Le problème du vendeur de journaux en horizon fini

J.-P. Chancelier, B. Lapeyre

http://cermics.enpc.fr/~jpc/decision-incertain

Plan

Rappel sur le problème à un pas de temps

Le vendeur de journaux : problème dynamique

$$\min_{u \in \mathcal{U}, \mathbf{X_1}, \mathbf{X_0}} \mathbb{E}_{\mathrm{W}}\left[\tilde{\jmath}(u, \mathrm{X}_1)\right]$$

sous contraintes

$$\mathrm{X}_0 = x \quad \mathrm{X}_1 = f(\mathrm{X}_0, u, \mathrm{W})$$

$$f(x,u,w) = x + u - w$$
$$\tilde{\jmath}(u,x) = c_F \mathbb{I}_{\{u>0\}} + cu + \alpha\left(c_S(x)_+ + c_M(-x)_+\right)$$

- Le « stock » $\mathrm{X}_t \in \mathbb{Z}$ peut être positif (c'est alors un stock physique) ou négatif (c'est alors moins la quantité de manquants)
- La loi de la demande W, est supposée connue (de moyenne finie)
- Taux d'actualisation $\alpha \in (0, 1]$ (valoriser des gains futurs à l'instant présent)

Le problème à horizon fini T

Le problème sur un horizon fini

$$\min_{\mathsf{U}\in\mathcal{U},\mathsf{X}} \mathbb{E}_{\mathsf{W}}\Big[\sum_{t=0}^{T-1} \alpha^t \tilde{\jmath}(\mathsf{U}_t, \mathsf{X}_{t+1})\Big]$$

sous contraintes

$$\mathsf{X}_0 = x \quad \text{et} \quad \forall t \in [0, T-1]\,,\ \mathsf{X}_{t+1} = f(\mathsf{X}_t, \mathsf{U}_t, \mathsf{W}_{t+1})$$

- La loi de la demande W_t, est supposée connue (de moyenne finie) pour tout t.
- On sera amené à supposer que les demandes $(\mathsf{W}_1, \mathsf{W}_2, \ldots, \mathsf{W}_T)$ sont indépendantes.
- $\alpha \in (0, 1]$ est un taux d'actualisation (valoriser des gains futurs à l'instant présent).

Le problème à horizon T sous une forme canonique

Le problème sur un horizon fini

$$\min_{\mathsf{U} \in \mathcal{U}, \mathsf{X}} \mathbb{E}_{\mathsf{W}} \Big[\sum_{t=0}^{T-1} \alpha^t c_t(\mathsf{U}_t, \mathsf{X}_t) + \alpha^T c_T(\mathsf{X}_T) \Big]$$

sous contraintes

$$\mathsf{X}_0 = x \quad \text{et} \quad \forall t \in [0, T-1]\,,\ \mathsf{X}_{t+1} = f(\mathsf{X}_t, \mathsf{U}_t, \mathsf{W}_{t+1})$$

avec

$$\begin{aligned} c_t(u,x) &= c_F \mathbb{I}_{\{u>0\}} + cu + c_S(x)_+ + c_M(-x)_+ \\ c_0(u,x) &= c_F \mathbb{I}_{\{u>0\}} + cu \\ c_T(x) &= c_S(x)_+ + c_M(-x)_+ \end{aligned}$$

Ré-écriture du coût

$$
\begin{aligned}
\tilde{\jmath}(u_0, x_1) + \alpha \tilde{\jmath}(u_1, x_2) &= \overbrace{c_F \mathbb{I}_{\{u_0 > 0\}} + c u_0 + \alpha \left(c_S (x_1)_+ + c_M (-x_1)_+ \right)}^{\tilde{\jmath}(u_0, x_1)} \\
&+ \alpha \overbrace{\left(c_F \mathbb{I}_{\{u_1 > 0\}} + c u_1 + \alpha \left(c_S (x_2)_+ + c_M (-x_2)_+ \right) \right)}^{\tilde{\jmath}(u_1, x_2)} \\
&= \overbrace{c_F \mathbb{I}_{\{u_0 > 0\}} + c u_0}^{c_0(u_0, x_0)} \\
&+ \alpha \overbrace{\left(c_F \mathbb{I}_{\{u_1 > 0\}} + c u_1 + c_S (x_1)_+ + c_M (-x_1)_+ \right)}^{c_1(u_1, x_1)} \\
&+ \alpha^2 \overbrace{\left(c_S (x_2)_+ + c_M (-x_2)_+ \right)}^{c_T(x_2)} \\
&= c_0(u_0, x_0) + \alpha c_1(u_1, x_1) + \alpha^2 c_T(x_2)
\end{aligned}
$$

La contrainte de non anticipativité

- On suppose que le vendeur de journaux note chaque jour la valeur de la demande qui s'est réalisée et conserve cette valeur.
  - À l'instant t, il connaît $(W_1, \cdots, W_t)$ et X_0 et peut utiliser cette information pour calculer sa commande de journaux U_t
  - On pourrait donc choisir pour $\mathcal{U}$ les commandes de journaux qui à l'instant t sont fonction de $(X_0, W_1, \cdots, W_t)$

$$\mathcal{U}_t = \{\phi : \mathbb{X} \times \mathbb{W}_1 \times \cdots \times \mathbb{W}_t \to \mathbb{N}\}$$

- Si les bruits sont indépendants et indépendants de X_0, on peut montrer que la stratégie optimale pour des fonctions de $(X_0, W_1, \cdots, W_t)$ ne dépend à l'instant t que du stock X_t. La forme de la fonction coût à son importance dans ce résultat.

$$\mathcal{U}_t = \{\phi : \mathbb{X}_t \to \mathbb{N}\}$$

- Noter que, dans le cas des chaînes de Markov données par leur matrices de transition le, « bruit » n'est pas observé.
- On parle d'observation complète si on observe l'état de la chaîne de Markov.

Equation de Programmation Dynamique (Bellman)

On pose pour simplifier $L_t(x, u, w) = \alpha^t c_t(x, u)$ et $K = \alpha^T c_T$

Définition de $V_0(x)$

$$
\begin{aligned}
V_0(x) = \min_{\mathrm{X},\mathrm{U}\in\mathcal{U}} \quad & \mathbb{E}\Big[\sum_{s=0}^{T-1} L_s\left(\mathrm{X}_s, \mathrm{U}_s, \mathrm{W}_{s+1}\right) + K\left(\mathrm{X}_T\right)\Big], \\
\text{s.c.} \quad & \mathrm{X}_0 = x \text{ fixé}, \\
& \mathrm{X}_{s+1} = f_s\left(\mathrm{X}_s, \mathrm{U}_s, \mathrm{W}_{s+1}\right), \qquad \forall s = 0, \ldots, T-1,
\end{aligned}
$$

Pour obtenir $V_0(x)$ on calcule la famille de fonctions valeurs $\{V_t\}_{t\in[0,T]}$

$$
\begin{aligned}
V_T &= K \\
V_t(x) &= \min_{u\in\mathbb{U}} \mathbb{E}\Big[L_t(x, u, \mathrm{W}_{t+1}) + V_{t+1}\big(f_t(x, u, \mathrm{W}_{t+1})\big)\Big]
\end{aligned}
$$

Commande optimale

- Pour $t \in [0, T-1]$

$$V_t(x) = \min_{u \in \mathbb{U}} \mathbb{E}\Big[L_t(x, u, \mathrm{W}_{t+1}) + V_{t+1}\big(f_t(x, u, \mathrm{W}_{t+1})\big)\Big]$$

$$\phi_t^\sharp(x) \in \arg\min_{u \in \mathbb{U}} \mathbb{E}\Big[L_t(x, u, \mathrm{W}_{t+1}) + V_{t+1}\big(f_t(x, u, \mathrm{W}_{t+1})\big)\Big]$$

- A t fixé, autant de problème d'optimisation qu'il y a de x possibles
- Pour chaque x possible, on trouve une commande $u^\sharp$ optimale
- On construit ainsi une stratégie $\phi_t^\sharp : \mathbb{X} \to \mathbb{U}$
- La commande optimale pour le problème demarrant en $t = 0$ avec $X_0 = x$ est pour $t \in [0, T-1]$, $\mathrm{U}_t^\sharp = \phi_t^\sharp(\mathrm{X}_t^\sharp)$ avec

$$\mathrm{X}_0^\sharp = x_0$$

$$\mathrm{X}_{s+1}^\sharp = f_s\left(\mathrm{X}_s^\sharp, \phi_t^\sharp(\mathrm{X}_t^\sharp), \mathrm{W}_{s+1}\right), \qquad \forall s = 0, \ldots, T-1,$$

Résumé

Sous une hypothèse d'indépendance des bruits $\{\mathrm{W}_t\}_{t \in [1,T]}$
l'équation de Programmation Dynamique $V_T = K$ et

$$V_t(x) = \min_{u \in \mathbb{U}} \mathbb{E}\Big[L_t(x,u,\mathrm{W}_{t+1}) + V_{t+1}\big(f_t(x,u,\mathrm{W}_{t+1})\big)\Big] \qquad \forall t \in [0, T-1]$$

donne la valeur du problème d'optimisation

$$\begin{aligned} V_0(x) = \min_{\mathrm{X},\mathrm{U} \in \mathcal{U}} \quad & \mathbb{E}\Big(\sum_{s=0}^{T-1} L_s\left(\mathrm{X}_s, \mathrm{U}_s, \mathrm{W}_{s+1}\right) + K\left(\mathrm{X}_T\right)\Big), \\ \text{s.c.} \quad & \mathrm{X}_0 = x \text{ fixé}, \\ & \mathrm{X}_{s+1} = f_s\left(\mathrm{X}_s, \mathrm{U}_s, \mathrm{W}_{s+1}\right), \qquad \forall s = 0, \ldots, T-1, \end{aligned}$$

et la commande optimale à chaque instant est une fonction de l'état

$$\phi_t^\sharp \in \mathcal{U}_t = \{\phi : \mathbb{X}_t \to \mathbb{N}\}$$

Difficultés de la programmation dynamique

- Difficultés mathématiques
  - Existence d'un minimum pour le problème initial
  - Existence de fonctions de Bellman au moins mesurables solution de l'Equation de Bellman et eventuellement plu srégulières.
  - Existence de stratégies optimales mesurable, problèmes de selections mesurables

D. P. Bertsekas and S. E. Shreve
Stochastic Optimal Control : The Discrete-Time Case (1996)

  - Le cas convexe s.c.i : Algorithme particulier SDDP qui utilise la convexité.
- Difficultés numériques, malédiction de la dimension. Si le cardinal de $\mathbb{X}$ est trop grand on ne sait plus résoudre
  - Tétris : espace d'état énorme 10^{61} pour hauteur 20 largeur 10 et 7 pièces.
  - Vallée hydraulique : problème quand on à plus de quatre stocks (barrages).
  - Gestion de portefeuille : problème quand on à plus de quatre actifs.

Autre Example : Processus en “dents de scie”

$(W_t, t \geq 1)$ une suite de tirages à pile ou face indépendants, $0 \leq p \leq 1$
$W_t = P$ avec probabilité p, $W_t = F$ avec probabilité $(1-p)$.

$$V_0(x) = \mathbb{E}\Big[\sum_{t=0}^{T-1} 1_{[5,\infty[}(X_t) + 1_{[5,\infty[}(X_T)\Big],$$

$$\text{avec} \quad X_0 = x \text{ fixé},$$

$$X_{t+1} = (1 + X_t) 1_{\{W_{t+1} = P\}}, \quad \forall t = 0, \ldots, T-1 .$$

Pour obtenir $V_0(x)$ on calcule la famille de fonctions valeurs $\{V_t\}_{t \in [0,T]}$

$$V_T(.) = 1_{[5,\infty[}(\cdot)$$

$$V_t(x) = \mathbb{E}\Big[1_{[5,\infty[}(x) + V_{t+1}\big((1+x)1_{[5,\infty[}(W_{t+1})\big)\Big] .$$

Soit aussi

$$V_T(.) = 1_{[5,\infty[}(\cdot)$$

$$V_t(x) = 1_{[5,\infty[}(x) + pV_{t+1}((1+x)) + (1-p)V_{t+1}(0) .$$

Autre Example : Gestion d'un stock d'eau dans un barrage

$$V_0(x) = \max_{U \in \mathcal{U}, X} \mathbb{E}\Big[\sum_{t=0}^{T-1} \underbrace{f(S_t, U_t)}_{\text{gain à turbiner } U_t} + \underbrace{K(X_T)}_{\text{valeur du stock à } T}\Big],$$

$$S_0 = x \text{ fixé},$$

$$S_{t+1} = \min\big(\max(0, S_t - U_t + W_{t+1}), \overline{S}\big), \ \forall t = \{0, \dots, T-1\},$$

- $(W_t)_{t \in \{1,\dots,T\}}$ est une suite de v.a indépendantes (apports d'eaux).
- Les contrôles sont non-anticipatif $U \in \mathcal{U}$ à valeur dans $\mathbb{U}$.

$V_0(x)$ est solution de l'equation de Bellman

$$V_T(.) = K(\cdot)$$

$$V_t(x) = \max_{u \in \mathbb{U}} \mathbb{E}\Big[f(x,u) + V_{t+1}\big(\min\big(\max(0, S_t - u + W_{t+1})\big)\big)\Big].$$

La commande optimale est donnée pour $t \in \{0, \dots, T-1\}$ par

$$U_t^\sharp(x) \in \max_{u \in u} \mathbb{E}\Big[f(x,u) + V_{t+1}\big(\min\big(\max(0, S_t - u + W_{t+1})\big)\big)\Big].$$

Preuve

On introduit $(\mathcal{P}_t)$ démarrant en x à l'instant t :

$$
\begin{aligned}
V_t(x) = \min_{\mathrm{X},\mathrm{U}\in\mathcal{U}} \quad & \mathbb{E}\Big[\sum_{s=t}^{T-1} L_s\left(\mathrm{X}_s, \mathrm{U}_s, \mathrm{W}_{s+1}\right) + K\left(\mathrm{X}_T\right)\Big], \\
\text{s.c.} \quad & \mathrm{X}_t = x \text{ fixé}, \\
& \mathrm{X}_{s+1} = f_s\left(\mathrm{X}_s, \mathrm{U}_s, \mathrm{W}_{s+1}\right), \qquad \forall s = t, \ldots, T-1,
\end{aligned}
$$

- V_t est appelé fonction valeur à l'instant t.
- V_T est connu $V_T = K$
- On cherche V_0 la valeur du problème du vendeur de journaux $(\mathcal{P}_0)$
- On cherche la stratégie optimale du vendeur de journaux

On commence par le cas où $L_s \equiv 0$

On commence par un problème avec juste un coût final

$(\mathcal{P}_t)$ démarrant en x à l'instant t :

$$
\begin{aligned}
V_t(x) = \min_{\mathrm{X},\mathrm{U}\in\mathcal{U}} \quad & \mathbb{E}(K(\mathrm{X}_T)),\\
\text{s.c.} \quad & \mathrm{X}_t = x \text{ fixé},\\
& \mathrm{X}_{s+1} = f_s(\mathrm{X}_s, \mathrm{U}_s, \mathrm{W}_{s+1}), \qquad \forall s = t, \dots, T-1,
\end{aligned}
$$

- Les bruits $\mathrm{W} = (\mathrm{W}_t)_{t=1,\dots,T}$.
- Les contrôles $\mathrm{U} = (\mathrm{U}_t)_{t=0,\dots,T-1}$.
- Les états $(\mathrm{X}_t)_{t=0,\dots,T-1}$.

On se limite aussi à

$$\mathcal{U}_t = \{\phi : \mathbb{X}_t \to \mathbb{N}\}$$

Dynamique Markovienne

Hypothèse Markovienne : bruits $\mathrm{X}_0, \mathrm{W}_1, \dots, \mathrm{W}_T$ indépendants et les W_t ont même loi (homogénéité).

- Matrice de transition : dynamique non contrôlée $f : \mathbb{X} \times \mathbb{W} \to \mathbb{X}$

$$\mathrm{X}_{t+1} = f(\mathrm{X}_t, \mathrm{W}_{t+1}), \quad P(x,y) = \mathbb{P}(f(x, \mathrm{W}_1) = y)$$

- Matrice de transition : dynamique contrôlée $f : \mathbb{X} \times \mathbb{U} \times \mathbb{W} \to \mathbb{X}$ avec politique Markovienne $(\phi_s)_{s \in [0, T-1]}$, $\mathrm{U}_t = \phi_t(\mathrm{X}_t)$

$$\mathrm{X}_{t+1} = f(\mathrm{X}_t, \phi_t(\mathrm{X}_t), \mathrm{W}_{t+1}), \quad P_t(x,y) = \mathbb{P}(f(x, \phi_t(x), \mathrm{W}_1) = y)$$

Pour $u \in \mathbb{U}$, soit $P^u = \mathbb{P}(f(x, u, \mathrm{W}_1) = y)$. Pour une politique $(\phi_s)_{s \in [0, T-1]}$ Markovienne, P_t^{ϕ} définie par $P_t^{\phi}(x,y) = P^{\phi_t(x)}(x,y)$

$$\mathrm{X}_{t+1} = f(\mathrm{X}_t, \phi_t(\mathrm{X}_t), \mathrm{W}_{t+1}), \quad P_t^{\phi}(x,y) = P^{\phi_t(x)}(x,y)$$

Deux familles de problèmes

$(\mathcal{P}_t)$ démarrant en x à l'instant t

$$
\begin{aligned}
V_t(x) = \min_{X,\phi} \quad & V_t^{\phi}(x) \\
& V_t^{\phi}(x) = \mathbb{E}(K(X_T)), \\
& \qquad \text{s.c. } X_t = x \text{ fixé}, \\
& \qquad\qquad X_{s+1} = f_s\left(X_s, \phi_s(X_s), W_{s+1}\right), \ \forall s \in [t, T-1]
\end{aligned}
$$

$(\mathcal{P}'_t)$ démarrant en μ (loi) à l'instant t

$$
\begin{aligned}
\mathcal{V}_t(\mu) = \min_{\mu_t,\dots,\mu_T,\phi(\cdot)} \quad & \mathcal{V}_t^{\phi}(\mu) \\
& \mathcal{V}_t^{\phi}(\mu) = \sum_x \mu_T(x) K(x) \\
& \text{avec} \quad \mu_t = \mu, \quad \mu_{s+1} = \mu_s P_s^{\phi}, \ \forall s \in [t, T-1]
\end{aligned}
$$

Liens entre les fonctions valeurs $\mathcal{V}_t^{\phi}(\cdot)$ et $V_t^{\phi}(\cdot)$

$\mathcal{V}_t^{\phi}(\mu)$ est la moyenne de la fonction V_t^{ϕ} pour une loi μ sur $\mathbb{X}$

$$\mathcal{V}_t^{\phi}(\mu) = \left\langle \mu,\ V_t^{\phi} \right\rangle \text{ avec } \langle \mu,\ F \rangle = \sum_{x \in \mathbb{X}} \mu(x) F(x)$$

$V_t^{\phi}(x)$ est la valeur de $\mathcal{V}_t^{\phi}(\mu)$ quand $\mu = \delta_x(\cdot)$

$$V_t^{\phi}(x) = \mathcal{V}_t^{\phi}(\delta_x(\cdot))$$

$\rightarrow$ Les mêmes relations sont vraies entre $\mathcal{V}_t$ et v_t.

En effet on obtient

- Problème $\mathcal{P}_t$ démarrant en x à t :

$$V_t^\phi(x) = (P_t^\phi \cdots P_{T-1}^\phi K)(x)$$

$$\text{(Par exemple } V_{T-1}^\phi(x) = \sum_{y \in \mathbb{X}} P_{T-1}^\phi(x,y)K(y))$$

- Problème $\mathcal{P}'_t$ démarrant en μ à t :

$$\mathcal{V}_t^\phi(\mu) = \mu P_t^\phi \cdots P_{T-1}^\phi K$$

$$\text{(Par exemple } \mathcal{V}_{T-1}^\phi(\mu) = \sum_{x,y \in \mathbb{X}} \mu(x) P_{T-1}^\phi(x,y)K(y))$$

- Les deux formules s'en déduisent

Lien entre V et $\mathcal{V}$ sur un pb d'horizon $T = 1$

$V(x)$

$$V(x) = \mathbb{E}_{X_0=x}\left[K(X_1)\right] = \sum_y P_{x,y} K(y)$$

$\mathcal{V}(\mu)$

$$\mathcal{V}(\mu) = \mathbb{E}_\mu\left[K(X_1)\right] = \sum_x \mu(x) \sum_y P_{x,y} K(y)$$

On obtient

$$\mathcal{V}(\mu) = \sum_y \mu(x) V(x) = \langle \mu,\ V \rangle \ ,$$

$$\mathcal{V}(\delta_{x'}) = \sum_x \delta_{x'}(x) \sum_y P_{x,y} K(y) = \sum_y P_{x',y} K(y) = V(x') \, .$$

Formule Récursive pour $\mathcal{V}_t$

On a la relation suivante :

$$\mathcal{V}_t(\mu) = \min_{\phi_t} \mathcal{V}_{t+1}(\mu P_t^{\phi})$$

Preuve : Le problème $\mathcal{P}'_t$ démarrant en μ à t :

$$\mathcal{V}_t(\mu) = \min_{\phi(\cdot)} \quad \mu P_t^{\phi} \cdots P_{T-1}^{\phi} K = \min_{\phi(\cdot)} \left\langle \mu P_t^{\phi},\ P_{t+1}^{\phi} \cdots P_{T-1}^{\phi} K \right\rangle$$

- À l'instant t, P_t^{ϕ} ne dépend que de la fonction ϕ_t.
- Le vecteur ligne μP_t^{ϕ} est à éléments positifs (loi de probabilité)

$$\begin{aligned} \mathcal{V}_t(\mu) &= \min_{\phi_t} \left\langle \mu P_t^{\phi},\ \min_{(\phi_s)_{s>t}} P_{t+1}^{\phi} \cdots P_{T-1}^{\phi} K \right\rangle \\ &= \min_{\phi_t} \left\langle \mu P_t^{\phi},\ V_{t+1}(\cdot) \right\rangle = \min_{\phi_t} \mathcal{V}_{t+1}(\mu P_t^{\phi}) \end{aligned}$$

Formule Récursive pour V_t

Équation de Bellman

$$V_t(x) = \min_{u \in \mathbb{U}} \mathbb{E}\left[V_{t+1}\left(f_t(x, u, \mathrm{W}_{t+1})\right)\right]$$

$$u^\sharp(x) \in \arg\min_{u \in \mathbb{U}} \mathbb{E}\left[V_{t+1}\left(f_t(x, u, \mathrm{W}_{t+1})\right)\right]$$

$$V_T(x) = K(x)$$

Preuve : On a obtenu pour $\mathcal{V}_t$

$$\mathcal{V}_t(\mu) = \min_{\phi_t} \mathcal{V}_{t+1}(\mu P_t^\phi)$$

- $V_t(x) = \mathcal{V}_t(\delta_x(\cdot))$
- $V_{t+1}\left(\delta_x(\cdot) P_t^\phi\right) = \mathbb{E}\left[V_{t+1}(f_1(x, \phi_t, \mathrm{W}_{t+1}))\right]$

Fin de la preuve

Écriture avec la matrice de transition de la chaîne de Markov

Équation de Bellman

$$V_t(x) = \min_{u \in \mathbb{U}} \sum_y P^u_{x,y} V_{t+1}(y)$$

$$V_T(x) = K(x)$$

$$u^\sharp(x) \in \arg\min_{u \in \mathbb{U}} \sum_y P^u_{x,y} V_{t+1}(y)$$

Preuve :

$$\begin{aligned}\sum_y P^u_{x,y} V_{t+1}(y) &= \sum_y \mathbb{P}(f(x, u, W_{t+1}) = y) V_{t+1}(y) \\ &= \mathbb{E}\left[V_{t+1}(f_t(x, u, W_{t+1}))\right]\end{aligned}$$

Horizon fini avec coût instantané

Le problème $(\mathcal{P}_0)$ démarrant en x à l'instant initial $t = 0$:

$$
\begin{aligned}
V_0(x) = \min_{\mathrm{X},\mathrm{U} \in \mathcal{U}} \quad & \mathbb{E}\Big(\sum_{t=0}^{T-1} L_t\left(\mathrm{X}_t, \mathrm{U}_t, \mathrm{W}_{t+1}\right) + K\left(\mathrm{X}_T\right)\Big), \\
\text{s.c.} \quad & \mathrm{X}_0 = x \text{ fixé}, \\
& \mathrm{X}_{t+1} = f_t\left(\mathrm{X}_t, \mathrm{U}_t, \mathrm{W}_{t+1}\right), \qquad \forall t = 0, \ldots, T-1
\end{aligned}
$$

Équation de programmation dynamique

$$
V_t(x) = \min_{u \in \mathbb{U}} \mathbb{E}\left[L_t(x, u, \mathrm{W}_{t+1}) + V_{t+1}(f_t(x, u, \mathrm{W}_{t+1}))\right]
$$

Preuve

Le problème $(\mathcal{P}_0)$ démarrant en x à $t = 0$ à un coût $\tilde{V}(0, x)$ où :

$$
\begin{aligned}
\tilde{V}_0(z, x) = \min_{\mathrm{X},\mathrm{U} \in \mathcal{U}} \quad & \mathbb{E}(Z_T + K(\mathrm{X}_T)), \\
\text{s.c.} \quad & \mathrm{X}_0 = x, \mathrm{Z}_0 = z, \text{ fixées}, \\
& \mathrm{X}_{t+1} = f_t(\mathrm{X}_t, \mathrm{U}_t, \mathrm{W}_{t+1}), \\
& \mathrm{Z}_{t+1} = \mathrm{Z}_t + L_t(\mathrm{X}_t, \mathrm{U}_t, \mathrm{W}_{t+1}), \qquad \forall t = 0, \dots, T-1,
\end{aligned}
$$

Le couple (Z, X) est Markovien pour des feedbacks $\phi_t(z, x)$.

Équation de Bellman

$$
\tilde{V}_t(z, x) = \min_{u \in \mathbb{U}} \mathbb{E}\left[\tilde{V}_{t+1}(z + L_t(x, u, \mathrm{W}_{t+1}), f_t(x, u, \mathrm{W}_{t+1}))\right]
$$

$$
\tilde{V}_T(z, x) = z + K(x)
$$

Preuve

- On montre que $\tilde{V}_t(z,x) = z + V_t(x)$
- C'est vrai pour $t = T$, en effet $\tilde{V}_T(z,x) = z + K(x)$. Puis :

$$\tilde{V}_t(z,x) = \min_{u \in \mathbb{U}} \mathbb{E}\big[z + L_t(x,u,\mathrm{W}_{t+1}) + V_{t+1}(f_t(x,u,\mathrm{W}_{t+1}))\big]$$

$$\tilde{V}_t(z,x) = z + \underbrace{\min_{u \in \mathbb{U}} \mathbb{E}\big[L_t(x,u,\mathrm{W}_{t+1}) + V_{t+1}(f_t(x,u,\mathrm{W}_{t+1}))\big]}_{V_t(x)}$$

- On remarque alors que le min en $u \in \mathbb{U}$ ne dépend que de x. Le contrôle optimal est en feedback seulement sur x.
- On note aussi que $\tilde{V}_t(0,x) = V_t(x)$, ce qui donne l'équation de Bellman pour le problème avec coût instantané (V).

Plus court chemin dans un graphe

Si le chemin $A \to B \to C \to D \to E$ est optimal alors $C \to D \to E$ est optimal.

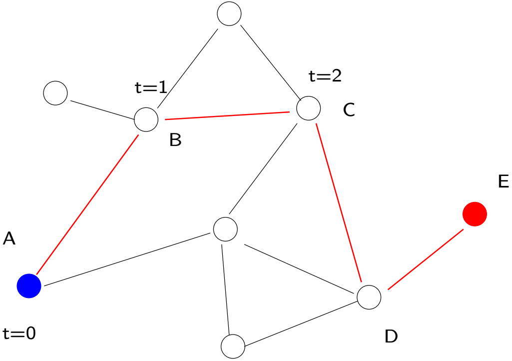

Figure – Plus court chemin

Principe de programmation dynamique

- Plus court chemin pour toutes les paires (i,j) du graphes ($w_{i,j}$ poids de l'arc (i,j)).
- $d_{ij}^{(m)}$: valeur du plus court chemin allant de i à j avec au plus m arcs.
- $d_{i,j}^{(0)} = 0$ si $i = j$ et $+\infty$ sinon.
- Principe de programmation dynamique

$$\begin{aligned} d_{ij}^{(m)} &= \min\left(d_{ij}^{(m-1)}, \min_{1\leq k\leq n, k\neq j}(d_{ik}^{(m-1)} + w_{kj})\right) \\ &= \min_{1\leq k\leq n}(d_{ik}^{(m-1)} + w_{kj}) \end{aligned}$$

- Si pas de cycles de poids négatif alors $d^{(n-1)}$ ou n est le nombre de nœuds du graphe donne la solution du problème.

Quelques éléments toujours présents

- Un problème de départ : trouver $d^{(n-1)}$ est remplacé par une famille de problèmes $(d^{(m)},\ m = 0, \ldots, n-1)$.
- Le problème de départ est bien sûr l'un des problèmes.
- On établit une relation de récurrence entre les $d^{(m)}$.
- Dans la récurrence apparaît un problème d'optimisation local.
- On a remplacé le calcul du plus court chemin par le calcul de la valeur du plus court chemin.
- Le plus court chemin se retrouve en gardant en mémoire l'argmin des problèmes récursifs.

Retour sur le plus court chemin

- Fonction valeur :

$$d_{ij}^{(m)} = \min_{1 \le k \le n} (d_{ik}^{(m-1)} + w_{kj})$$

- Comment aller de i à j par le plus court chemin. On regarde l'argmin du problème

$$d_{ij}^{(n-2)} = \min_{1 \le k \le n} (d_{ik}^{(n-1)} + w_{kj})$$

  Si cet argmin vaut $k^\sharp$ cela veut dire que le plus court chemin $i \rightsquigarrow j = i \rightsquigarrow k^\sharp \to j$
- On regarde alors comment aller de i à $k^\sharp$ en au plus $n-2$ arcs.
- Noter que $d^{(n-1)} = (W)^{(n-1)}$ (dans l'algèbre $(\max, +)$) : carré itérés.

Récursion

- $Fib(n) = (n <= 1) ? 1 : Fib(n-1) + Fib(n-2)$.

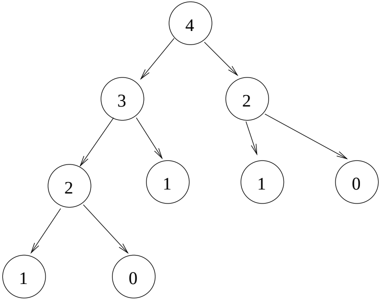

Figure – Fibonacci

Récursion

- Complexité exponentielle si l'on y prend garde !
- Solution 1 : partir de $(Fib(0), Fib(1))$ et calculer itérativement (plus de récursion).
- Solution 2 : utiliser la récursion mais garder en mémoire les valeurs déjà calculées (fonction à mémoire ou memoization (en anglais)).

POUR ALLER PLUS LOIN : COURS DE L'ÉCOLE DES PONTS

- cours de « Recherche Opérationnelle »
- cours d'« Optimisation et contrôle »
- cours « Modéliser l'Aléa »
- cours « Finance : aspects mathématiques et numériques ».