Chapitre 5: état partiellement observé

Fabian Bastin

DIRO, Université de Montréal

IFT-6521 – Hiver 2013

Information partielle et conversion au modèle avec information parfaite

Dans plusieurs situations pratiques, l'état du système n'est pas observable complètement, de sorte qu'une politique admissible ne peut pas être n'importe quelle fonction de l'état. Les décisions ne doivent dépendre que de ce qui est observable.

Comment traiter cette situation? La réponse courte: simplement remplacer (redéfinir) l'état (partiellement observable) par l'information disponible, ou encore par une fonction de l'information disponible qui nous donne autant d'information utile mais sous une forme plus agrégée (une statistique exhaustive). On se ramène alors au cadre connu, mais avec un état défini différemment.
L'aggrégation des états (via une statistique exhaustive ou une autre méthode d'approximation) est souvent essentielle pour limiter la dimension de l'espace d'états, pour pouvoir résoudre.

C'est essentiellement ce que raconte la chapitre 5 du livre, avec quelques détails en plus.

Modèle

à l'étape k, le système est dans l'état x_k, mais on ne peut observer que

$$z_k = h_k(x_k, u_{k-1}, v_k),$$

où $z_k \in Z_k$ et $v_k \in V_k$ est une v.a. dont la loi

$$\mathbb{P}[v_k \in \cdot \mid x_k, \ldots, x_0, u_{k-1}, \ldots, u_0, w_{k-1}, \ldots, w_0, v_{k-1}, \ldots, v_0]$$

dépend de la suite des états, décisions, et aléas précédents.

L'état initial peut aussi être aléatoire, de loi $\mathbb{P}[x_0 \in \cdot]$.

L'information disponible à l'étape k est

$$I_k = (z_0, z_1, \ldots, z_k, u_0, u_1, \ldots, u_{k-1}), \quad k = 0, 1, \ldots, N-1,$$

et la décision décision $u_k \in U_k$ ne peut dépendre que de cette information. On suppose ici que U_k ne dépend pas de x_k. Ensuite une variable aléatoire w_k est "générée" selon une loi $P_k(\cdot \mid x_k, u_k)$, on doit payer un coût $g_k(x_k, u_k, w_k)$, et l'état à la prochaine étape est $x_{k+1} = f_k(x_k, u_k, w_k)$.

On cherche une politique admissible de la forme $\pi = (\mu_0, \ldots, \mu_{N-1})$, où $\mu_k(I_k) \in U_k$, qui minimise

$$\mathbb{E}\left[g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k, w_k) \right],$$

sous les contraintes

$$\begin{aligned} x_{k+1} &= f_k(x_k, \mu_k(I_k), w_k), \\ z_0 &= h_0(x_0, v_0), \\ z_k &= h_k(x_k, \mu_{k-1}(I_{k-1}), v_k), \quad k = 1, \ldots, N-1. \end{aligned}$$

Reformulation

Si on remplace l'état par I_k et la fonction de coût par étape par

$$\tilde{g}_k(I_k, u_k) = \mathbb{E}[g_k(x_k, u_k, w_k) \mid I_k, u_k],$$

on se retrouve dans cadre "standard" où l'état est complètement observé.

L'équation de récurrence se réecrit alors comme

$$\begin{aligned} J_k(I_k) &= \text{coût espéré total optimal de l'étape } k \text{ à la fin,} \\ &\quad \text{si l'information disponible à l'étape } k \text{ est } I_k \\ &= \min_{u_k \in U_k} [\tilde{g}_k(I_k, u_k) + \mathbb{E}[J_{k+1}(I_{k+1})]] \end{aligned}$$

où $z_{k+1} = h_{k+1}(x_{k+1}, u_k, v_{k+1})$ et $I_{k+1} = (I_k, z_{k+1}, u_k)$.

DPOC traite en détail le cas des systèmes linéaires à coût quadratique, puis examine plusieurs examples.

Statistique exhaustive

Une statistique exhaustive est une fonction S_k qui associe à chaque I_k une valeur $S_k = S_k(I_k)$, souvent plus compacte, telle que l'on peut réécrire

$$J_k(I_k) = \min_{u_k \in U_k} H_k(S_k(I_k), u_k)$$

pour une certaine fonction H_k. En d'autres mots, on peut écrire J_k et une politique optimale comme fonctions de $S_k = S_k(I_k)$ au lieu de I_k.

Dans ce cas, on peut remplacer l'état I_k par S_k.

Loi conditionnelle de l'état x_k.

Dans le cas fréquent où

$$\begin{aligned} & \mathbb{P}[v_k \in \cdot \mid x_k, \ldots, x_0, u_{k-1}, \ldots, u_0, w_{k-1}, \ldots, w_0, v_{k-1}, \ldots, v_0] \\ = \quad & \mathbb{P}[v_k \in \cdot \mid x_k, x_{k-1}, u_{k-1}, w_{k-1}], \end{aligned}$$

on peut prendre $S_k = S_k(I_k) = \mathbb{P}[x_k \in \cdot \mid I_k]$, la loi de probabilité de x_k conditionnelle à l'information connue I_k.

On peut mettre à jour

$$S_{k+1} = \Phi_k(\mathbb{P}[x_k \in \cdot \mid I_k], u_k, z_{k+1}) = \Phi_k(S_k, u_k, z_{k+1})$$

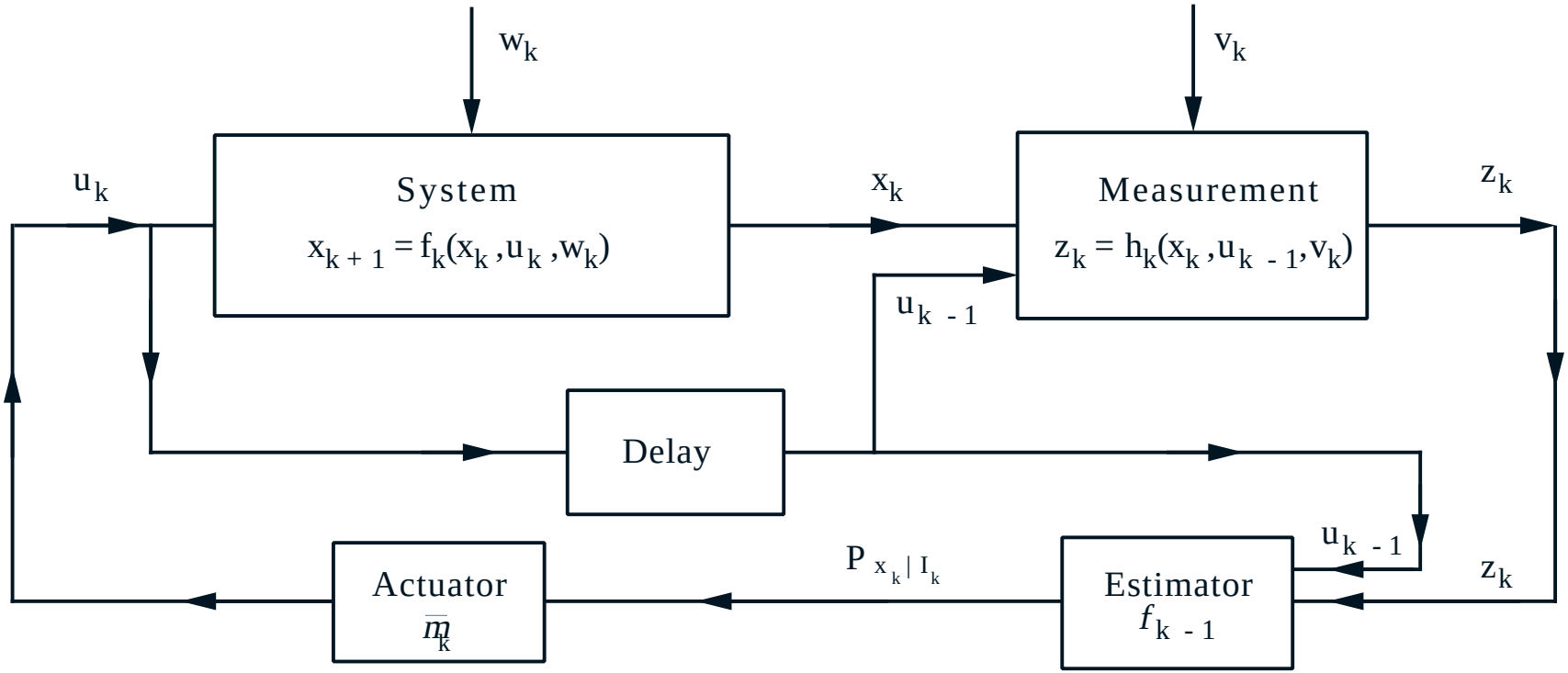
pour une certaine fonction Φ_k. Le coût par étape est remplacée par

$$\tilde{g}_k(S_k, u_k) = \mathbb{E}[g_k(x_k, u_k, w_k) \mid S_k, u_k] = \int g_k(x_k, u_k, w_k) d\mathbb{P}[x_k, w_k \mid S_k, u_k].$$

et on peut alors écrire

$$J_k(S_k) = \min_{u_k \in U_k} \left[\tilde{g}_k(S_k, u_k) + \mathbb{E}[J_{k+1}(S_{k+1})]\right].$$

La commande optimale se décompose alors en deux parties: (a) estimation de la loi conditionnelle de l'état; (b) choix de la décision. En pratique, de nombreuses heuristiques (sous-optimales) sont basées sur des versions approximatives de ce schéma.



Exemple: prises de décision sous un modèle Bayesien

Une loterie bien particulière vend des billets C dollars. On pense que chaque billet permet de gagner V dollars avec probabilité $\beta > 0$ (cas A) mais il est aussi possible que la probabilité de gagner soit de zéro à tous les tirages (cas B). Notre probabilité a priori que l'on soit dans le cas A est $p_0 > 0$. Dès que l'on a gagné une fois, on ne peut plus jouer.

Soit p_k la probabilité que l'on soit dans le cas A après avoir acheté k billets sans gagner. De par la formule de Bayes, nous avons

$$\begin{aligned} p_k &= \mathbb{P}[\text{cas A} \mid k \text{ échecs}] = \frac{\mathbb{P}[\text{cas A et } k \text{ échecs}]}{\mathbb{P}[k \text{ échecs}]} \\ &= \frac{\mathbb{P}[k \text{ échecs} \mid \text{cas A}]p_0}{\mathbb{P}[k \text{ échecs} \mid \text{cas A}]p_0 + \mathbb{P}[k \text{ échecs} \mid \text{cas B}](1-p_0)} \\ &= \frac{(1-\beta)^k p_0}{(1-\beta)^k p_0 + 1 - p_0} \end{aligned}$$

Probabilité de jeu non truqué

Nous avons

$$\begin{aligned}
\frac{(1-\beta)^k p_0}{(1-\beta)^k p_0 + 1 - p_0} &= \frac{(1-\beta)\frac{(1-\beta)^{k-1}p_0}{(1-\beta)^{k-1}p_0+1-p_0}}{(1-\beta)\frac{(1-\beta)^{k-1}p_0}{(1-\beta)^{k-1}p_0+1-p_0} + \frac{1-p_0}{(1-\beta)^{k-1}p_0+1-p_0}} \\
&= \frac{(1-\beta)p_{k-1}}{(1-\beta)p_{k-1} + 1 - \frac{(1-\beta)^{k-1}p_0}{(1-\beta)^{k-1}p_0+1-p_0}} \\
&= \frac{(1-\beta)p_{k-1}}{(1-\beta)p_{k-1} + 1 - p_{k-1}}
\end{aligned}$$

On voit que p_k est décroissant en k et $p_k \to 0$ quand $k \to \infty$.

Comme état à l'étape k, on peut prendre p_k (statistique exhaustive), ou même simplement k, puisque l'état n'est utile que lorsqu'on n'a pas encore gagné. Si J_k est le gain espéré optimal après k échecs, alors on a

$$J_k = \max\{0,\ p_k\beta V - C + (1 - p_k\beta)J_{k+1}]\}$$

pour $k = 0, 1, 2, \ldots$.

Nous avons comme contrainte $J_k \leq V$, vu que l'on ne peut gagner plus que V dollars. On montre alors facilement $J_k \geq J_{k+1}$.

Par récurrence sur k.

Le coût terminal est nul: $J_N = 0$. Par conséquent,

$$J_{N-1} \geq 0 = J_N$$

Supposons que le résultat tient pour J_{k+1}. Nous avons

$$\begin{aligned} J_k &= \max\{0,\ p_k\beta V - C + (1 - p_k\beta)J_{k+1}\} \\ &\geq \max\{0,\ p_k\beta V - C + (1 - p_k\beta)J_{k+2}\} \\ &= \max\{0,\ p_k\beta[V - J_{k+2}] - C + J_{k+2}\} \\ &\geq \max\{0,\ p_{k+1}\beta[V - J_{k+2}] - C + J_{k+2}\} \\ &= \max\{0,\ p_{k+1}\beta V - C + (1 - p_{k+1})J_{k+2}\} \\ &= J_{k+1} \end{aligned}$$

Puisque pour tout k, $J_k \geq 0$, il est clair que $J_k > 0$ dès que $p_k\beta V - C > 0$. Si au contraire, $p_k\beta V - C \leq 0$, nous avons $J_k = 0$. En effet,

$$J_k = p_k\beta V - C + (1 - p_k\beta)J_{k+1} \leq p_k\beta V - C + (1 - p_k\beta)J_k$$

et donc

$$0 \leq p_k\beta J_k \leq p_k\beta V - C \leq 0.$$

La politique optimale est donc d'acheter au maximum k^* billets, où

$$k^* = \max\{k : p_k\beta V > C\}.$$