

# IFT-6521

## PROGRAMMATION DYNAMIQUE

### Chapitre 4:

### Modèle stochastique sur horizon fini

Pierre L'Ecuyer

DIRO, Université de Montréal

Février 2015

## Modèle stochastique sur horizon fini

À l'étape  $k$ , on observe l'état  $x_k$  et prend une décision  $u_k \in U_k(x_k)$ . Puis une variable aléatoire  $\omega_k$  est générée selon une loi de probabilité  $P_k(\cdot | x_k, u_k)$ . On observe  $\omega_k$ , on paye un coût  $g_k(x_k, u_k, \omega_k)$ , et l'état à la prochaine étape est  $x_{k+1} = f_k(x_k, u_k, \omega_k)$ .

À l'étape  $k$ , on a :

$x_k$  = état du système à l'étape  $k$ ;

$u_k$  = décision prise à l'étape  $k$ .

$\omega_k$  = perturbation (var. aléatoire) produite à l'étape  $k$ .

$g_k$  = fonction de coût;

$f_k$  = fonction de transition;

On cherche une politique admissible  $\pi = (\mu_0, \dots, \mu_{N-1})$  qui minimise l'espérance mathématique du coût total,

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

On a

$$\begin{aligned}
 J_N(x) &= g_N(x) \quad \text{pour } x \in X_N, \\
 J_k(x) &= \text{coût espéré total optimal de l'étape } k \\
 &\quad \text{à la fin, si on est dans l'état } x \text{ à l'étape } k \\
 &= \inf_{u \in U_k(x)} \mathbb{E}_{\omega_k} [g_k(x, u, \omega_k) + J_{k+1}(f_k(x, u, \omega_k))] \\
 &\quad \text{pour } 0 \leq k \leq N - 1, x \in X_k,
 \end{aligned}$$

où  $\omega_k$  suit la loi  $P_k(\cdot \mid x, u)$ , et

$$\mu_k^*(x) = \arg \min_{u \in U_k(x)} \mathbb{E}_{\omega_k} [g_k(x, u, \omega_k) + J_{k+1}(f_k(x, u, \omega_k))].$$

Dans ce chapitre, on va examiner des modèles (ou exemples) plus particuliers et caractériser la forme de la politique optimale  $\mu_k^*$  et de la fonction de coût anticipé  $J_k$ .

## Système linéaire à coût quadratique

$x_k$  = vecteur de dimension  $n$ ,

$u_k$  = vecteur de dimension  $m$ ,

$w_k$  = vecteur de dimension  $n$ ,  $\mathbb{E}[w_k] = 0$ ,  $\text{Var}[w_k] < \infty$ ,  
indép. de  $(x_k, u_k)$ , et les  $w_k$  sont indépendants.

$$x_{k+1} = f_k(x_k, u_k, w_k) = A_k x_k + B_k u_k + w_k,$$

$$g_N(x_N) = x_N^t Q_N x_N,$$

$$g_k(x_k, u_k) = x_k^t Q_k x_k + u_k^t R_k u_k,$$

où  $A_k$ ,  $B_k$ ,  $Q_k$ , et  $R_k$  sont des matrices, les  $Q_k$  sont symétriques et définies semi-positives, les  $R_k$  sont symétriques et définies positives.

Le modèle se généralise facilement au cas où  $\mathbb{E}[w_k] \neq 0$  et où les formes quadratiques sont décentrées. On ne le fera pas pour éviter d'alourdir la notation. On se ramène au cas centré par simple changement de variable.

Les équations de la PD deviennent:

$$J_N(x_N) = g_N(x_N) = x_N^t Q_N x_N,$$

$$J_k(x_k) = \min_{u_k} \mathbb{E} [x_k^t Q_k x_k + u_k^t R_k u_k + J_{k+1}(A_k x_k + B_k u_k + w_k)], \quad k < N.$$

On montre que les fonctions  $J_k$  sont quadratiques en  $x_k$  et  $u_k$ .

Cela implique que  $\mu_k^*(x_k)$  est linéaire en  $x_k$ , et ne dépend pas des lois de probabilité des  $w_k$ . Ainsi, pour ce modèle, on peut remplacer les  $w_k$  par leurs espérances sans changer la politique optimale!

C'est le principe du modèle déterministe équivalent.

Posons  $K_N = Q_N$ , et pour  $k < N$ ,

$$L_k = -(B_k^t K_{k+1} B_k + R_k)^{-1} B_k^t K_{k+1} A_k,$$

$$\begin{aligned} K_k &= A_k^t (K_{k+1} - K_{k+1} B_k (B_k^t K_{k+1} B_k + R_k)^{-1} B_k^t K_{k+1}) A_k + Q_k \\ &= A_k^t K_{k+1} (A_k + B_k L_k) + Q_k. \end{aligned} \quad (\text{Eq. de Riccati})$$

La proposition qui suit donne la forme de la politique optimale, qui est très facile à calculer et à implanter via une rétroaction ("feedback") linéaire, avec matrices de gain  $L_k$ .

**Proposition.** Les matrices  $K_k$  sont symétriques et définies semi-positives et les  $L_k$  sont donc bien définies.

La politique optimale et la fonction de coût optimal  $J_k$ , pour  $0 \leq k < N$ , sont donnés par

$$\begin{aligned}\mu_k^*(x_k) &= L_k x_k, \\ J_k(x_k) &= x_k^t K_k x_k + \sum_{j=k}^{N-1} \mathbb{E}[w_j^t K_{j+1} w_j].\end{aligned}$$

**Preuve.** Par induction arrière sur  $k$ .

Pour  $k = N - 1$ , en utilisant le fait que  $\mathbb{E}[w_{N-1}] = 0$ ,

$$\begin{aligned}
 J_{N-1}(x_{N-1}) &= \min_{u_{N-1}} \mathbb{E} \left[ x_{N-1}^t Q_{N-1} x_{N-1} + u_{N-1}^t R_{N-1} u_{N-1} \right. \\
 &\quad \left. + (A_{N-1} x_{N-1} + B_{N-1} u_{N-1} + w_{N-1})^t \right. \\
 &\quad \left. K_N (A_{N-1} x_{N-1} + B_{N-1} u_{N-1} + w_{N-1}) \right] \\
 &= x_{N-1}^t Q_{N-1} x_{N-1} \\
 &\quad + x_{N-1}^t A_{N-1}^t K_N A_{N-1} x_{N-1} + \mathbb{E}[w_{N-1}^t K_N w_{N-1}] \\
 &\quad + \min_{u_{N-1}} \left[ u_{N-1}^t R_{N-1} u_{N-1} + u_{N-1}^t B_{N-1}^t K_N B_{N-1} u_{N-1} \right. \\
 &\quad \left. + 2x_{N-1}^t A_{N-1}^t K_N B_{N-1} u_{N-1} \right].
 \end{aligned}$$

En mettant à 0 la dérivée p.r. à  $u_{N-1}$ , on obtient

$$\begin{aligned}
 2R_{N-1} u_{N-1} + 2B_{N-1}^t K_N B_{N-1} u_{N-1} + 2(x_{N-1}^t A_{N-1}^t K_N B_{N-1})^t &= 0, \text{ i.e.,} \\
 (R_{N-1} + B_{N-1}^t K_N B_{N-1}) u_{N-1} &= -(x_{N-1}^t A_{N-1}^t K_N B_{N-1})^t.
 \end{aligned}$$

Puisque  $R_{N-1}$  est définie positive et  $B_{N-1}^t K_N B_{N-1}$  est définie semi-positive,  $(R_{N-1} + B_{N-1}^t K_N B_{N-1})$  est inversible. Donc la commande optimale est la valeur de  $u_{N-1}$  qui satisfait cette équation:

$$\begin{aligned}\mu_{N-1}^*(x_{N-1}) &= -(R_{N-1} + B_{N-1}^t K_N B_{N-1})^{-1} B_{N-1}^t K_N A_{N-1} x_{N-1} \\ &= L_{N-1} x_{N-1}.\end{aligned}$$

En remplaçant  $u_{N-1}$  par  $\mu_{N-1}^*(x_{N-1}) = L_{N-1} x_{N-1}$  dans l'expression pour  $J_{N-1}$  et en simplifiant, on obtient:

$$\begin{aligned}J_{N-1}(x_{N-1}) &= x_{N-1}^t Q_{N-1} x_{N-1} + \mathbb{E}[w_{N-1}^t K_N w_{N-1}] \\ &\quad + \min_{u \in \mathbb{R}^m} [u^t R_{N-1} u + (A_{N-1} x_{N-1} + B_{N-1} u)^t K_N (A_{N-1}^t x_{N-1} + B_{N-1} u)] \\ &= x_{N-1}^t K_{N-1} x_{N-1} + \mathbb{E}[w_{N-1}^t K_N w_{N-1}].\end{aligned}$$

La matrice  $K_{N-1}$  est symétrique et elle est définie semi-positive, puisque cette dernière expression est  $\geq 0$  car  $Q_{N-1}$ ,  $R_{N-1}$ , et  $K_N$  sont définies semi-positives. Cela prouve le résultat pour  $k = N - 1$ .

On suppose maintenant que le résultat tient pour  $k + 1$  et on prouve que cela implique qu'il tient pour  $k$ . On a

$$J_{k+1}(x_{k+1}) = x_{k+1}^t K_{k+1} x_{k+1} + \sum_{j=k+1}^{N-1} \mathbb{E}[w_j^t K_{j+1} w_j]$$

où la somme est simplement une constante qui n'a pas d'influence sur la politique optimale.

On peut refaire exactement le même raisonnement que pour  $k + 1 = N$ , en remplaçant  $N$  par  $k + 1$ , et on obtient le résultat.  $\square$

## État partiellement observé.

Un résultat semblable tient aussi si on ne peut pas observer l'état  $x_k$  au complet, mais seulement une transformation linéaire

$$z_k = C_k x_k + v_k$$

où les  $v_k$  sont des vecteurs aléatoires indépendants entre eux et indépendants des  $w_k$ . La politique optimale a alors la forme

$$\mu_k^*(I_k) = L_k \mathbb{E}[x_k | I_k]$$

où  $I_k = (u_0, \dots, u_{k-1}, z_0, \dots, z_k)$  est l'information disponible à l'étape  $k$ .

## Modèle stationnaire sur horizon infini.

Supposons que  $(A_k, B_k, Q_k, R_k) = (A, B, Q, R)$  pour tout  $k$ .

$$\begin{aligned}x_{k+1} &= Ax_k + Bu_k + w_k, \\g_k(x_k, u_k) &= x_k^t Q x_k + u_k^t R u_k.\end{aligned}$$

Lorsque  $k \rightarrow -\infty$ , on s'attend à ce que  $K_k \rightarrow K$  où  $K$  satisfait l'équation de Riccati algébrique

$$K = A^t(K - KB(B^tKB + R)^{-1}B^tK)A + Q. \quad (1)$$

La politique optimale sur horizon infini sera alors stationnaire:

$$\mu^*(x) = Lx \quad \text{où} \quad L = -(B^tKB + R)^{-1}B^tKA.$$

On prouve que cela est vrai sous les conditions qui suivent.  
D'abord quelques définitions.

Décomposons  $Q = C^tC$  où  $C$  est  $r \times n$  si  $r = \text{rang}(Q)$ .

La paire  $(A, B)$  est dite **contrôlable** si la matrice  $n \times nm$

$$(B, AB, A^2B, \dots, A^{n-1}B)$$

est de rang  $n$ .

La **contrôlabilité** implique qu'en l'absence de bruit (si  $w_k = 0$ ), à partir de n'importe quelle position initiale, on peut ramener le système en  $n$  étapes à  $x_n = 0$  par un choix approprié de  $u_0, \dots, u_{n-1}$ :

$$x_n = A^n x_0 + (B, AB, \dots, A^{n-1}B) \begin{pmatrix} u_{n-1} \\ \vdots \\ u_0 \end{pmatrix}.$$

La paire  $(A, C)$  est dite **observable** si la matrice  $n \times nm$

$$\begin{pmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{pmatrix} = (C^t, A^t C^t, \dots, (A^{n-1})^t C^t)^t$$

est de rang  $n$ .

L'**observabilité** implique qu'en observant  $Cx_0, \dots, Cx_{n-1}$ , on peut en déduire l'état initial  $x_0$ , car

$$\begin{pmatrix} Cx_0 \\ \vdots \\ Cx_{n-1} \end{pmatrix} = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{pmatrix} x_0.$$

Cela implique aussi que si  $x_k^t Q x_k \rightarrow 0$ , alors  $x_k \rightarrow 0$ .

En l'absence de bruit, on a

$$x_k = (A + BL)x_{k-1} = (A + BL)^k x_0. \quad (2)$$

Ce système en boucle fermée est dit **stable** si et seulement si  $x_k \rightarrow 0$  lorsque  $k \rightarrow \infty$ , i.e., si et seulement si toutes les valeurs propres de la matrice  $A + BL$  (qui sont en général des nombres complexes) sont **strictement** à l'intérieur du cercle unité.

**Proposition.** Si  $(A, B)$  est contrôlable et  $(A, C)$  est observable, alors

$$\lim_{k \rightarrow -\infty} K_k = K$$

où  $K$  est une matrice définie positive qui est l'unique solution de (1) dans l'espace des matrices définies semi-positives, et le système (2) est stable.

**Preuve:** Voir DPOC.

## Matrices aléatoires.

Supposons que les matrices  $A_k$  et  $B_k$  ne sont plus constantes, mais sont aléatoires. La politique optimale est encore de la forme

$$\mu_k^*(x_k) = L_k x_k,$$

si on prend soin de remplacer les formes quadratiques  $A_k^t K_{k+1} A_k$ ,  $A_k^t K_{k+1} B_k$ , etc., par leurs **espérances conditionnelles**  $\mathbb{E}[A_k^t K_{k+1} A_k \mid K_{k+1}]$ ,  $\mathbb{E}[A_k^t K_{k+1} B_k \mid K_{k+1}]$ , etc., dans les expressions pour  $K_k$  et  $L_k$ .

Par ailleurs, dans le **cas stationnaire**, le système est stable seulement si la mesure d'incertitude

$$T = \mathbb{E}[A^2]\mathbb{E}[B^2] - (\mathbb{E}[A]\mathbb{E}[B])^2$$

n'est pas trop grande. Si  $T$  dépasse un certain seuil (trop d'incertitude),  $K_k$  diverge lorsque  $k \rightarrow -\infty$  et l'optimisation sur horizon infini n'a plus de sens.

## Modèle d'inventaire

$x_k$	=	Niveau d'inventaire au début du mois $k$ , avant de commander (entier ou <b>réel</b> ici);
$u_k$	=	Quantité commandée au début du mois $k$ ;
$y_k = x_k + u_k$	=	Niveau d'inventaire après avoir commandé;
$w_k$	=	Demande durant le mois $k$ (sont indép.);
$K + cu$	=	Coût d'une commande de taille $u$ ;
$r(x_{k+1})$	=	Coût d'inventaire payé à la fin du mois $k$ ;

L'algorithme de la PD vu au chapitre 1 devient très coûteux lorsque le nombre de valeurs de  $x_k$  possibles est très grand.

**Hypothèse:** Supposons que  $K = 0$  et que  $r$  est une fonction **continue**, **convexe** et telle que

$$\lim_{x \rightarrow \infty} r(x) = \lim_{x \rightarrow -\infty} (cx + r(x)) = \infty.$$

**Exemple:**  $r(x) = p \max(0, -x) + h \max(0, x)$  où  $p > c$  et  $h > 0$ .

**Exemple:**  $r(x) = p \max(0, -x) + h \max(0, x)$  où  $p > c$  et  $h > 0$ .

On suppose que les inventaires négatifs sont permis:

$$x_{k+1} = x_k + u_k - w_k = y_k - w_k.$$

On n'a pas à tenir compte des revenus de vente. On a alors

$$\begin{aligned} J_N(x_N) &= 0; \\ J_k(x_k) &= \text{coût espéré total optimal pour les mois } k \text{ à } N \\ &= \min_{u_k \geq 0} (cu_k + \mathbb{E}[r(x_k + u_k - w_k) + J_{k+1}(x_k + u_k - w_k)]) \\ &= \min_{y_k \geq x_k} (cy_k + \mathbb{E}[r(y_k - w_k) + J_{k+1}(y_k - w_k)]) - cx_k \\ &= \min_{y_k \geq x_k} G_k(y_k) - cx_k \end{aligned}$$

où

$$G_k(y) = cy + \mathbb{E}[r(y - w_k) + J_{k+1}(y - w_k)]$$

est le coût espéré pour  $k$  à  $N$  si  $x_k = 0$  et on commande  $y$ .

La proposition qui suit montre que la politique optimale est de toujours ramener le niveau d'inventaire à la valeur  $S_k$  qui minimise  $G_k$  si on est en dessous et de ne rien commander si on est au dessus.

Il suffit donc de trouver le min pour le cas où  $x_k = 0$ .

**Proposition.** Pour chaque  $k < N$ ,  $G_k$  et  $J_k$  sont des fonctions convexes telles que  $G_k(y) \rightarrow \infty$  et  $J_k(y) \rightarrow \infty$  lorsque  $y \rightarrow \pm\infty$ . La politique optimale est définie par

$$\mu_k^*(x) = \max(0, S_k - x)$$

où  $S_k$  est la valeur de  $y$  qui minimise  $G_k(y)$ .

**Preuve.** Par induction arrière sur  $k$ .

Pour  $k = N - 1$ ,  $G_{N-1}(y) = cy + \mathbb{E}[r(y - w_{N-1})]$  est convexe car  $r$  est convexe, et  $G_{N-1}(y) \rightarrow \infty$  quand  $|y| \rightarrow \infty$  en supposant que la loi de  $w_{N-1}$  est "raisonnable". (Pour être rigoureux, il faudrait ajouter des hypothèses sur cette loi et les détails techniques. Bertsekas suppose que les  $w_k$  prennent leurs valeurs dans un intervalle borné, mais on peut être moins restrictif.)

**Preuve.** Par induction arrière sur  $k$ .

Pour  $k = N - 1$ ,  $G_{N-1}(y) = cy + \mathbb{E}[r(y - w_{N-1})]$  est convexe car  $r$  est convexe, et  $G_{N-1}(y) \rightarrow \infty$  quand  $|y| \rightarrow \infty$  en supposant que la loi de  $w_{N-1}$  est "raisonnable". (Pour être rigoureux, il faudrait ajouter des hypothèses sur cette loi et les détails techniques. Bertsekas suppose que les  $w_k$  prennent leurs valeurs dans un intervalle borné, mais on peut être moins restrictif.)

Il y a donc une valeur finie et unique de  $y$  qui minimise  $G_{N-1}(y)$ .

Appelons-la  $S_{N-1}$ . Si  $S_{N-1} \geq x_{N-1}$ , cette valeur est la valeur optimale de  $y_{N-1}$ , i.e., la valeur optimale de  $u_{N-1}$  est  $S_{N-1} - x_{N-1}$ .

On aura  $J_{N-1}(x_{N-1}) = (S_{N-1} - x_{N-1})c + G_{N-1}(S_{N-1})$ .

**Preuve.** Par induction arrière sur  $k$ .

Pour  $k = N - 1$ ,  $G_{N-1}(y) = cy + \mathbb{E}[r(y - w_{N-1})]$  est convexe car  $r$  est convexe, et  $G_{N-1}(y) \rightarrow \infty$  quand  $|y| \rightarrow \infty$  en supposant que la loi de  $w_{N-1}$  est "raisonnable". (Pour être rigoureux, il faudrait ajouter des hypothèses sur cette loi et les détails techniques. Bertsekas suppose que les  $w_k$  prennent leurs valeurs dans un intervalle borné, mais on peut être moins restrictif.)

Il y a donc une valeur finie et unique de  $y$  qui minimise  $G_{N-1}(y)$ .

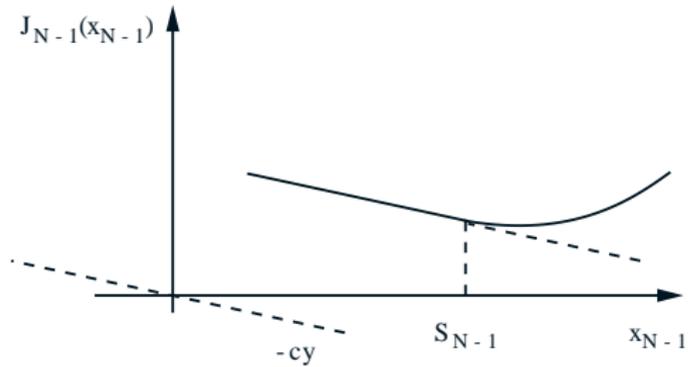
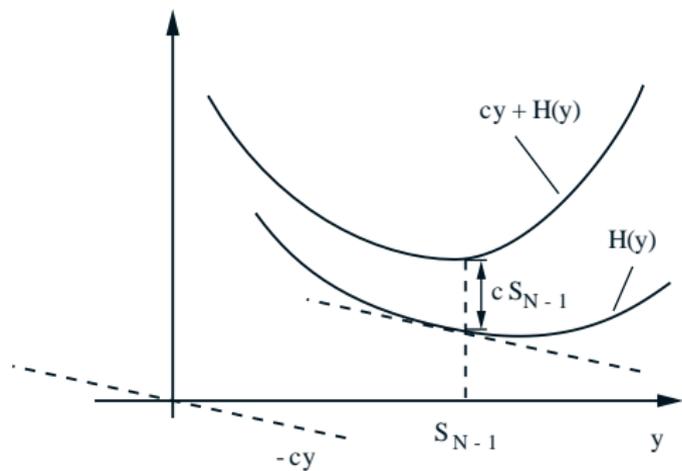
Appelons-la  $S_{N-1}$ . Si  $S_{N-1} \geq x_{N-1}$ , cette valeur est la valeur optimale de  $y_{N-1}$ , i.e., la valeur optimale de  $u_{N-1}$  est  $S_{N-1} - x_{N-1}$ .

On aura  $J_{N-1}(x_{N-1}) = (S_{N-1} - x_{N-1})c + G_{N-1}(S_{N-1})$ .

Si  $S_{N-1} < x_{N-1}$ , on est bloqué par la contrainte  $y_{N-1} \geq x_{N-1} > S_{N-1}$ .

On prendra  $y_{N-1} = x_{N-1}$ , i.e.,  $u_{N-1} = 0$ . Cela donne

$$\mu_{N-1}^*(x) = \max(0, S_{N-1} - x).$$



Il reste à montrer les propriétés pour  $J_{N-1}$ . On a

$$J_{N-1}(x) = \begin{cases} (S_{N-1} - x)c + \mathbb{E}[r(S_{N-1} - w_{N-1})] & \text{if } x < S_{N-1}, \\ \mathbb{E}[r(x - w_{N-1})] & \text{if } x \geq S_{N-1}, \end{cases}$$

qui est continue, convexe, et tend vers l'infini lorsque  $|x| \rightarrow \infty$ , grâce aux propriétés de  $r$ .

Il reste à montrer les propriétés pour  $J_{N-1}$ . On a

$$J_{N-1}(x) = \begin{cases} (S_{N-1} - x)c + \mathbb{E}[r(S_{N-1} - w_{N-1})] & \text{if } x < S_{N-1}, \\ \mathbb{E}[r(x - w_{N-1})] & \text{if } x \geq S_{N-1}, \end{cases}$$

qui est continue, convexe, et tend vers l'infini lorsque  $|x| \rightarrow \infty$ , grâce aux propriétés de  $r$ .

Si on suppose que les propriétés tiennent pour  $k + 1$ , on peut montrer par les mêmes arguments et en utilisant aussi l'hypothèse que  $J_{k+1}$  est convexe, que cela implique qu'elles tiennent pour  $k$ .  $\square$

Exercice: complétez les détails de la preuve pour passer de  $k + 1$  à  $k$ .

## Coût fixe positif.

Supposons maintenant qu'il y a un **coût fixe**  $K > 0$  pour commander. On définit  $G_k$  de la même façon:

$G_k(y)$  = coût espéré pour  $k$  à  $N$  si  $x_k = 0$ , on a déjà payé le coût fixe, et on commande  $y$ . On a

$$\begin{aligned} J_k(x_k) &= \min \left[ G_k(x_k), \min_{u_k \geq 0} (K + G_k(x_k + u_k)) \right] - cx_k \\ &= \min \left[ G_k(x_k), \min_{y_k \geq x_k} (K + G_k(y_k)) \right] - cx_k. \end{aligned}$$

Cette fonction  $J_k$  n'est **pas convexe**. Et  $G_k$  n'est pas nécessairement convexe non plus. Mais on peut prouver la forme de la politique optimale en utilisant une notion plus générale de convexité.

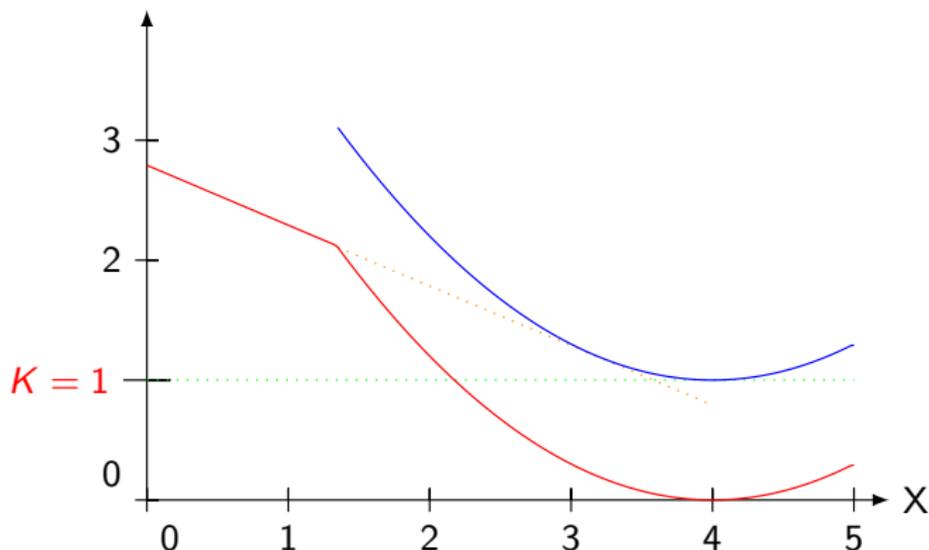
Ce genre de technique peut être utile en général pour déterminer et prouver la forme de la politique optimale dans le cas où il y a des coûts fixes pour les transactions.

**Définition.** Une fonction  $g : \mathbb{R} \rightarrow \mathbb{R}$  est  $K$ -convexe, pour  $K \geq 0$ , si

$$\frac{K + g(y + b_2) - g(y)}{b_2} \geq \frac{g(y) - g(y - b_1)}{b_1} \quad \forall y, \forall b_1 > 0, \forall b_2 > 0.$$

Ici,  $g(\cdot)$  est  $K$ -convexe pour  $K = 1$ , mais pas pour  $K < 1$ .

En bleu:  $g(\cdot) + K$ .



**Proposition.** Pour chaque  $k < N$ ,  $G_k$  et  $J_k$  sont des fonctions continues et  $K$ -convexes telles que  $G_k(y) \rightarrow \infty$  et  $J_k(y) \rightarrow \infty$  lorsque  $y \rightarrow \pm\infty$ . La politique optimale est une politique de type  $(s, S)$  non stationnaire, définie par

$$\mu_k^*(x_k) = \begin{cases} S_k - x_k & \text{si } x_k < s_k; \\ 0 & \text{si } x_k \geq s_k; \end{cases}$$

où  $S_k$  est la valeur de  $y$  qui minimise  $G_k(y)$  et  $s_k$  est la plus petite valeur de  $y$  telle que  $G_k(y) = K + G_k(S_k)$ .

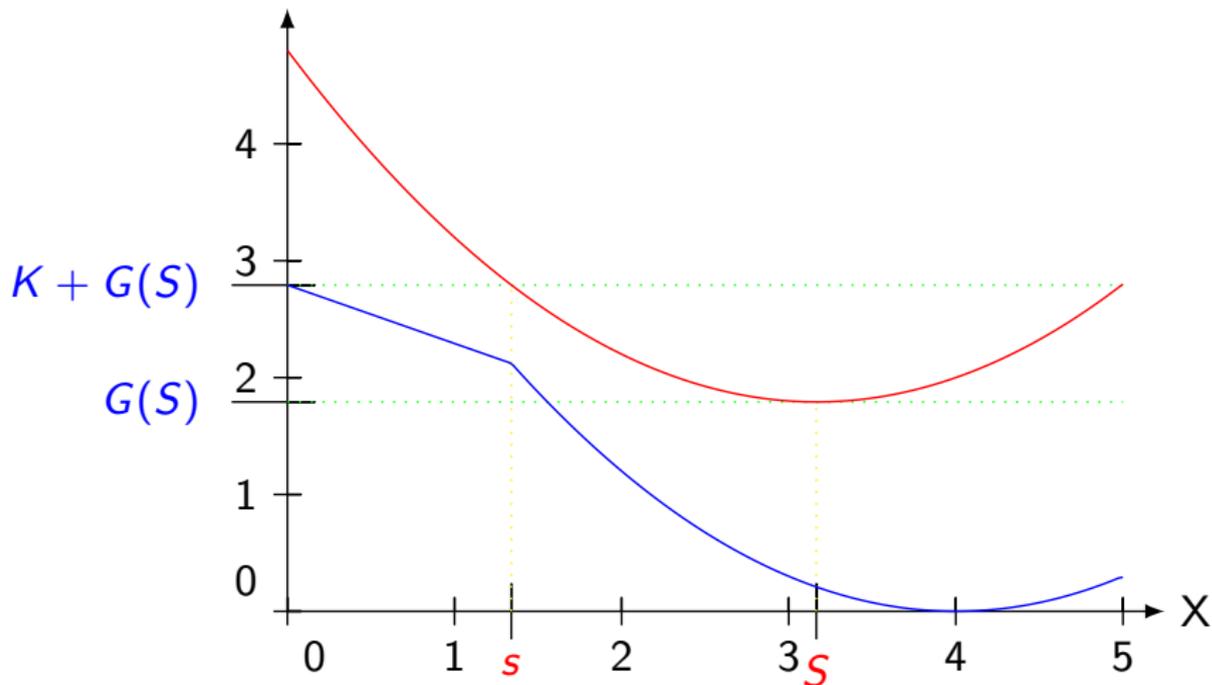
Si  $K = 0$ , alors  $s_k = S_k$ .

Intérêt du résultat: simplifie beaucoup les calculs!

$G_{N-1}(x) = cx + \mathbb{E}[r(x - w_{N-1})]$  et

$$J_{N-1}(x) = \min [G_{N-1}(x), K + \min_{y \geq x} G_{N-1}(y)] - cx$$

$$= \min [G_{N-1}(x), K + G_{N-1}(S)] - cx.$$



**Définition.** Une fonction  $g : \mathbb{R} \rightarrow \mathbb{R}$  est  $K$ -convexe, pour  $K \geq 0$ , si

$$\frac{K + g(y + b_2) - g(y)}{b_2} \geq \frac{g(y) - g(y - b_1)}{b_1} \quad \forall y, \forall b_1 > 0, \forall b_2 > 0.$$

**Définition.** Une fonction  $g : \mathbb{R} \rightarrow \mathbb{R}$  est  $K$ -convexe, pour  $K \geq 0$ , si

$$\frac{K + g(y + b_2) - g(y)}{b_2} \geq \frac{g(y) - g(y - b_1)}{b_1} \quad \forall y, \forall b_1 > 0, \forall b_2 > 0.$$

**Lemme 4.2.1** (page 167).

- (a) Si  $g$  est convexe, alors  $g$  est  $K$ -convexe pour tout  $K \geq 0$ .
- (b) Si  $g_1$  est  $K$ -convexe et  $g_2$   $L$ -convexe, alors  $\alpha g_1 + \beta g_2$  est  $(\alpha K + \beta L)$ -convexe.
- (c) Si  $g$  est  $K$ -convexe et  $W$  une v.a. aléatoire, alors  $h(y) = \mathbb{E}[|g(y - W)|]$  est  $K$ -convexe si  $h(y) < \infty$  pour tout  $y$ .
- (d) Si  $g$  est continue et  $K$ -convexe et si  $\lim_{y \rightarrow \pm\infty} g(y) = \infty$ , alors il existe deux constantes  $s \leq S$  telles que:
  - (i)  $g(y) \geq g(S)$  pour tout  $y$ ;
  - (ii)  $g(y) > g(s) = g(S) + K$  pour  $y < s$ ;
  - (iii)  $g(y)$  est décroissante pour  $y < s$ ;
  - (iv)  $g(y) \leq g(z) + K$  si  $s \leq y \leq z$ .

Preuve du lemme. (a), (b), (c) découlent directement de la définition.  
Prouvons (d).

Preuve du lemme. (a), (b), (c) découlent directement de la définition.  
Prouvons (d).

(i): Puisque  $g(y) \rightarrow \infty$  lorsque  $|y| \rightarrow \infty$ ,  $g$  possède un minimum. Soit  $S$  un endroit où il est atteint.

Preuve du lemme. (a), (b), (c) découlent directement de la définition.  
Prouvons (d).

(i): Puisque  $g(y) \rightarrow \infty$  lorsque  $|y| \rightarrow \infty$ ,  $g$  possède un minimum. Soit  $S$  un endroit où il est atteint.

(ii): Soit  $s \leq S$  le plus petit  $y$  tel que  $g(y) = g(S) + K$ .  
Pour  $y < s$ , on a par la  $K$ -convexité

$$\frac{K + g(S) - g(s)}{S - s} \geq \frac{g(s) - g(y)}{s - y}.$$

Puisque  $K + g(S) - g(s) = 0$ , cela donne  $g(y) \geq g(s) = K + g(S)$ . Mais par définition de  $s$ , on a  $g(y) > g(S) + K = g(s)$ .

Preuve du lemme. (a), (b), (c) découlent directement de la définition.  
Prouvons (d).

(i): Puisque  $g(y) \rightarrow \infty$  lorsque  $|y| \rightarrow \infty$ ,  $g$  possède un minimum. Soit  $S$  un endroit où il est atteint.

(ii): Soit  $s \leq S$  le plus petit  $y$  tel que  $g(y) = g(S) + K$ .  
Pour  $y < s$ , on a par la  $K$ -convexité

$$\frac{K + g(S) - g(s)}{S - s} \geq \frac{g(s) - g(y)}{s - y}.$$

Puisque  $K + g(S) - g(s) = 0$ , cela donne  $g(y) \geq g(s) = K + g(S)$ . Mais par définition de  $s$ , on a  $g(y) > g(S) + K = g(s)$ .

(iii) Pour  $y_1 < y_2 < s$ , on a

$$0 > \frac{K + g(S) - g(y_2)}{S - y_2} \geq \frac{g(y_2) - g(y_1)}{y_2 - y_1}$$

et donc  $g(y_1) > g(y_2)$ .

(iv) On veut montrer que  $K + g(z) \geq g(y)$  si  $s \leq y \leq z$ .

Si  $y = s$  ou  $y = S$  ou  $y = z$ , facile à vérifier.

Si  $y > S$ , alors par la  $K$ -convexité,

$$\frac{K + g(z) - g(y)}{z - y} \geq \frac{g(y) - g(S)}{y - S} \geq 0.$$

Si  $s < y < S$ , alors

$$\frac{g(s) - g(y)}{S - y} = \frac{K + g(S) - g(y)}{S - y} \geq \frac{g(y) - g(s)}{y - s}$$

qui implique

$$(g(s) - g(y))(y - s) \geq (g(y) - g(s))(S - y)$$

et donc

$$(g(s) - g(y))(S - s) \geq 0.$$

Ainsi

$$K + g(z) \geq K + g(S) = g(s) \geq g(y).$$

**Preuve de la proposition.** Par induction sur  $k$ , comme d'habitude.

Pour  $k = N - 1$ ,  $G_{N-1}$  est convexe, et donc  $K$ -convexe. On a

$$\begin{aligned} J_{N-1}(x) &= \min \left[ G_{N-1}(x), \min_{y \geq x} (K + G_{N-1}(y)) \right] - cx \\ &= \begin{cases} K + G_{N-1}(S_{N-1}) - cx & \text{si } x < s_{N-1}, \\ G_{N-1}(x) - cx & \text{si } x \geq s_{N-1}. \end{cases} \end{aligned}$$

La politique  $\mu_{N-1}^*$  qui fait atteindre le min a la forme voulue.

**Preuve de la proposition.** Par induction sur  $k$ , comme d'habitude.

Pour  $k = N - 1$ ,  $G_{N-1}$  est convexe, et donc  $K$ -convexe. On a

$$\begin{aligned} J_{N-1}(x) &= \min \left[ G_{N-1}(x), \min_{y \geq x} (K + G_{N-1}(y)) \right] - cx \\ &= \begin{cases} K + G_{N-1}(s_{N-1}) - cx & \text{si } x < s_{N-1}, \\ G_{N-1}(x) - cx & \text{si } x \geq s_{N-1}. \end{cases} \end{aligned}$$

La politique  $\mu_{N-1}^*$  qui fait atteindre le min a la forme voulue.

Pour montrer que  $J_{N-1}$  est  $K$ -convexe, on considère 3 cas:

(1)  $y \geq s_{N-1}$ , (2)  $y < y + b_2 \leq s_{N-1}$ , et (3)  $y < s_{N-1} < y + b_2$ .

Dans chaque cas, on montre que la définition de  $K$ -convexité tient. On montre facilement que  $G_{N-1}$  est continue, car  $r$  l'est, et que  $J_{N-1}$  est continue.  $G_{N-1}(y) \rightarrow \infty$  et  $J_{N-1}(y) \rightarrow \infty$  lorsque  $y \rightarrow \pm\infty$  de la même façon que pour le cas où  $K = 0$ . On a donc le résultat pour  $k = N - 1$ .

**Preuve de la proposition.** Par induction sur  $k$ , comme d'habitude.

Pour  $k = N - 1$ ,  $G_{N-1}$  est convexe, et donc  $K$ -convexe. On a

$$\begin{aligned} J_{N-1}(x) &= \min \left[ G_{N-1}(x), \min_{y \geq x} (K + G_{N-1}(y)) \right] - cx \\ &= \begin{cases} K + G_{N-1}(s_{N-1}) - cx & \text{si } x < s_{N-1}, \\ G_{N-1}(x) - cx & \text{si } x \geq s_{N-1}. \end{cases} \end{aligned}$$

La politique  $\mu_{N-1}^*$  qui fait atteindre le min a la forme voulue.

Pour montrer que  $J_{N-1}$  est  $K$ -convexe, on considère 3 cas:

(1)  $y \geq s_{N-1}$ , (2)  $y < y + b_2 \leq s_{N-1}$ , et (3)  $y < s_{N-1} < y + b_2$ .

Dans chaque cas, on montre que la définition de  $K$ -convexité tient. On montre facilement que  $G_{N-1}$  est continue, car  $r$  l'est, et que  $J_{N-1}$  est continue.  $G_{N-1}(y) \rightarrow \infty$  et  $J_{N-1}(y) \rightarrow \infty$  lorsque  $y \rightarrow \pm\infty$  de la même façon que pour le cas où  $K = 0$ . On a donc le résultat pour  $k = N - 1$ .

On montre ensuite que si on suppose le résultat vrai pour  $k + 1$ , cela implique qu'il tient aussi pour  $k$ .  $\square$

**Preuve de la proposition.** Par induction sur  $k$ , comme d'habitude.

Pour  $k = N - 1$ ,  $G_{N-1}$  est convexe, et donc  $K$ -convexe. On a

$$\begin{aligned} J_{N-1}(x) &= \min \left[ G_{N-1}(x), \min_{y \geq x} (K + G_{N-1}(y)) \right] - cx \\ &= \begin{cases} K + G_{N-1}(s_{N-1}) - cx & \text{si } x < s_{N-1}, \\ G_{N-1}(x) - cx & \text{si } x \geq s_{N-1}. \end{cases} \end{aligned}$$

La politique  $\mu_{N-1}^*$  qui fait atteindre le min a la forme voulue.

Pour montrer que  $J_{N-1}$  est  $K$ -convexe, on considère 3 cas:

(1)  $y \geq s_{N-1}$ , (2)  $y < y + b_2 \leq s_{N-1}$ , et (3)  $y < s_{N-1} < y + b_2$ .

Dans chaque cas, on montre que la définition de  $K$ -convexité tient. On montre facilement que  $G_{N-1}$  est continue, car  $r$  l'est, et que  $J_{N-1}$  est continue.  $G_{N-1}(y) \rightarrow \infty$  et  $J_{N-1}(y) \rightarrow \infty$  lorsque  $y \rightarrow \pm\infty$  de la même façon que pour le cas où  $K = 0$ . On a donc le résultat pour  $k = N - 1$ .

On montre ensuite que si on suppose le résultat vrai pour  $k + 1$ , cela implique qu'il tient aussi pour  $k$ .  $\square$

Que se passe-t-il si les  $w_k$  sont i.i.d. et si  $k \rightarrow -\infty$ ?

## Un modèle de gestion de portefeuille.

$x_0$  = capital initial à investir;

$x_k$  = capital de l'investisseur au début de la période  $k$ ;

$n$  types d'actifs risqués et un type d'actif sans risque;

$s_k$  = taux de rendement sans risque pour la période  $k$ ;

$e_k = (e_{k,1}, \dots, e_{k,n})$  = taux de rendement des  $n$  actifs risqués pour la période  $k$ . Ce sont des vecteurs aléatoires indépendants, dont la loi est telle que les espérances que nous allons considérer sont finies;

$u_k = (u_{k,1}, \dots, u_{k,n})$  = montants investis dans les différents actifs risqués à la période  $k$ ; Pas de frais de transactions!

Le capital évolue selon:

$$\begin{aligned} x_{k+1} &= \sum_{i=1}^n e_{k,i} u_{k,i} + s_k (x_k - u_{k,1} - \dots - u_{k,n}) \\ &= s_k x_k + \sum_{i=1}^n (e_{k,i} - s_k) u_{k,i}. \end{aligned}$$

L'investisseur a une **fonction d'utilité**  $U$ , et veut maximiser  $\mathbb{E}[U(x_N)]$ , l'espérance de l'utilité de son capital final.

**Hypothèses:** On suppose que  $U$  est **concave** et possède une dérivée seconde continue, que  $\mathbb{E}[U(x_N)] < \infty$ , et que  $U$  satisfait:

$$-U'(x)/U''(x) = a + bx.$$

L'investisseur a une **fonction d'utilité**  $U$ , et veut maximiser  $\mathbb{E}[U(x_N)]$ , l'espérance de l'utilité de son capital final.

**Hypothèses:** On suppose que  $U$  est **concave** et possède une dérivée seconde continue, que  $\mathbb{E}[U(x_N)] < \infty$ , et que  $U$  satisfait:

$$-U'(x)/U''(x) = a + bx.$$

Par **exemple**, les fonctions suivantes satisfont ces conditions:

$$U(x) = 1 - e^{-x/a}, \quad U(x) = \ln(x + a).$$

L'investisseur a une **fonction d'utilité**  $U$ , et veut maximiser  $\mathbb{E}[U(x_N)]$ , l'espérance de l'utilité de son capital final.

**Hypothèses:** On suppose que  $U$  est **concave** et possède une dérivée seconde continue, que  $\mathbb{E}[U(x_N)] < \infty$ , et que  $U$  satisfait:

$$-U'(x)/U''(x) = a + bx.$$

Par **exemple**, les fonctions suivantes satisfont ces conditions:

$$U(x) = 1 - e^{-x/a},$$

$$U(x) = \ln(x + a).$$

$$U'(x) = e^{-x/a}/a,$$

$$U'(x) = 1/(x + a)$$

$$U''(x) = -e^{-x/a}/a^2,$$

$$U''(x) = -1/(x + a)^2$$

$$b = 0,$$

$$b = 1$$

Les équations de la PD:

$$\begin{aligned} J_N(x_N) &= U(x_N), \\ J_k(x_k) &= \max_{\pi} \mathbb{E}[U(x_N) \mid x_k] \\ &= \max_{u_{k,1}, \dots, u_{k,n}} \mathbb{E} \left[ J_{k+1} \left( s_k x_k + \sum_{i=1}^n (e_{k,i} - s_k) u_{k,i} \right) \right]. \end{aligned} \quad (3)$$

Les équations de la PD:

$$\begin{aligned}
 J_N(x_N) &= U(x_N), \\
 J_k(x_k) &= \max_{\pi} \mathbb{E}[U(x_N) \mid x_k] \\
 &= \max_{u_{k,1}, \dots, u_{k,n}} \mathbb{E} \left[ J_{k+1} \left( s_k x_k + \sum_{i=1}^n (e_{k,i} - s_k) u_{k,i} \right) \right]. \quad (3)
 \end{aligned}$$

**Proposition.** Pour  $0 \leq k < N$ , on a

$$\mu_k^*(x_k) = \left( \frac{a}{s_{N-1} \cdots s_{k+1}} + b s_k x_k \right) \alpha_k$$

où  $\alpha_k = (\alpha_{k,1}, \dots, \alpha_{k,n})$  dépend de la loi de probabilité de  $e_k$  mais pas de  $x_k$ . (On peut déterminer  $\alpha_k$  via (3).) De plus,  $J_k$  satisfait

$$-\frac{J'_k(x)}{J''_k(x)} = \frac{a}{s_{N-1} \cdots s_k} + bx.$$

**Preuve.** (Pas complètement rigoureuse, comme dans DPOC.)

On suppose que la politique optimale existe et est différentiable en  $x_k$  (vrai si  $U$  et les lois des  $e_k$  sont suffisamment régulières). Sans perte de généralité, posons

$$\mu_k^*(x) = \alpha_k(x) \left( \frac{a}{s_{N-1} \cdots s_{k+1}} + bs_k x \right)$$

où les  $\alpha_k(x) = (\alpha_{k,1}(x), \dots, \alpha_{k,n}(x))$  sont des vecteurs de fonctions différentiables. On va montrer que  $d\alpha_{k,i}(x)/dx = 0$ , ce qui impliquera que  $\alpha_{k,i}(x)$  est constante en  $x$ .

La preuve se fait par induction arrière sur  $k$ .

Prenons  $k = N - 1$ ,  $x_{N-1} = x$  et  $s_{N-1} = s$ .

Puisque  $\mu_{N-1}^*(x)$  est le portefeuille optimal, on a

$$\begin{aligned} 0 &= \left. \frac{d\mathbb{E}[U(x_N)]}{du_{N-1,i}} \right|_{u_{N-1}=\mu_{N-1}^*(x)} \\ &= \frac{d}{du_{N-1,i}} \mathbb{E} \left[ U \left( sx + \sum_{j=1}^n (e_{N-1,j} - s) u_{N-1,j} \right) \right] \\ &= \mathbb{E} \left[ (e_{N-1,i} - s) U' \left( sx + \sum_{j=1}^n (e_{N-1,j} - s) \alpha_{N-1,j}(x) (a + bsx) \right) \right] \end{aligned}$$

en supposant que l'on peut échanger la dérivée et l'espérance (le théorème de convergence dominée de Lebesgue donne des conditions suffisantes pour cela).

En dérivant la dernière équation par rapport à  $x$ , et en supposant encore<sup>35</sup> que l'on peut passer la dérivée à l'intérieur de  $\mathbb{E}$ , on obtient

$$0 = \mathbb{E} \left[ (e_{N-1,i} - s) U''(x_N) \left( s + \sum_{j=1}^n (e_{N-1,j} - s) \left( \alpha_{N-1,j}(x) bs + \frac{d\alpha_{N-1,j}}{dx}(x)(a + bsx) \right) \right) \right]$$

$$= \mathbb{E} \left[ U''(x_N)(a + bsx) \sum_{j=1}^n (e_{N-1,i} - s)(e_{N-1,j} - s) \frac{d\alpha_{N-1,j}(x)}{dx} \right] \quad (4)$$

$$+ \mathbb{E} \left[ U''(x_N)(e_{N-1,i} - s)s \left( 1 + \sum_{j=1}^n (e_{N-1,j} - s)\alpha_{N-1,j}(x)b \right) \right]. \quad (5)$$

Mais on sait que

$$-\frac{U'(x_N)}{U''(x_N)} = a + bx_N = a + b \left( sx + \sum_{j=1}^n (e_{N-1,j} - s)\alpha_{N-1,j}(x)(a + bsx) \right)$$

$$= (a + bsx) \left( 1 + \sum_{j=1}^n (e_{N-1,j} - s)\alpha_{N-1,j}(x)b \right). \quad (6)$$

Si on isole  $U''(x_N)$  dans cette expression et on remplace dans (5), on obtient

$$-\mathbb{E}[U'(x_N)(e_{N-1,i} - s)]s/(a + bsx),$$

et on a vu plus haut que cette espérance vaut 0.

On obtient ainsi, pour  $i = 1, \dots, n$ ,

$$\mathbb{E} \left[ (a + bsx)U''(x_N) \sum_{j=1}^n (e_{N-1,i} - s)(e_{N-1,j} - s) \frac{d\alpha_{N-1,j}(x)}{dx} \right] = 0.$$

ce système de  $n$  équations s'écrit sous forme matricielle:

$$(a + bsx)\mathbf{M}\nabla_x\alpha_{N-1}(x) = 0,$$

où  $\nabla_x\alpha_{N-1}(x)$  est un vecteur dont l'élément  $j$  est  $d\alpha_{N-1,j}(x)/dx$ , et  $\mathbf{M}$  est une matrice  $n \times n$  dont l'élément  $(i, j)$  est

$$\mathbb{E}[U''(x_N)(e_{N-1,i} - s)(e_{N-1,j} - s)].$$

Si on suppose que  $\mathbf{M}$  est inversible (ce qui est vrai sauf dans les cas dégénérés), on obtient que  $d\alpha_{N-1,j}(x)/dx = 0$  pour tout  $j$ .

Montrons maintenant que

$$-\frac{J'_{N-1}(x)}{J''_{N-1}(x)} = \frac{a}{s} + bx. \quad (7)$$

Notant  $J(x) = J_{N-1}(x) = \mathbb{E}[U(x_N)]$ , On a

$$J(x) = \mathbb{E} \left[ U \left( \left( 1 + \sum_{j=1}^n (e_{N-1,j} - s) \alpha_{N-1,j} b \right) sx + \sum_{j=1}^n (e_{N-1,j} - s) \alpha_{N-1,j} a \right) \right].$$

On dérive 2 fois par rapport à  $x$  pour obtenir  $J'(x)$  et  $J''(x)$ :

$$J'(x) = \mathbb{E} \left[ U'(x_N) \left( 1 + \sum_{j=1}^n (e_{N-1,j} - s) \alpha_{N-1,j} b \right) s \right]$$

$$J''(x) = \mathbb{E} \left[ U''(x_N) \left( 1 + \sum_{j=1}^n (e_{N-1,j} - s) \alpha_{N-1,j} b \right)^2 s^2 \right].$$

Puis, si on exprime  $U''(x)$  en fonction de  $U'(x)$  via (6) et si on remplace dans  $J''(x)$ , puis on divise l'expression de  $J'(x)$  par cette expression de  $J''(x)$ , on obtient (7).

Induction: on suppose le résultat vrai pour  $k + 1$ , et on montre que cela implique qu'il est vrai pour  $k$ . Les détails sont semblables au cas  $k = N - 1$ . □

**Remarque:** La politique optimale à l'étape  $k$  est la même que si c'était la dernière étape et que la fonction d'utilité était  $J_{k+1}$ . Cette fonction  $J_{k+1}$  a exactement la même forme que la fonction d'utilité, avec  $a$  remplacé par  $a/(s_{N-1} \cdots s_k)$  dans l'expression qui la définit. La politique optimale aura donc la même forme que si on était à la dernière étape, avec  $a$  remplacé par  $a/(s_{N-1} \cdots s_k)$ .

Cela équivaut à dire que sous les hypothèses de ce modèle, on peut optimiser à l'étape  $k$  en faisant l'hypothèse qu'à partir de la prochaine étape tout notre argent sera investi sans risque. Cela revient à maximiser, par rapport à  $u_k = (u_{k,1}, \dots, u_{k,n})$ ,

$$\mathbb{E} \left[ U \left( s_{k+1} \cdots s_{N-1} \left( s_k x_k + \sum_{i=1}^n (e_{k,i} - s_k) u_{k,i} \right) \right) \right].$$

Il s'agit d'une politique **partiellement myope**.

**Remarque:** La politique optimale à l'étape  $k$  est la même que si c'était la dernière étape et que la fonction d'utilité était  $J_{k+1}$ . Cette fonction  $J_{k+1}$  a exactement la même forme que la fonction d'utilité, avec  $a$  remplacé par  $a/(s_{N-1} \cdots s_k)$  dans l'expression qui la définit. La politique optimale aura donc la même forme que si on était à la dernière étape, avec  $a$  remplacé par  $a/(s_{N-1} \cdots s_k)$ .

Cela équivaut à dire que sous les hypothèses de ce modèle, on peut optimiser à l'étape  $k$  en faisant l'hypothèse qu'à partir de la prochaine étape tout notre argent sera investi sans risque. Cela revient à maximiser, par rapport à  $u_k = (u_{k,1}, \dots, u_{k,n})$ ,

$$\mathbb{E} \left[ U \left( s_{k+1} \cdots s_{N-1} \left( s_k x_k + \sum_{i=1}^n (e_{k,i} - s_k) u_{k,i} \right) \right) \right].$$

Il s'agit d'une politique **partiellement myope**.

Si  $a = 0$ , ou si  $s_k = 1$  pour tout  $k$ , alors la politique optimale est la même à toutes les étapes et maximise simplement l'utilité espérée de  $x_{k+1}$ , i.e., on agit toujours comme si on était à l'étape  $N - 1$ . Idem si  $s_k > 1$  et  $k \rightarrow -\infty$ . Cela s'appelle une politique **myope**.

## Temps d'arrêt optimal

Supposons qu'à chaque étape  $k$ , on doit décider si on arrête le système (et on encaisse un revenu ou un coût) ou si on continue. L'espace d'états  $X_k$  peut être partitionné en deux: les états où il est optimal de s'arrêter et ceux où il est optimal de continuer.

### Exemple: vente d'un actif.

Durant chaque période  $k$ ,  $0 \leq k \leq N - 1$ , on reçoit une offre  $w_k$  que l'on peut accepter à la fin de la période (au temps  $k + 1$ ) si aucune offre n'a encore été acceptée.

On suppose que les  $w_k$  sont des v.a. indépendantes, et indépendantes de nos décisions (pas nécessairement i.i.d.).

On pose  $u_k = 1$  si on vend à l'étape  $k$ ,  $u_k = 0$  sinon.

L'état du système à l'étape  $k + 1$  sera

$$x_{k+1} = \begin{cases} w_k & \text{si } u_0 = \dots = u_k = 0; \\ \Delta & \text{sinon;} \end{cases}$$

où  $\Delta$  est l'état dans lequel on a déjà vendu.

On pose  $x_0 = u_0 = 0$ .

Les décisions admissibles pour  $k \geq 1$  sont  $u_k = 0$  si  $x_k = \Delta$ ,  $u_N = 1$  si  $x_N \neq \Delta$ , et  $u_k \in \{0, 1\}$  sinon.

Si  $u_k = 1$ , on reçoit  $x_k = w_{k-1}$  au temps  $k$ , que l'on place au taux d'intérêt  $r$  pour les  $N - k$  périodes restantes. Le revenu à l'étape  $k$ , actualisé au temps  $N$ , est donc

$$g_k(x_k, u_k, w_k) = \begin{cases} (1 + r)^{N-k} x_k & \text{si } u_k = 1; \\ 0 & \text{sinon.} \end{cases}$$

Soit  $J_k(x_k)$  le **revenu espéré optimal** à partir du temps  $k$  jusqu'à la fin, actualisé au temps  $N$ , si on est dans l'état  $x_k$ . On obtient:

$$J_k(x_k) = \begin{cases} 0 & \text{si } x_k = \Delta; \\ x_N & \text{si } k = N \text{ et } x_N \neq \Delta; \\ \max((1+r)^{N-k}x_k, \mathbb{E}[J_{k+1}(w_k)]) & \text{si } k < N \text{ et } x_k \neq \Delta. \end{cases}$$

La **décision optimale**, lorsqu'on a encore le choix, est de vendre (accepter l'offre  $x_k = w_{k-1}$ ) si et seulement si

$$x_k \geq \alpha_k \stackrel{\text{def}}{=} \frac{\mathbb{E}[J_{k+1}(w_k)]}{(1+r)^{N-k}}.$$

La **politique optimale** est déterminée par ces seuils  $\alpha_1, \dots, \alpha_{N-1}$ . Voyons comment les calculer.

On pose  $\alpha_N = 0$ . Pour  $x_k \neq \Delta$ , le revenu total espéré actualisé au temps  $k$  est

$$\begin{aligned} V_k(x_k) &= J_k(x_k)/(1+r)^{N-k} \\ &= \begin{cases} x_N & \text{si } k = N; \\ \max(x_k, \mathbb{E}[V_{k+1}(w_k)]/(1+r)) & \text{si } k < N \end{cases} \\ &= \max(x_k, \alpha_k). \end{aligned}$$

On a

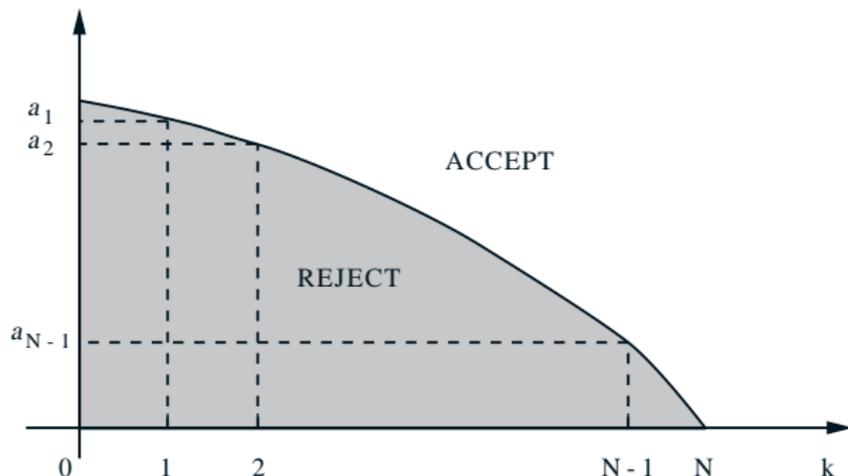
$$\begin{aligned} \alpha_k &= \mathbb{E}[J_{k+1}(w_k)]/(1+r)^{N-k} = \mathbb{E}[V_{k+1}(w_k)]/(1+r); \\ (1+r)\alpha_k &= \mathbb{E}[V_{k+1}(x_{k+1})] \\ &= \mathbb{E}[\max(x_{k+1}, \alpha_{k+1})] \\ &= \mathbb{E}[\max(w_k, \alpha_{k+1})] \\ &= \mathbb{E}[w_k \mathbb{I}[w_k > \alpha_{k+1}]] + \mathbb{E}[\alpha_{k+1} \mathbb{I}[w_k \leq \alpha_{k+1}]] \\ &= \mathbb{E}[w_k \mathbb{I}[w_k > \alpha_{k+1}]] + \alpha_{k+1} \mathbb{P}[w_k \leq \alpha_{k+1}], \end{aligned} \tag{8}$$

ce qui nous permet de calculer les  $\alpha_k$  par récurrence.

On a  $\alpha_N = 0$ ,  $\alpha_{N-1} = \mathbb{E}[w_{N-1}]/(1+r)$ , etc.

Peut-on caractériser davantage cette politique? Intuitivement, moins il reste de temps, moins on devrait être exigeant, car il nous reste moins d'opportunités; donc  $\alpha_k$  devrait diminuer avec  $k$ .

**Proposition.** Si les  $w_k$  sont i.i.d., alors  $\alpha_k \geq \alpha_{k+1}$  pour tout  $k$ .



**Proposition.** Si les  $w_k$  sont i.i.d., alors  $\alpha_k \geq \alpha_{k+1}$  pour tout  $k$ .

**Preuve.** On note par  $w$  une v.a. qui a la même loi que les  $w_k$ .

Il suffit de montrer que  $V_k(x) \geq V_{k+1}(x)$  pour  $x \geq 0$  et  $1 \leq k \leq N - 1$ .

Cela se fait aisément par **induction sur  $k$** .

Pour  $k = N - 1$ , on a  $V_{N-1}(x) \geq x = V_N(x)$ .

Si on suppose que  $V_{k+1}(x) \geq V_{k+2}(x)$  pour tout  $x$ , alors

$$\begin{aligned} V_k(x) &= \max(x, \mathbb{E}[V_{k+1}(w)]/(1+r)) \\ &\geq \max(x, \mathbb{E}[V_{k+2}(w)]/(1+r)) \\ &= V_{k+1}(x). \end{aligned}$$

Que se passe-t-il lorsque  $N \rightarrow \infty$ ?

Ou de façon équivalente, lorsque  $k \rightarrow -\infty$ ?

Si on peut borner la suite des  $\alpha_k$ , cela montrera que cette suite converge lorsque  $k \rightarrow -\infty$ . Par (8),

$$(1+r)\alpha_k \leq \mathbb{E}[w] + \alpha_{k+1}.$$

D'où l'on tire

$$\begin{aligned} \alpha_k &\leq \frac{\mathbb{E}[w] + \alpha_{k+1}}{1+r} \\ &\leq \frac{\mathbb{E}[w]}{1+r} + \frac{\mathbb{E}[w] + \alpha_{k+2}}{(1+r)^2} \\ &\leq \dots \\ &\leq \mathbb{E}[w] \sum_{i=1}^{\infty} \frac{1}{(1+r)^i} \\ &= \mathbb{E}[w] \frac{1+r}{r} < \infty. \end{aligned}$$

en supposant  $r > 0$ .

Ainsi, lorsque  $k \rightarrow -\infty$ ,  $\alpha_k \rightarrow \bar{\alpha}$  pour une constante  $\bar{\alpha}$  qui satisfait

$$\begin{aligned}(1+r)\bar{\alpha} &= \mathbb{E}[\max(w, \bar{\alpha})] \\ &= \mathbb{E}[w\mathbb{I}[w > \bar{\alpha}]] + \bar{\alpha}\mathbb{P}[w \leq \bar{\alpha}].\end{aligned}$$

Lorsque l'horizon tend vers l'infini, la politique optimale est une politique stationnaire déterminée par le seuil  $\bar{\alpha}$ .

Ainsi, lorsque  $k \rightarrow -\infty$ ,  $\alpha_k \rightarrow \bar{\alpha}$  pour une constante  $\bar{\alpha}$  qui satisfait

$$\begin{aligned}(1+r)\bar{\alpha} &= \mathbb{E}[\max(w, \bar{\alpha})] \\ &= \mathbb{E}[w\mathbb{I}[w > \bar{\alpha}]] + \bar{\alpha}\mathbb{P}[w \leq \bar{\alpha}].\end{aligned}$$

Lorsque l'horizon tend vers l'infini, la politique optimale est une politique stationnaire déterminée par le seuil  $\bar{\alpha}$ .

Dans ce modèle,  $r > 0$  fait qu'il devient plus attrayant de vendre plus tôt, à prix égal. Et si on avait  $r = 0$  pour le modèle sur horizon infini? Dans ce cas, à moins que l'on atteigne la valeur maximale que  $w$  peut prendre (ce qui est impossible pour plusieurs lois de prob.), on attendra toujours indéfiniment...

## Achat avant une date limite.

Au lieu de vendre un actif, on veut (ou on doit) l'acheter au plus tard au début de la période  $N$ . L'état  $x_k$  à l'étape  $k$  est  $\Delta$  si on a déjà acheté, et  $x_k = w_{k-1}$  si on n'a pas encore acheté, où  $w_{k-1}$  est le prix du marché pour le produit dont on a besoin, au début de la période  $k$ .

Ici on minimise au lieu de maximiser.

Le coût espéré total optimal à partir du temps  $k$ , actualisé au temps  $k$ , si on est dans l'état  $x_k \neq \Delta$ , est

$$\begin{aligned} V_k(x_k) &= \begin{cases} x_N & \text{si } k = N; \\ \min(x_k, \mathbb{E}[V_{k+1}(w_k)]/(1+r)) & \text{si } k < N \end{cases} \\ &= \min(x_k, \alpha_k) \end{aligned}$$

où  $\alpha_N = \infty$ ,  $\alpha_{N-1} = \mathbb{E}[w_{N-1}]/(1+r)$ , etc.

On a

$$(1+r)\alpha_k = \mathbb{E}[V_{k+1}(w_k)] \quad (9)$$

$$\begin{aligned} &= \mathbb{E}[\min(w_k, \alpha_{k+1})] \\ &= \mathbb{E}[w_k \mathbb{I}[w_k \leq \alpha_{k+1}]] + \alpha_{k+1} \mathbb{P}[w_k > \alpha_{k+1}], \end{aligned} \quad (10)$$

ce qui nous permet de calculer les  $\alpha_k$  par récurrence.

La **politique optimale** est ainsi déterminée par des **seuils**  $\alpha_k$  comme pour le problème du vendeur: on **achète si et seulement si**  $x_k \leq \alpha_k$ .

On peut montrer, de la même façon:

**Proposition.** Si les  $w_k$  sont i.i.d., alors  $\alpha_k \leq \alpha_{k+1}$  pour tout  $k$ .

## Prix corrélés.

On peut généraliser le modèle au cas où les  $w_k$  sont dépendants. Supposons, par exemple, que  $x_0 = 0$  et

$$x_{k+1} = w_k = \lambda x_k + \xi_k, \quad 0 \leq k \leq N - 1,$$

où  $\lambda \in [0, 1)$  est une constante et les  $\xi_k$  sont des v.a. i.i.d. à valeur dans  $[0, \infty)$ , avec  $\bar{\xi} = \mathbb{E}[\xi_k] > 0$ . Pour simplifier, supposons que  $r = 0$ . On a alors, pour le problème de l'acheteur,

$$J_k(x_k) = \begin{cases} x_N & \text{si } k = N \text{ et } x_N \neq \Delta; \\ \min(x_k, \mathbb{E}[J_{k+1}(\lambda x_k + \xi_k)]) & \text{si } k < N \text{ et } x_k \neq \Delta. \end{cases}$$

**Proposition.** Pour  $k \leq N - 1$ ,  $J_k(x)$  est croissante et concave en  $x$  et on a  $J_k(x) \leq J_{k+1}(x)$  pour tout  $x$ . La décision optimale à l'étape  $k$  est d'acheter si  $x_k \leq \alpha_k$  et d'attendre sinon, où  $\alpha_k$  est l'unique valeur positive qui satisfait

$$\alpha_k = \mathbb{E}[J_{k+1}(\lambda\alpha_k + \xi_k)].$$

De plus, on a  $\alpha_k \leq \alpha_{k+1}$ .

**Preuve.** Par induction sur  $k$ .

Pour  $k = N - 1$ , on a (faire graphique au tableau)

$$J_{N-1}(x) = \min(x, \lambda x + \bar{\xi}) \leq x = J_N(x)$$

et le résultat tient avec  $\alpha_{N-1} = \lambda\alpha_{N-1} + \bar{\xi}$ , i.e.,  $\alpha_{N-1} = \bar{\xi}/(1 - \lambda)$ .

Supposons que le résultat tient pour  $k + 1$  et montrons qu'il tient alors aussi pour  $k$ . On a

$$\begin{aligned} J_k(x) &= \min(x, \mathbb{E}[J_{k+1}(\lambda x + \xi_k)]) \\ &\leq \min(x, \mathbb{E}[J_{k+2}(\lambda x + \xi_{k+1})]) \\ &= J_{k+1}(x) \end{aligned}$$

et  $J_k(x)$  est croissante et concave en  $x$  car  $J_{k+1}(x)$  l'est.

Supposons que le résultat tient pour  $k + 1$  et montrons qu'il tient alors aussi pour  $k$ . On a

$$\begin{aligned} J_k(x) &= \min(x, \mathbb{E}[J_{k+1}(\lambda x + \xi_k)]) \\ &\leq \min(x, \mathbb{E}[J_{k+2}(\lambda x + \xi_{k+1})]) \\ &= J_{k+1}(x) \end{aligned}$$

et  $J_k(x)$  est croissante et concave en  $x$  car  $J_{k+1}(x)$  l'est.

Par ailleurs, pour  $x = 0$ ,  $\mathbb{E}[J_{k+1}(\xi_k)] > 0$ , donc  $J_k(x) = x$  pour  $x$  proche de 0. Et puisque

$$\alpha_{k+1} = \mathbb{E}[J_{k+2}(\lambda \alpha_{k+1} + \xi_{k+1})] \geq \mathbb{E}[J_{k+1}(\lambda \alpha_{k+1} + \xi_k)],$$

pour  $x \geq \alpha_{k+1}$  on a  $J_k(x) = \varphi(x) \stackrel{\text{def}}{=} \mathbb{E}[J_{k+1}(\lambda x + \xi_k)]$ .

Supposons que le résultat tient pour  $k + 1$  et montrons qu'il tient alors aussi pour  $k$ . On a

$$\begin{aligned} J_k(x) &= \min(x, \mathbb{E}[J_{k+1}(\lambda x + \xi_k)]) \\ &\leq \min(x, \mathbb{E}[J_{k+2}(\lambda x + \xi_{k+1})]) \\ &= J_{k+1}(x) \end{aligned}$$

et  $J_k(x)$  est croissante et concave en  $x$  car  $J_{k+1}(x)$  l'est.

Par ailleurs, pour  $x = 0$ ,  $\mathbb{E}[J_{k+1}(\xi_k)] > 0$ , donc  $J_k(x) = x$  pour  $x$  proche de 0. Et puisque

$$\alpha_{k+1} = \mathbb{E}[J_{k+2}(\lambda \alpha_{k+1} + \xi_{k+1})] \geq \mathbb{E}[J_{k+1}(\lambda \alpha_{k+1} + \xi_k)],$$

pour  $x \geq \alpha_{k+1}$  on a  $J_k(x) = \varphi(x) \stackrel{\text{def}}{=} \mathbb{E}[J_{k+1}(\lambda x + \xi_k)]$ .

Puisque  $J_{k+1}(x)$  est concave et  $J'_{k+1}(\cdot) \leq 1$ , on a  $\varphi'(x) \leq \lambda < 1$ . La valeur de  $x$  où  $\varphi(x) = x$  est donc unique et doit se trouver dans l'intervalle  $(0, \alpha_{k+1})$ , puisque  $\alpha_{k+1} \geq \varphi(\alpha_{k+1})$ . C'est  $x = \alpha_k$ . Cela complète la preuve.  $\square$

## Option de type américaine pour une action dont le prix suit une marche aléatoire.

(Ross 1983, Section 1.3.)

On suppose que le **prix de l'action** de la firme ABC est  $x_k$  au jour  $k$ , et évolue selon une marche aléatoire:

$$x_{k+1} = x_k + w_k = x_0 + \sum_{i=0}^k w_i$$

où les  $w_i$  sont des v.a. i.i.d. de moyenne  $\mu$ .

On a une **option d'achat** pour une action au prix fixe  $K$ .

On peut l'exercer à l'un des  $N$  premiers jours.

Si on l'exerce au jour  $k$ , notre **profit** est  $\max(0, x_k - K)$ .

On veut maximiser notre profit espéré.

Différence avec l'exemple précédent: ici on n'est pas obligé d'acheter à la fin.

Soit  $J_k(x_k)$  le profit espéré optimal si le prix est  $x_k$  au jour  $k$  et que l'on n'a pas encore exercé l'option. On a

$$J_k(x_k) = \begin{cases} \max(0, x_N - K) & \text{si } k = N; \\ \max(x_k - K, \mathbb{E}[J_{k+1}(x_k + w_k)]) & \text{si } k < N. \end{cases}$$

**Proposition.**  $J_k(x)$  est continue et croissante en  $x$  et décroissante en  $k$ , tandis que  $J_k(x) - x$  est décroissante en  $x$ .

La **politique optimale** consiste à exercer l'option au jour  $k$  si et seulement si  $x_k \geq \alpha_k$ , où  $\alpha_k$  est la plus petite valeur qui satisfait  $J_k(\alpha_k) = \alpha_k - K$ . De plus,  $\alpha_{k+1} \leq \alpha_k$  pour  $k < N$ .

Soit  $J_k(x_k)$  le profit espéré optimal si le prix est  $x_k$  au jour  $k$  et que l'on n'a pas encore exercé l'option. On a

$$J_k(x_k) = \begin{cases} \max(0, x_N - K) & \text{si } k = N; \\ \max(x_k - K, \mathbb{E}[J_{k+1}(x_k + w_k)]) & \text{si } k < N. \end{cases}$$

**Proposition.**  $J_k(x)$  est continue et croissante en  $x$  et décroissante en  $k$ , tandis que  $J_k(x) - x$  est décroissante en  $x$ .

La **politique optimale** consiste à exercer l'option au jour  $k$  si et seulement si  $x_k \geq \alpha_k$ , où  $\alpha_k$  est la plus petite valeur qui satisfait  $J_k(\alpha_k) = \alpha_k - K$ . De plus,  $\alpha_{k+1} \leq \alpha_k$  pour  $k < N$ .

**Preuve.** On montre la première partie par induction sur  $k$ .

$J_N(x) = \max(0, x - K)$  est continue et croissante en  $x$ , et  
 $J_N(x) - x = \max(-x, -K)$  est décroissante en  $x$ .

$J_N(x) = \max(0, x - K)$  est continue et croissante en  $x$ , et  $J_N(x) - x = \max(-x, -K)$  est décroissante en  $x$ . De plus,

$$\begin{aligned} J_{N-1}(x) &= \max(x - K, \mathbb{E}[J_N(x + w_{N-1})]) \\ &= \max(x - K, \mathbb{E}[\max(0, x + w_{N-1} - K)]) \\ &\geq \max(x - K, 0, \mathbb{E}[x + w_{N-1} - K]) \geq J_N(x). \end{aligned}$$

$J_N(x) = \max(0, x - K)$  est continue et croissante en  $x$ , et  $J_N(x) - x = \max(-x, -K)$  est décroissante en  $x$ . De plus,

$$\begin{aligned} J_{N-1}(x) &= \max(x - K, \mathbb{E}[J_N(x + w_{N-1})]) \\ &= \max(x - K, \mathbb{E}[\max(0, x + w_{N-1} - K)]) \\ &\geq \max(x - K, 0, \mathbb{E}[x + w_{N-1} - K]) \geq J_N(x). \end{aligned}$$

Si on suppose que  $J_{k+1}$  a les propriétés désirées, pour  $k < N$ ,

$$J_k(x) = \max(x - K, \mathbb{E}[J_{k+1}(x + w_k)])$$

est continue et croissante en  $x$  car  $J_{k+1}$  l'est,

$$\begin{aligned} J_k(x) &\geq \max(x - K, \mathbb{E}[J_{k+2}(x + w_k)]) = J_{k+1}(x), \text{ et} \\ J_k(x) - x &= \max(-K, \mathbb{E}[J_{k+1}(x + w_k) - (x + w_k)] + \mathbb{E}[w_k]) \end{aligned}$$

est décroissante en  $x$  grâce à l'hypothèse sur  $J_{k+1}$ .

Il est optimal d'exercer l'option à l'étape  $k$  si et seulement si  $J_k(x_k) = x_k - K$ . Posons

$$\alpha_k = \inf\{x : J_k(x) - x = -K\}.$$

( $\alpha_k = \infty$  si  $J_k(x) - x$  n'atteint jamais  $-K$ .)

Pour  $x \geq \alpha_k$ , on a  $J_k(x) - x \leq J_k(\alpha_k) - \alpha_k = -K$  et il est donc optimal d'exercer. (C'est pour montrer ceci que l'on a montré que  $J_k(x) - x$  est décroissant.)

La politique optimale a donc la forme spécifiée.

Le fait que  $\alpha_{k+1} \leq \alpha_k$  découle du fait que  $J_{k+1}(x) \leq J_k(x)$ .  $\square$

## Ensemble d'arrêt absorbant et règle du "un coup à l'avance".

On considère un modèle **stationnaire** général où en plus (ou à l'intérieur) de la décision  $u_k$  à l'étape  $k$ , on peut arrêter le système et payer un coût terminal  $t(x_k)$  fonction de l'état  $x_k$ , ou bien continuer. Si on se rend à l'étape  $N$ , on doit payer  $t(x_N)$ . On a  $J_N(x_N) = t(x_N)$  et

$$J_k(x_k) = \min \left( t(x_k), \min_{u \in U(x_k)} \mathbb{E}_{w_k} [g(x_k, u, w_k) + J_{k+1}(f(x_k, u, w_k))] \right),$$

pour  $k < N$ , et il est **optimal de s'arrêter** au temps  $k$  si et seulement si  $x_k$  est dans

$$T_k = \left\{ x : t(x) \leq \min_{u \in U(x)} \mathbb{E} [g(x, u, w) + J_{k+1}(f(x, u, w))] \right\}.$$

On voit que  $J_{N-1}(x) \leq J_N(x)$  et on montre facilement par induction que  $J_k(x) \leq J_{k+1}(x)$ , pour tout  $x$  et  $k < N$ . Il en découle que pour  $0 \leq k < N$ ,  $x \in T_k \Rightarrow x \in T_{k+1}$ , et donc  $T_k \subseteq T_{k+1}$ .

**Proposition.** Si l'ensemble  $T_{N-1}$  est absorbant tant que l'on ne s'arrête pas, i.e.,  $\{x \in T_{N-1}\} \Rightarrow \{f(x, u, w) \in T_{N-1}\}$ , alors  $T_k = T_{N-1}$  pour tout  $k < N$ .

**Preuve.** Si  $x_{N-2} = x \in T_{N-1}$ , alors  $x_{N-1} = f(x, u, w) \in T_{N-1}$ , de sorte que  $J_{N-1}(x_{N-1}) = t(x_{N-1})$ , et donc

$$\begin{aligned} t(x) &\leq \min_{u \in U(x)} \mathbb{E} [g(x, u, w) + J_N(f(x, u, w))] \\ &= \min_{u \in U(x)} \mathbb{E} [g(x, u, w) + t(f(x, u, w))] \\ &= \min_{u \in U(x)} \mathbb{E} [g(x, u, w) + J_{N-1}(f(x, u, w))], \end{aligned}$$

ce qui veut dire que  $x \in T_{N-2}$ . On a donc  $T_{N-1} \subseteq T_{N-2}$ , et donc  $T_{N-2} = T_{N-1}$ . On montre de la même manière que  $T_k = T_{N-1}$  pour tout  $k$ . (Ou encore, par induction sur  $k$ .)  $\square$

Ainsi, si  $T_{N-1}$  est absorbant, l'ensemble d'arrêt optimal est le même à toutes les étapes. Il suffit donc de le calculer à l'étape  $N - 1$ .

En d'autres mots, la politique optimale à chaque étape est de s'arrêter si et seulement s'il est préférable de s'arrêter maintenant plutôt que de continuer et de s'arrêter obligatoirement à la prochaine étape. Autrement dit, **il suffit de regarder un coup à l'avance** (“one-step look ahead policy”).

Ainsi, si  $T_{N-1}$  est absorbant, l'ensemble d'arrêt optimal est le même à toutes les étapes. Il suffit donc de le calculer à l'étape  $N - 1$ .

En d'autres mots, la politique optimale à chaque étape est de s'arrêter si et seulement s'il est préférable de s'arrêter maintenant plutôt que de continuer et de s'arrêter obligatoirement à la prochaine étape. Autrement dit, **il suffit de regarder un coup à l'avance** (“one-step look ahead policy”).

Cette proposition s'applique dans les deux exemples qui suivent.

## Exemple: Problème du vendeur avec retention des offres.

Revenons au problème de vente d'un actif. On suppose maintenant que les offres qui ne sont pas acceptées immédiatement ne sont pas rejetées, elles peuvent être acceptées plus tard.

L'état  $x_k$  indique l'offre courante et on a

$$x_{k+1} = \max(x_k, w_k).$$

## Exemple: Problème du vendeur avec retention des offres.

Revenons au problème de vente d'un actif. On suppose maintenant que les offres qui ne sont pas acceptées immédiatement ne sont pas rejetées, elles peuvent être acceptées plus tard.

L'état  $x_k$  indique l'offre courante et on a

$$x_{k+1} = \max(x_k, w_k).$$

Question: Que fera-t-on si (le taux d'intérêt)  $r = 0$ ?

## Exemple: Problème du vendeur avec retention des offres.

Revenons au problème de vente d'un actif. On suppose maintenant que les offres qui ne sont pas acceptées immédiatement ne sont pas rejetées, elles peuvent être acceptées plus tard.

L'état  $x_k$  indique l'offre courante et on a

$$x_{k+1} = \max(x_k, w_k).$$

Question: Que fera-t-on si (le taux d'intérêt)  $r = 0$ ?

Les récurrences pour  $V_k$  deviennent

$$V_N(x_N) = x_N,$$

$$V_k(x_k) = \max(x_k, \mathbb{E}[V_{k+1}(\max(x_k, w_k))]/(1+r)), \quad k < N.$$

L'ensemble d'arrêt optimal à l'étape  $N - 1$  est

$$T_{N-1} = \{x : x \geq \mathbb{E}[\max(x, w)] / (1 + r)\} = \{x : x \geq \bar{\alpha}\},$$

où  $\bar{\alpha}$  est la solution de

$$(1 + r)\bar{\alpha} = \mathbb{E}[\max(\bar{\alpha}, w)] = \bar{\alpha}\mathbb{P}[w \leq \bar{\alpha}] + \mathbb{E}[w\mathbb{I}[w > \bar{\alpha}]]$$

si la solution existe,  $\bar{\alpha} = \infty$  sinon.

Cet ensemble  $T_{N-1}$  est **absorbant**, car  $x_k$  ne peut jamais diminuer lorsque  $k$  augmente. On a donc  $T_k = T_{N-1}$  pour tout  $k$ , i.e., on **accepte la première offre qui atteint  $\bar{\alpha}$** .

## Exemple: Un voleur qui sait calculer.

À chaque période  $k$  (e.g., chaque nuit), un voleur peut tenter un nouveau vol ou prendre sa retraite avec son profit déjà accumulé,  $x_k$ .

S'il tente un vol, avec probabilité  $p$  il se fait prendre et perd tout, et avec probabilité  $1 - p$  il fait un gain  $w_k$  (aléatoire).

Après  $N$  périodes, il doit nécessairement se retirer avec son profit  $x_N$ , s'il ne l'a pas fait avant et s'il n'a pas été pris.

Il veut maximiser son profit espéré total,  $\mathbb{E}[x_N]$ .

Les équations de la PD:

$$J_N(x_N) = x_N,$$

$$J_k(x_k) = \max(x_k, (1 - p)\mathbb{E}[J_{k+1}(x_k + w_k)]), \quad k < N.$$

On a ici

$$\begin{aligned}
 T_{N-1} &= \{x : x \geq (1 - \rho)(x + \mathbb{E}[w])\} \cup \{\Delta\} \\
 &= \{x : x \geq \mathbb{E}[w](1 - \rho)/\rho\} \cup \{\Delta\}
 \end{aligned}$$

où  $\Delta$  est l'état d'arrestation. Cet ensemble est **absorbant** au sens de la proposition. La politique optimale est donc de se retirer dès que le profit accumulé  $x_k$  atteint  $\bar{\alpha} = \mathbb{E}[w](1 - \rho)/\rho$ , peu importe la valeur de  $k$ .

Ce modèle s'applique aussi à d'autres types de situations.

## Problèmes d'ordonnancement et argument d'échange de voisins.

On a un ensemble de  $N$  tâches à accomplir et on doit les ordonner. Par exemple,  $N$  pièces à fabriquer dans un atelier, où  $N$  voitures à réparer, ou  $N$  dossiers à étudier, ou  $N$  articles à arbitrer, etc. On veut minimiser un critère de performance qui s'exprime comme l'espérance de la somme des coûts pour les différentes tâches.

Algorithme de PD standard: l'état à l'étape  $k$  est l'ensemble des tâches restantes. Il y a  $2^N$  états en tout.

## Problèmes d'ordonnancement et argument d'échange de voisins.

On a un ensemble de  $N$  tâches à accomplir et on doit les ordonner. Par exemple,  $N$  pièces à fabriquer dans un atelier, où  $N$  voitures à réparer, ou  $N$  dossiers à étudier, ou  $N$  articles à arbitrer, etc. On veut minimiser un critère de performance qui s'exprime comme l'espérance de la somme des coûts pour les différentes tâches.

Algorithme de PD standard: l'état à l'étape  $k$  est l'ensemble des tâches restantes. Il y a  $2^N$  états en tout.

Les modèles considérés ici sont stochastiques, mais l'information obtenue au cours des premières étapes n'est pas utile pour améliorer les décisions futures, de sorte que la politique optimale sera une politique en **boucle ouverte** (on peut trouver l'ordonnancement optimal dès le départ).

L'argument d'échange de voisins dit que si

$$L = \{i_0, i_1, \dots, i_{k-1}, i, j, i_{k+2}, \dots, i_{N-1}\}$$

est un ordonnancement optimal et

$$L' = \{i_0, i_1, \dots, i_{k-1}, j, i, i_{k+2}, \dots, i_{N-1}\}$$

alors le coût espéré total pour  $L$  ne doit pas dépasser celui pour  $L'$ .

En général, cela ne donne que des conditions **nécessaires** d'optimalité, mais dans certains cas il devient évident que ces conditions sont aussi suffisantes.

## Exemple: Ordonnancement des questions d'un quiz.

Il y a  $N$  questions, auxquelles on peut répondre dans l'ordre que l'on veut. On répondra correctement à la question  $i$  avec probabilité  $p_i$ , et si on le fait on gagne  $R_i$ . Dès que l'on échoue à une question, c'est terminé. On veut maximiser notre gain total espéré.

À noter que la politique optimale est en boucle ouverte, car une fois que l'on aura répondu aux  $k$  premières questions, on n'aura pas davantage d'information qu'au départ qui puisse justifier de changer notre ordonnancement des  $N - k$  questions restantes.

Soit  $J(S)$  le revenu espéré pour une suite ordonnée de questions  $S$ , et soient  $L$  et  $L'$  comme à la page précédente, où  $L$  est optimal:

$$L = \{i_0, i_1, \dots, i_{k-1}, i, j, i_{k+2}, \dots, i_{N-1}\},$$

$$L' = \{i_0, i_1, \dots, i_{k-1}, j, i, i_{k+2}, \dots, i_{N-1}\},$$

$$J(L) = J(\{i_0, \dots, i_{k-1}\}) + p_{i_0} \cdots p_{i_{k-1}} (p_i R_i + p_i p_j R_j) \\ + p_{i_0} \cdots p_{i_{k-1}} p_i p_j J(\{i_{k+2}, \dots, i_{N-1}\});$$

$$J(L') = J(\{i_0, \dots, i_{k-1}\}) + p_{i_0} \cdots p_{i_{k-1}} (p_j R_j + p_j p_i R_i) \\ + p_{i_0} \cdots p_{i_{k-1}} p_j p_i J(\{i_{k+2}, \dots, i_{N-1}\}).$$

Puisque  $L$  est optimale, on doit avoir  $J(L) \geq J(L')$ , i.e.,

$$p_i R_i + p_i p_j R_j \geq p_j R_j + p_j p_i R_i, \quad \text{i.e.,}$$

$$p_i R_i (1 - p_j) \geq p_j R_j (1 - p_i), \quad \text{i.e.,}$$

$$p_i R_i / (1 - p_j) \geq p_j R_j / (1 - p_i).$$

Il faut ordonner les tâches par ordre décroissant de  $p_i R_i / (1 - p_i)$ .

## Ordonnement sur un processeur, avec revenu actualisé.

$N$  tâches. La tâche  $i$  requiert un temps  $T_i$  aléatoire de loi connue et rapporte un revenu  $\alpha^t R_i$  si on la complète au temps  $t$ , où  $\alpha < 1$  est un facteur d'actualisation. Les v.a.  $T_i$  sont supposées indépendantes.

Les valeurs des  $T_i$  des tâches déjà réalisées n'affectent les revenus futurs que via un facteur d'actualisation commun, donc n'affectent pas l'optimisation de l'ordonnement des tâches qui restent. On peut ainsi déterminer l'ordonnement optimal dès le départ.

Soient  $L$  et  $L'$  comme dans l'exemple précédent, et soit  $t_k$  l'instant où on termine la tâche  $k - 1$  et débute la suivante.

Puisque les  $T_i$  sont indépendants, on a

$$\begin{aligned}
 J(L) &= J(\{i_0, \dots, i_{k-1}\}) + \mathbb{E}[\alpha^{t_k+T_i} R_i + \alpha^{t_k+T_i+T_j} R_j] \\
 &\quad + \mathbb{E}[\alpha^{t_k+T_i+T_j} J(\{i_{k+2}, \dots, i_{N-1}\})] \\
 &= J(\{i_0, \dots, i_{k-1}\}) + \mathbb{E}[\alpha^{t_k}] \mathbb{E}[\alpha^{T_i} R_i + \alpha^{T_i+T_j} R_j] \\
 &\quad + \mathbb{E}[\alpha^{t_k+T_i+T_j}] J(\{i_{k+2}, \dots, i_{N-1}\}).
 \end{aligned}$$

Puisque les  $T_i$  sont indépendants, on a

$$\begin{aligned}
 J(L) &= J(\{i_0, \dots, i_{k-1}\}) + \mathbb{E}[\alpha^{t_k+T_i} R_i + \alpha^{t_k+T_i+T_j} R_j] \\
 &\quad + \mathbb{E}[\alpha^{t_k+T_i+T_j} J(\{i_{k+2}, \dots, i_{N-1}\})] \\
 &= J(\{i_0, \dots, i_{k-1}\}) + \mathbb{E}[\alpha^{t_k}] \mathbb{E}[\alpha^{T_i} R_i + \alpha^{T_i+T_j} R_j] \\
 &\quad + \mathbb{E}[\alpha^{t_k+T_i+T_j}] J(\{i_{k+2}, \dots, i_{N-1}\}).
 \end{aligned}$$

Tout comme dans l'exemple précédent,  $J(L) \geq J(L')$  implique que

$$\begin{aligned}
 \mathbb{E}[\alpha^{T_i} R_i + \alpha^{T_i+T_j} R_j] &\geq \mathbb{E}[\alpha^{T_j} R_j + \alpha^{T_j+T_i} R_i], \quad \text{i.e.,} \\
 R_i \mathbb{E}[\alpha^{T_i}] + R_j \mathbb{E}[\alpha^{T_i}] \mathbb{E}[\alpha^{T_j}] &\geq R_j \mathbb{E}[\alpha^{T_j}] + R_i \mathbb{E}[\alpha^{T_j}] \mathbb{E}[\alpha^{T_i}] \quad \text{i.e.,} \\
 \frac{p_i R_i}{1 - p_i} &\geq \frac{p_j R_j}{1 - p_j}
 \end{aligned}$$

où  $p_k = \mathbb{E}[\alpha^{T_k}]$ . On doit ordonner les tâches par ordre décroissant de leur valeur de  $p_i R_i / (1 - p_i)$ , exactement comme dans l'exemple précédent.

## Ordonnement sur 2 processeurs en série.

$N$  tâches. La tâche  $i$  requiert un temps  $a_i$  sur la machine  $A$ , puis un temps  $b_i$  sur la machine  $B$  (déterministe). On veut minimiser le temps total de traitement.

## Ordonnancement sur 2 processeurs en série.

$N$  tâches. La tâche  $i$  requiert un temps  $a_i$  sur la machine  $A$ , puis un temps  $b_i$  sur la machine  $B$  (déterministe). On veut minimiser le temps total de traitement.

Soit  $X_k$  l'ensemble des tâches qui restent à traiter sur  $A$  et  $\tau_k$  la quantité de travail en attente ("workload") sur la machine  $B$ , à l'instant où l'on termine la  $k$ -ième tâche sur la machine  $A$  (étape  $k$ ).

L'état à l'étape  $k$  est  $(X_k, \tau_k)$ .

## Ordonnement sur 2 processeurs en série.

$N$  tâches. La tâche  $i$  requiert un temps  $a_i$  sur la machine  $A$ , puis un temps  $b_i$  sur la machine  $B$  (déterministe). On veut minimiser le temps total de traitement.

Soit  $X_k$  l'ensemble des tâches qui restent à traiter sur  $A$  et  $\tau_k$  la quantité de travail en attente ("workload") sur la machine  $B$ , à l'instant où l'on termine la  $k$ -ième tâche sur la machine  $A$  (étape  $k$ ).

L'état à l'étape  $k$  est  $(X_k, \tau_k)$ .

Si  $i$  est la prochaine tâche sur  $A$ , on aura à l'étape  $k + 1$ :

$$X_{k+1} = X_k - \{i\}, \quad \tau_{k+1} = \max(0, \tau_k - a_i) + b_i.$$

Lorsqu'on termine la dernière tâche sur  $A$ , il reste un temps  $\tau_N$  avant de tout terminer. La récurrence de la PD s'écrit donc

$$\begin{aligned} J_N(\phi, \tau_N) &= \tau_N, \\ J_k(X_k, \tau_k) &= \min_{i \in X_k} [a_i + J_{k+1}(X_k - \{i\}, \max(0, \tau_k - a_i) + b_i)] \end{aligned}$$

pour  $k < N$ .

Soient  $L$  et  $L'$  comme dans les exemples précédents. L'argument d'échange des voisins  $i$  et  $j$  nous donne la **condition d'optimalité**

$$J_{k+2}(X_k - \{i, j\}, \tau_{i,j}) \leq J_{k+2}(X_k - \{i, j\}, \tau_{j,i})$$

où  $\tau_{ij}$  et  $\tau_{ji}$  sont les valeurs de  $\tau_{k+2}$  pour  $L$  et  $L'$ , respectivement. Ceci implique que  $\tau_{ij} \leq \tau_{ji}$  car  $J_k(X_k, \tau_k)$  est croissante en  $\tau_k$ .

Pour  $L$ , on a

$$\begin{aligned} \tau_{ij} &= \tau_{k+2} = \max(0, \tau_{k+1} - a_j) + b_j \\ &= \max(0, \max(0, \tau_k - a_i) + b_i - a_j) + b_j \\ &= \max(0, \max(a_i, \tau_k) + b_i - a_j - a_i) + b_j \\ &= \max(a_i + a_j - b_i, \max(a_i, \tau_k)) - a_j - a_i + b_i + b_j \\ &= \max(\tau_k, a_i, a_i + a_j - b_i) - a_i - a_j + b_i + b_j. \end{aligned}$$

De même,

$$\tau_{ji} = \max(\tau_k, a_j, a_i + a_j - b_j) - a_i - a_j + b_i + b_j.$$

Si  $\tau_k \geq \max(a_j, a_i + a_j - b_j)$ , alors  $\tau_k = \tau_{ji}$ . Mais on sait que  $\tau_{ji} \geq \tau_{ij}$ , donc  $\tau_{ij} = \tau_k = \tau_{ji}$ , ce qui veut dire que  $L$  et  $L'$  ont le même coût.

Si  $\tau_k < \max(a_j, a_i + a_j - b_j)$ , alors on doit avoir  $\max(a_i, a_i + a_j - b_i) \leq \tau_{ji} = \max(a_j, a_i + a_j - b_j)$ , et on peut montrer que ceci est équivalent à

$$\min(a_i, b_j) \leq \min(a_j, b_i).$$

De même,

$$\tau_{ji} = \max(\tau_k, a_j, a_i + a_j - b_j) - a_i - a_j + b_i + b_j.$$

Si  $\tau_k \geq \max(a_j, a_i + a_j - b_j)$ , alors  $\tau_k = \tau_{ji}$ . Mais on sait que  $\tau_{ji} \geq \tau_{ij}$ , donc  $\tau_{ij} = \tau_k = \tau_{ji}$ , ce qui veut dire que  $L$  et  $L'$  ont le même coût.

Si  $\tau_k < \max(a_j, a_i + a_j - b_j)$ , alors on doit avoir  $\max(a_i, a_i + a_j - b_i) \leq \tau_{ji} = \max(a_j, a_i + a_j - b_j)$ , et on peut montrer que ceci est équivalent à

$$\min(a_i, b_j) \leq \min(a_j, b_i).$$

Cela détermine un ordre unique pour chaque paire  $(i, j)$ , sauf s'il y a égalité ci-haut. Dans le cas où il y a égalité, l'ordre n'a pas d'importance. Mis à part ces cas d'égalité, l'ordre unique pour chaque paire détermine un ordre unique pour l'ensemble des tâches.

**Intuition:** Il faut éviter de laisser la machine B innocupée, car c'est là que le temps est perdu. Pour cela, on donne priorité aux tâches qui demandent très peu de temps sur la machine A et/ou beaucoup de temps sur la machine B.

**Algorithme d'ordonnement:**

$T \leftarrow \{1, \dots, N\}; \quad L1 \leftarrow \phi; \quad L2 \leftarrow \phi;$

TANTQUE  $T \neq \phi$  FAIRE

$k \leftarrow \arg \min_{i \in T} \min(a_i, b_i);$

    Si  $a_k < b_k$ , mettre  $k$  à la fin de  $L1$

    SINON mettre  $k$  au début de  $L2$ ;

L'ordonnement choisi est la liste  $L1$  suivie de la liste  $L2$ .

**Proposition.** Cet algorithme retourne un ordonnancement optimal.

**Preuve:** Exercice.

## Ordonnement des pages web par un engin de recherche.

Voir: “Revenue-Maximizing Rankings for Online Platforms with Quality-Sensitive Consumers”, par L’Ecuyer, Maillé, Stier-Moses, and Tuffin, <http://www.iro.umontreal.ca/~lecuyer/papers.html>.

Un engin de recherche sur Internet (EBay, Google, etc.) reçoit des requêtes par mots clés. Elles arrivent selon un processus de Poisson de taux  $\lambda$ .

On représente une requête par  $Y = (M, R_1, G_1, \alpha_1, \dots, R_M, G_M, \alpha_M)$  où

$M$  est le nombre de pages pertinentes,

$R_i$  est une mesure de pertinence (estimée) de la page  $i$ ,

$G_i$  est le revenu espéré (estimé) par clic pour cette page, et

$\alpha_i$  est la fraction du revenu perçu par l’engin, pour cette page.

Pour chaque  $Y$ , l’engin doit choisir une permutation

$\pi = (\pi(1), \dots, \pi(M)) = \mu(Y)$  qui détermine l’ordre de présentation des pages. La fonction  $\mu$  est la **politique**.

On suppose que pour une paire  $(Y, \pi)$ , le **click-through-rate** de chaque page  $i$  (la probabilité de cliquer cette page) est

$$\text{CTR}(i) = \theta_{\pi(i)} \psi(R_i),$$

où  $1 \geq \theta_1 \geq \theta_2 \geq \dots \geq \theta_{m_0} > 0$  sont des constantes qui dépendent de la position des pages, et  $\psi : [0, 1] \rightarrow [0, 1]$  pondère la pertinence.

La **pertinence locale** pour une paire  $(Y, \pi)$  est

$$r(\pi, Y) = \sum_{i=1}^M \text{CTR}(i) R_i = \sum_{i=1}^M \theta_{\pi(i)} \psi(R_i) R_i.$$

Le **revenu espéré** pour cette requête est

$$g_0(\pi, Y) = \sum_{i=1}^M \text{CTR}(i) G_i = \sum_{i=1}^M \theta_{\pi(i)} \psi(R_i) G_i,$$

et le **revenu espéré pour l'engin** est

$$g(\pi, Y) = \sum_{i=1}^M \text{CTR}(i) \alpha_i G_i = \sum_{i=1}^M \theta_{\pi(i)} \psi(R_i) \alpha_i G_i.$$

La mesure de pertinence moyenne par requête, le revenu moyen, et le revenu moyen pour l'engin, pour une politique d'ordonnancement  $\mu$ , sont (l'espérance est p.r. à  $Y$ ):

$$\begin{aligned}r &= r(\mu) = \mathbb{E}[r(\mu(Y), Y)], \\g_0 &= g_0(\mu) = \mathbb{E}[g_0(\mu(Y), Y)], \\g &= g(\mu) = \mathbb{E}[g(\mu(Y), Y)].\end{aligned}$$

Supposons que le taux d'arrivé des requêtes est une fonction croissante de  $r$ , soit  $\lambda = \lambda(r)$ . L'engin peut augmenter son taux de revenu espéré en augmentant  $g$  sans trop diminuer  $r$ , ou en augmentant  $r$  sans trop diminuer  $g$ . Il faut trouver l'équilibre!

La mesure de pertinence moyenne par requête, le revenu moyen, et le revenu moyen pour l'engin, pour une politique d'ordonnancement  $\mu$ , sont (l'espérance est p.r. à  $Y$ ):

$$\begin{aligned} r &= r(\mu) = \mathbb{E}[r(\mu(Y), Y)], \\ g_0 &= g_0(\mu) = \mathbb{E}[g_0(\mu(Y), Y)], \\ g &= g(\mu) = \mathbb{E}[g(\mu(Y), Y)]. \end{aligned}$$

Supposons que le taux d'arrivé des requêtes est une fonction croissante de  $r$ , soit  $\lambda = \lambda(r)$ . L'engin peut augmenter son taux de revenu espéré en augmentant  $g$  sans trop diminuer  $r$ , ou en augmentant  $r$  sans trop diminuer  $g$ . Il faut trouver l'équilibre!

Sous certaines conditions (raisonnables), on peut montrer par un argument d'échanges de voisins que la politique optimale a la forme: toujours ordonner les pages selon leur valeur de  $\psi(R_i)R_i + \rho\alpha_i\psi(R_i)G_i$ , où  $\rho \in \mathbb{R}$  est une constante. On peut trouver la valeur optimale de  $\rho$  par recherche binaire.