linéaire est $H : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times U \to \mathbb{R}$ t.q.

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

- Le Hamiltonien minimisé est $H_\flat(t,x,p) = \inf_{u \in U} H(t,x,p,u)$
- L'équation HJB prend la forme suivante :

$$\frac{\partial V}{\partial s}(s,\xi) + H_\flat\big(s,\xi,\frac{\partial V}{\partial \xi}(s,\xi)\big) = 0$$

  avec la condition à l'instant final $V(T,\xi) = h(\xi)$

- Dans l'exemple précédent, $H(x,p,u) = pu$, $U = [-1,1]$
  - $H_\flat(x,p) = -|p|$
  - là où $V(s,\xi) = h(T - s - |\xi|)$ est régulière ($\xi \neq 0$), elle satisfait l'équation HJB $\frac{\partial V}{\partial s} - |\frac{\partial V}{\partial \xi}| = 0$ (le vérifier!)

Équation HJB (2)

- **(Thm. 8.4)** En tout point $(s, \xi) \in [0, T[\times \mathbb{R}^d$ où la fonction valeur est différentiable, elle satisfait l'équation HJB
  - la preuve repose sur le **principe d'optimalité de Bellman** : un contrôle $\overline{u}$ est optimal ssi à tout instant $s \in [0, T]$ sur la trajectoire associée $\overline{x}$, le contrôle restreint aux instants ultérieurs $\overline{u}|_{[s,T]}$ est optimal pour le nouveau problème ayant l'état courant $\overline{x}(s)$ comme état initial
- Si une fonction W régulière (C^0 sur $[0, T] \times \mathbb{R}^d$ et C^1 sur $[0, T[\times \mathbb{R}^d$) satisfait l'équation HJB et la condition en temps final et si U est borné, on a $W \equiv V$
  - ce résultat montre que l'équation HJB a **au plus une solution régulière**
  - il s'étend au cas U non borné sous hypothèse de décroissance de W quand $|\xi| \to +\infty$ uniformément en t

Preuve du Thm. 8.4 (1)

- On se restreint au cas où U est borné
- Soit $(s,\xi) \in [0,T[\times \mathbb{R}^d$ un point où la fonction valeur est différentiable
- On applique le principe d'optimalité ci-dessus avec $s' = s+\delta$, $0 < \delta < T-s$

$$V(s,\xi) = \inf_{u \in \mathcal{U}_s^{s+\delta}} \left\{ \int_{I_s^{s+\delta}} g(t, x_u(t), u(t))\,\mathrm{d}t + V(s+\delta, x_u(s+\delta)) \right\}$$

- Puisque U est borné, les hypothèses sur f impliquent que
  - $\sup_{t \in I_s^{s+\delta}} |x_u(t) - \xi| \leq C\delta$, uniformément en δ et en u
  - on en déduit que $x_u(s+\delta) = \xi + \int_{I_s^{s+\delta}} f(s,\xi,u(t))\,\mathrm{d}t + o(\delta)$ unif. en u
- Comme V est supposée différentiable en (s,ξ), on obtient

$$V(s+\delta, x_u(s+\delta)) = V(s,\xi) + \delta \left\{ \frac{\partial V}{\partial s}(s,\xi) + \frac{\partial V}{\partial \xi}(s,\xi)^\dagger \left(\frac{1}{\delta} \int_{I_s^{s+\delta}} f(s,\xi,u(t))\,\mathrm{d}t \right) \right\} + o(\delta)$$

- Les hypothèses sur g impliquent que

$$\int_{I_s^{s+\delta}} g(t, x_u(t), u(t))\,\mathrm{d}t = \int_{I_s^{s+\delta}} g(s,\xi,u(t))\,\mathrm{d}t + o(\delta)$$

Preuve du Thm. 8.4 (2)

- En reportant dans le principe d'optimalité, en simplifiant par $V(s,\xi)$ et en divisant par δ, et comme les termes en $o(\delta)$ sont uniformes en u, il vient

$$\frac{\partial V}{\partial s}(s,\xi)+\inf_{u\in\mathcal{U}_s^{s+\delta}}\left\{\frac{1}{\delta}\int_{I_s^{s+\delta}} g(s,\xi,u(t))\,\mathrm{d}t+\frac{\partial V}{\partial \xi}(s,\xi)^\dagger\left(\frac{1}{\delta}\int_{I_s^{s+\delta}} f(s,\xi,u(t))\,\mathrm{d}t\right)\right\}=o(1)$$

- On conclut en remarquant que (s et ξ sont fixés)

$$\begin{aligned}&\inf_{u\in\mathcal{U}_s^{s+\delta}}\left\{\frac{1}{\delta}\int_{I_s^{s+\delta}} g(s,\xi,u(t))\,\mathrm{d}t+\frac{\partial V}{\partial \xi}(s,\xi)^\dagger\left(\frac{1}{\delta}\int_{I_s^{s+\delta}} f(s,\xi,u(t))\,\mathrm{d}t\right)\right\}\\ &\quad=\inf_{v\in U}\left\{g(s,\xi,v)+\frac{\partial V}{\partial \xi}(s,\xi)^\dagger f(s,\xi,v)\right\}=\inf_{v\in U} H\left(s,\xi,\frac{\partial V}{\partial \xi}(s,\xi),v\right)\end{aligned}$$

  - on a $\inf_{u\in\mathcal{U}_s^{s+\delta}} \frac{1}{\delta}\int_{I_s^{s+\delta}} \Phi(u(t))\,\mathrm{d}t = \inf_{v\in U}\Phi(v)$ où $\Phi:\mathbb{R}^k\to\mathbb{R}$
  - on note I_1 le premier infimum et I_2 le deuxième
  - $\forall u\in\mathcal{U}_s^{s+\delta}$, on a $\Phi(u(t))\geq \inf_{v\in U}\Phi(v)$, $\forall t\in I_s^{s+\delta}$, d'où $I_1\geq I_2$
  - $\forall v\in U$, en considérant la fonction constante égale à v, on obtient $\Phi(v)\geq \inf_{u\in\mathcal{U}_s^{s+\delta}} \frac{1}{\delta}\int_{I_s^{s+\delta}} \Phi(u(t))\,\mathrm{d}t$, d'où $I_2\geq I_1$

Synthèse d'un feedback optimal

- On suppose que la fonction valeur est régulière $(C^0([0,T]\times\mathbb{R}^d)\cap C^1([0,T[\times\mathbb{R}^d))$
- On suppose que $\forall(s,\xi)\in[0,T]\times\mathbb{R}^d$, on peut trouver un feedback optimal

$$\tilde{u}(s,\xi)\in\arg\min_{v\in U} H(s,\xi,\tfrac{\partial V}{\partial\xi}(s,\xi),v)$$

  - l'existence est assurée par les hypothèses sur f et g
  - en général, on n'a pas unicité, ni dépendance continue en (s,ξ)
  - on suppose que l'on peut choisir $\tilde{u}(s,\xi)$ pour résoudre le système différentiel

$$\tfrac{d}{dt}\overline{x}(t)=f(t,\overline{x}(t),\tilde{u}(t,\overline{x}(t))),\ t\in[0,T],\quad \overline{x}(0)=x_0$$

- **(Prop. 8.7)** $\overline{u}(t)=\tilde{u}(t,\overline{x}(t))$ est un contrôle optimal sur $[0,T]$
  - on a $\frac{d}{dt}V(t,\overline{x}(t))=\frac{\partial V}{\partial s}(t,\overline{x}(t))+\frac{\partial V}{\partial\xi}(t,\overline{x}(t))^\dagger f(t,\overline{x}(t),\overline{u}(t))$, p.p. $t\in[0,T]$
  - comme V satisfait l'équation HJB, $\frac{\partial V}{\partial s}(t,\overline{x}(t))+H_\flat(t,\overline{x}(t),\frac{\partial V}{\partial\xi}(t,\overline{x}(t)))=0$
  - par définition de $\overline{u}$, $H_\flat(t,\overline{x}(t),\frac{\partial V}{\partial\xi}(t,\overline{x}(t)))=H(t,\overline{x}(t),\frac{\partial V}{\partial\xi}(t,\overline{x}(t)),\overline{u}(t))$
  - comme $H(t,x,p,u)=p^\dagger f(t,x,u)+g(t,x,u)$, il vient

$$\frac{d}{dt}V(t,\overline{x}(t))=-g(t,\overline{x}(t),\overline{u}(t)),\quad \text{p.p. } t\in[0,T]$$

  d'où $V(0,x_0)=\int_0^T g(t,\overline{x}(t),\overline{u}(t))\,\mathrm{d}t+h(\overline{x}(T))=J(0,x_0;\overline{u})$, i.e., $\overline{u}$ est bien un contrôle optimal

Fonction valeur et état adjoint

- On suppose qu'il existe un contrôle optimal $\overline{u} : [0,T] \to U$
  - on note $\overline{x} : [0,T] \to \mathbb{R}^d$ la trajectoire correspondante et $\overline{p} : [0,T] \to \mathbb{R}^d$ l'état adjoint introduit dans le PMP
  - on suppose que V est différentiable en tout $(s, \overline{x}(s))$, $\forall s \in [0,T]$
- Dans ces conditions, on a $\overline{p}(s) = \frac{\partial V}{\partial \xi}(s, \overline{x}(s))$, $\forall s \in [0,T]$
  - on a $V(s, \overline{x}(s)) = J(s, \overline{x}(s); \overline{u}|_{I_s})$ (principe d'optimalité de Bellman)
  - pour tout $\xi \in \mathbb{R}^d$, $V(s,\xi) = \inf_{u \in \mathcal{U}_s} J(s,\xi;u) \leq J(s,\xi;\overline{u}|_{I_s})$
  - la fonction $\xi \mapsto V(s,\xi) - J(s,\xi,\overline{u}|_{I_s})$ est donc maximale en $\xi = \overline{x}(s)$; par suite, $\frac{\partial}{\partial \xi_i} V(s, \overline{x}(s)) = \frac{\partial}{\partial \xi_i} J(s, \overline{x}(s), \overline{u}|_{I_s})$, $\forall i \in \{1{:}d\}$
  - on vérifie aisément que

$$\frac{\partial}{\partial \xi_i} J(s, \overline{x}(s), \overline{u}|_{I_s}) = \int_{I_s} \frac{\partial g}{\partial x}(t, \overline{x}(t), \overline{u}(t))^\dagger y_i(t) \, \mathrm{d}t + \frac{\partial h}{\partial x}(\overline{x}(T))^\dagger y_i(T)$$

    où $\dot{y}_i(t) = \frac{\partial f}{\partial x}(t, \overline{x}(t), \overline{u}(t)) y_i(t)$, $\forall t \in I_s$, $y_i(s) = e_i = (\delta_{ij})_{j \in \{1:d\}}$
  - en introduisant l'état adjoint t.q. $\frac{d}{dt}\overline{p} = -\frac{\partial f}{\partial x}^\dagger \overline{p} - \frac{\partial g}{\partial x}$, $\overline{p}(T) = \frac{\partial h}{\partial x}(\overline{x}(T))$, et en intégrant par parties en temps, il vient

$$\frac{\partial}{\partial \xi_i} J(s, \overline{x}(s), \overline{u}|_{I_s}) = -\int_{I_s} \frac{d}{dt}\left(\overline{p}(t)^\dagger y_i(t)\right) \mathrm{d}t + \overline{p}(T)^\dagger y_i(T) = \overline{p}(s)^\dagger e_i$$

Application au système LQ

- On considère la famille de systèmes LQ

$$\dot{x}_u(t) = Ax_u(t) + Bu(t),\ \forall t \in I_s, \quad x_u(s) = \xi, \quad u \in L^2(I_s;\mathbb{R}^k)$$
$$J(s,\xi;u) = \int_{I_s} \left\{\frac{1}{2}u(t)^\dagger Ru(t) + \frac{1}{2}x_u(t)^\dagger Qx_u(t)\right\}\mathrm{d}t + \frac{1}{2}x_u(T)^\dagger Dx_u(T)$$

- Le Hamiltonien est $H(x,p,u) = p^\dagger(Ax+Bu) + \frac{1}{2}u^\dagger Ru + \frac{1}{2}x^\dagger Qx$ et le Hamiltonien minimisé est

$$H_\flat(x,p) = \min_{u\in\mathbb{R}^k} H(x,p,u) = p^\dagger Ax - \frac{1}{2}p^\dagger BR^{-1}B^\dagger p + \frac{1}{2}x^\dagger Qx$$

l'unique minimiseur étant $\tilde{u} = -R^{-1}B^\dagger p$

- La fonction valeur satisfait $V(T,\xi) = \frac{1}{2}\xi^\dagger D\xi$ et l'équation HJB

$$\frac{\partial V}{\partial s} + \left(\frac{\partial V}{\partial \xi}\right)^\dagger A\xi - \frac{1}{2}\left(\frac{\partial V}{\partial \xi}\right)^\dagger BR^{-1}B^\dagger\frac{\partial V}{\partial \xi} + \frac{1}{2}\xi^\dagger Q\xi = 0$$

On vérifie que $V(s,\xi) = \frac{1}{2}\xi^\dagger P(s)\xi$ (où P vérifie l'équation de Riccati) est bien solution

  - V est ici **régulière** sur $[0,T]\times\mathbb{R}^d$; on notera l'**unicité** du contrôle optimal

Exemple : planification du mouvement

- Mouvement d'un point matériel avec critère quadratique

$$\dot{x}_u(t) = u(t), \quad J(s,\xi;u) = \int_s^T \frac{1}{2}(u(t)^2 + x_u(t)^2)\,\mathrm{d}t + \frac{1}{2}x_u(T)^2$$

- Hamiltonien $H(x,p,u) = pu + \frac{1}{2}(x^2+u^2)$, d'où $H_\flat(x,p) = \frac{1}{2}(x^2 - p^2)$ avec $\tilde{u} = -p$; on obtient l'équation HJB

$$\frac{\partial V}{\partial s} + \frac{1}{2}\left(\xi^2 - \left(\frac{\partial V}{\partial \xi}\right)^2\right) = 0, \qquad V(T,\xi) = \frac{1}{2}\xi^2$$

  - on cherche une solution de la forme $V(s,\xi) = \frac{1}{2}\mu(s)\xi^2$
  - il vient $\mu'(s) = \mu(s)^2 - 1$, $\mu(T) = 1$, d'où $\mu \equiv 1$
  - le contrôle optimal (sous forme de feedback) est

$$\tilde{u}(s,\xi) = -\frac{\partial V}{\partial \xi}(s,\xi) = -\mu(s)\xi = -\xi$$

si bien que $\frac{d}{dt}\overline{x}(t) = -\overline{x}(t)$; d'où $\overline{x}(t) = x_0 e^{-t}$ et $\overline{u}(t) = -x_0 e^{-t}$

Bilan : PMP vs. HJB

- Le **PMP**
  - fournit une CN d'optimalité
  - fournit le contrôle optimal en boucle ouverte (fonction du temps)
  - repose sur la résolution d'équations différentielles ordinaires
  - ne s'applique (sauf rares exceptions) qu'aux systèmes déterministes
- La **programmation dynamique** via la résolution de l'équation HJB
  - fournit une CS d'optimalité
  - fournit le contrôle optimal en boucle fermée (fonction de l'état)
  - repose sur la résolution d'une équation aux dérivées partielles (ce qui devient intractable lorsque la dimension d de l'espace des états est grande ...)
  - s'applique aux systèmes déterministes et stochastiques