Ordonnancement dynamique d'une machine flexible

Formulation en processus de Bandits-manchots

Fabrice Dusonchet[1] **— Max-Olivier Hongler**

Institut de Production Microtechnique (IPM)
Département de Microtechnique
EPFL
CH-1015 Lausanne Suisse
Fabrice.Dusonchet@epfl.ch
Max.Hongler@epfl.ch

RÉSUMÉ. *Nous utilisons une classe de processus de décisions séquentielles (i.e. « Restless Bandits ») pour déterminer de manière sous-optimale un ordonnancement dynamique d'une machine flexible à l'aide d'indices de priorité. Nous dérivons explicitement la forme de ces indices pour une dynamique sous-jacente décrite par des chaînes de Markov en temps continu et pour des fonctions coût convexes et linéaires par morceaux. On observe, sur un cas particulier, que l'ordonnancement dynamique obtenu par des tels indices de priorité donne des résultats très proches de l'optimum.*

ABSTRACT. *We explore the scheduling rules and the hedging levels that can be obtained by using a Restless Bandit Problem formulation of a make-to-stock production. The underlying dynamics are markov chain in continuous time and the associate reward are piecewise linear. We observe that, the use of priority indices to sub-optimally solve the Restless Bandit problem yields, on a particular example, results close to the optimal.*

MOTS-CLÉS : *Ordonnancement dynamique, Production sur stocks, Restless Bandit, Indices de priorité.*

KEYWORDS: *Dynamic Scheduling, Make-to-stock Production, Restless Bandit, Priority Indices.*

1. En partie supporté par le « Fonds National Suisse pour la Recherche Scientifique »

1. Introduction

Considérons le problème classique d'une machine flexible en flux poussé : une machine peut fabriquer N produits différents, chaque produit fini est placé dans un stock de couverture. La population de ce stock est donnée par $X_k(t)$. Dans le cas où des commandes en retard (« backorder ») sont acceptées, ce qui sera le cas dans la suite, $X_k(t)$ peut prendre des valeurs négatives et nous désignerons par $X_k(t)$ la position du surplus de type k. On aura donc :

$$\vec{X}(t) = (X_1(t), \ldots, X_k(t), \ldots, X_N(t)) \in \mathbb{Z}^N. \quad [1]$$

Le flux de production $\vec{P}(t)$ et le flux de demandes $\vec{D}(t)$ seront décrits par des processus stochastiques (supposés ici markoviens). On écrira :

$$\vec{P}(t) \cdot \vec{u}(t) = (P_1(t)u_1(t), \ldots, P_k(t)u_k(t), \ldots, P_N(t)u_N(t)) \quad [2]$$

et

$$\vec{D}(t) = (D_1(t), \ldots D_k(t), \ldots D_N(t)) \in \mathbb{Z}^N. \quad [3]$$

où $u_k(t) \in \{0, 1\}$ pour $k = 1, \ldots N$, est un contrôle de type « bang-bang » (i.e. $u_k(t) = 1$ signifie que nous produisons la variante de type k au temps t). Nous supposerons que la production n'est **pas « préemptive » :** on a l'obligation de finir la production de la pièce en cours avant de pouvoir en produire une autre et que la machine a une **capacité limitée** : elle ne peut fabriquer qu'un seul type de pièces à la fois ou ne rien produire. Si nous exprimons l'état du surplus en fonction des processus $\vec{P}(t) \cdot \vec{u}(t)$ et $\vec{D}(t)$ nous obtenons :

$$\vec{X}(t) = \int_0^t \vec{P}(s)\vec{u}(s) - \vec{D}(s)\, ds, \quad [4]$$

$$\vec{X}(t = 0) = \vec{x}_0.$$

A un état de surplus $\vec{X}(t)$, nous associons le coût instantané :

$$\vec{h}(\vec{X}(t)) = (h_1(X_1(t)), \ldots, h_N(X_N(t))) : \mathbb{Z}^N \to \mathbb{R}^+.$$

où $h_k(x)$ seront, dans la suite, des fonctions convexes.

Sous ces hypothèses, le coût total de stockage et de retard de livraison sur un horizon $\mathcal{H}$ est décrit par :

$$J_{\vec{u}(t)}(\vec{x}_0, \vec{u}_0) := E_{(\vec{x}_0, \vec{u}_0)} \int_0^{\mathcal{H}} e^{-\delta t} \vec{h}(\vec{X}(t))\, dt, \quad [5]$$

où $E_{(\vec{x}_0,\vec{u}_0)}$ désigne l'espérance sous la condition initiale $(\vec{x}_0, \vec{u}_0)$. Le paramètre $\delta > 0$ est un facteur d'actualisation. Nous voulons déterminer le contrôle optimal $\vec{u}^*(t)$ qui minimise le coût total de stockage et de retard de livraison, parmi tous les contrôles admissibles $\vec{u}(t)$ (i.e. dépendant uniquement de l'état présent et/ou passé du système) :

$$J^*(\vec{x}_0, \vec{u}_0) := J_{\vec{u}^*(t)}(\vec{x}_0, \vec{u}_0) = \min_{\vec{u}(t)} J_{\vec{u}(t)}(\vec{x}_0, \vec{u}_0).$$

Dans la suite, nous nous intéresserons uniquement aux problèmes pour lesquels les coûts et les temps de changement de variantes sont négligeables. L'absence de ces coûts simplifie considérablement notre problème. En effet, c'est en l'absence de coûts de changement, que le problème de l'ordonnancement dynamique d'une machine flexible peut être approché en découplant le problème original en N sous-problèmes indépendants à une seule variante. Plusieurs contributions traitent de ce sujet, parmi les plus récentes mentionnons : [VEA 96], [PEN 97], [VER 00]. Dans ces articles, les auteurs proposent des ordonnancements basés sur des indices de priorité, comme ceux introduits par Gittins pour le problème dit des « Bandits classiques » ([WHI 82]). En termes d'indice de priorité, la politique d'ordonnancement s'exprime comme suit : on attribue à l'état de chaque surplus de produits finis, un indice mesurant un coût effectif encouru, en fonction de l'état du surplus, si l'on décide de fabriquer ce produit. A l'aide de ces indices, la politique d'ordonnancement **prescrit de fabriquer à chaque instant t le produit ayant l'indice le plus bas**.

Les indices de Gittins permettent de résoudre de manière optimale tous les problèmes d'ordonnancement dynamique qui peuvent être modélisés par les processus connus sous le nom de « Bandits classiques » . Les Bandits classiques décrivent des systèmes où seul le projet engagé (par exemple ici le type de production engagée) voit son état évoluer, les autres projets restent figés dans l'état où on les a laissé. De plus, seul le projet engagé engendre un coût. Un exemple de système qui peut être naturellement décrit par les Bandits classiques est celui des machines à sous : un joueur se trouvant devant N machines à sous, dont il connaît partiellement les processus stochastiques décrivant leur évolution, doit décider, sur la base de ses observations antérieures et présentes, sur quelle machine jouer, afin de maximiser son gain sur un horizon infini. L'engagement d'un projet lui apporte un gain d'information sur sa dynamique stochastique.

Dans le cas d'un système de production, la demande pour un type de produit donné continue d'arriver pendant qu'on fabrique un autre type de produit. Nous sommes donc face à un problème dit du type « Bandit en évolution permanente » communément appelé « Restless Bandit » (introduit par [WHI 88]). Les « Restless Bandits » généralisent les Bandits classiques dans plusieurs directions :

a) tous les projets (engagés ou non) évoluent simultanément,

b) tout les projets (engagés ou non) engendrent des coûts,

c) on doit engager M projets à chaque instant ($0 < M < N$).

Nous savons qu'une politique basée sur des indices de priorité résout généralement le problème des Restless Bandits de façon sous optimale mais constitue souvent une stratégie performante. Il est donc intéressant de calculer explicitement de tels indices pour un choix de dynamique donné et d'évaluer la pertinence de la politique d'ordonnancement qui en découle.

Pour la suite nous adopterons le plan suivant : dans la section 2, nous introduisons la formulation générale du problème des « Restless Bandits » en temps continu et états continus puis nous exposons brièvement un problème approché connu sous le nom de *relaxation de Whittle*. Dans la section 3, nous formulons le problème de l'ordonnancement d'une machine flexible en termes de processus de « Restless Bandits » . Nous calculons explicitement les indices de priorité pour le cas spécial où les processus décrivant la fabrication des produits et l'arrivée des demandes sont des processus de renouvellement et où les fonctions coût sont linéaires par morceaux. Nous concluons par un exemple numérique qui illustre la performance de l'ordonnancement résultant de l'utilisation de ces indices de priorité.

2. « Restless Bandits » en temps et espace continu et la relaxation de Whittle [WHI 88]

Considérons une collection de N projets, (i.e. systèmes dynamiques), dont l'évolution dans le temps est décrite par $X_k(t) \in \mathbb{R},\ t \in \mathbb{R}^+$, pour $k = 1, 2, ...N$. Chaque projet $X_k(t)$ peut être dans une phase active ou passive dont la dynamique est décrite pas les processus stochastiques $X_{a,k}(t)$ et $X_{p,k}(t)$,

$$X_k(t) = \begin{cases} X_{a,k}(t) & \text{si le projet } k \text{ est actif au temps t,} \\ X_{p,k}(t) & \text{si le projet } k \text{ est passif au temps t.} \end{cases} \qquad [6]$$

Au temps $t \in \mathbb{R}^+$, **exactement** $0 < M < N$ projets doivent être en activité. Quand le projet k dans l'état x_k, est activé il engendre un coût $h_k^a(x_k)$ par unité de temps. Ce coût est actualisé dans le temps par un terme de la forme $e^{-\delta t}$, ($\delta > 0$). Les autres $(N - M)$ projets restent passifs et engendrent un coût $h_k^p(x_k)$ par unité de temps qui est actualisé par le même terme.

Problème. Ordonnancement optimal :

Pour des conditions initiales données $\vec{x}_0$ et $\vec{u}_0$, choisir quels projets engager à chaque instant, (i.e. déterminer l'ordonnancement des tâches) afin de minimiser le coût total, dont la forme est donnée par :

$$J^*(\vec{x}_0, \vec{u}_0) = \min_{\vec{u} \in \mathcal{U}} E_{(\vec{u}_0, \vec{x}_0)} \left[\int_0^\infty \sum_{k=1}^{N} h_k^{\mathcal{P}_k(s)}(X_k(s)) e^{-\delta s} ds \right] \qquad [7]$$

avec $\mathcal{P}_k(t) \in \{a, p\}$ dénotant l'action (active ou passive) correspondant au projet k au temps t lorsque la politique $\vec{u}(t)$ est adoptée ; $E_{(\vec{u}_0, \vec{x}_0)}$ dénote l'espérance pour un état initial du système $(\vec{x}_0, \vec{u}_0)$. Rappelons que l'optimisation de l'équation [7] doit être faite sous la condition qu'*exactement M projets doivent être engagés à chaque instant t*.

La contribution de [PAP 94] montre que le problème des « Bandits Restless » est PSPACE-hard, ce qui compromet généralement les chances de le résoudre de manière algorithmique. Dès lors, il est nécessaire d'étudier des méthodes d'approximations pour ce problème. Un cadre d'approximation naturel a été introduit par P. Whittle [WHI 88] qui considère une version affaiblie du problème des « Restless Bandits » pouvant être résolue par des algorithmes de complexité polynômiale (cf. [NIN 99] et [NIN 00]). Cette approximation connue sous le nom de *relaxation de Whittle* consiste en l'étude d'une version modifiée du problème original où l'on affaiblit la condition qu'**exactement M projets** doivent être activés à chaque instant t en la remplaçant par la condition plus faible que M **projets doivent être activés en moyenne**.

Pour cette version affaiblie, une approche analytique peut être envisagée. A l'instar de P. Whittle [WHI 88], nous utilisons le formalisme du multiplicateur de Lagrange et découplons le problème en N sous-problèmes à un seul projet, puis nous discutons chaque projet indépendamment. Le multiplicateur de Lagrange γ est interpreté comme étant une taxe systématiquement perçue lorsque le projet est en phase passive. Ce découplage nous permet de définir un *indice de priorité* $\nu_k(x_k)$ pour chaque projet k. La valeur de cet indice correspond à la valeur de la taxe qui rend les phases actives et passives du projet k, également attractives. Inversement, pour une valeur donnée de γ, il est optimal d'engager le projet k (optimal d'être actif) dans tous les états x pour lesquels $\nu_k(x) \leq \gamma$. Clairement, pour que l'indice $\nu_k(x_k)$ puisse être utilisé comme indice de priorité résolvant le problème relaxé de Whittle, nous devons tout d'abord vérifier une condition d'indexabilité. On dit que le problème relaxé de Whittle est indexable si, en accord avec l'intuition, la cardinalité de l'ensemble des états où il est optimal d'être passif diminue de manière monotone lorsque la taxe γ croît (pour plus de détail se reporter à l'annexe A).

Si l'indice $\nu_k(x)$ est croissant et indexable, et si d^* est tel que $\nu_k(d^*) = 0$, on en déduit qu'il est optimal d'être actif pour tous les états x satisfaisant $x < d^*$ (inversement, si $x \geq d^*$, il est optimal d'être passif). Pour l'ordonnancement d'une machine flexible, la position d^* détermine un stock de couverture (« hedging stock »). Une fois les indices de priorité calculés, l'heuristique pour résoudre le problème des Restless Bandits prend la forme : **engager à chaque temps $\mathbf{t}$ les $\mathbf{M}$ projets présentant les valeurs d'indice $\boldsymbol{\nu_k(x_k)}$ les plus petites**.

3. Ordonnancement dynamique d'une machine flexible en flux poussé - (« Make-to-Stock Production »).

Formulons maintenant le problème de production exposé en section 1 en termes de « Restless Bandits » . Nous étudions le cas où la demande et la production du produit k sont modélisées par des marches aléatoires (i.e. des chaînes de Markov en temps continu) ayant respectivement les paramètres μ_k et λ_k.

Le problème de l'ordonnancement dynamique d'une machine flexible peux s'écrire de manière naturelle en terme de « Restless Bandits » avec $N+1$ projets de la façon suivante :

i) Identifions la position des surplus $X_k(t)$, $k = 1, 2, \ldots N$ avec l'état des N premiers projets du Bandit Restless.

ii) Pour inclure la possibilité que la machine soit à l'arrêt, ajoutons un projet d'indice $N+1$ dont la dynamique est $X_{N+1}(t) \equiv x_c$ $\forall t \in \mathbb{R}$ (i.e. l'état de ce projet est figé). De plus $X_{N+1}(t)$ n'engendre aucun coût (i.e. $h_{N+1}(x) \equiv 0$).

iii) Le coût en phase active $h_k^a(x)$ est égal au coût en phase passive $h_k^p(x)$, $\forall$ k, et nous supposons ici que ces coûts sont linéaires par morceaux pour chaque variante :

$$\begin{gathered} h_k^a(x) = h_k^p(x) = h_k(x) = A_k x^+ + B_k x^-, \\ x^{\pm} = \max(\pm x, 0). \end{gathered} \qquad [8]$$

Découplons le problème en utilisant la méthode de Whittle, supprimons l'argument k et supposons que les indices résultant vérifie la propriété d'indexabilité. Pour un surplus à un niveau x, deux actions sont possibles : soit produire une pièce ou ne pas produire. Ces décisions engendrent des coûts, respectifs $J_a(x)$ et $J_p(x)$ déterminés par le principe de la programmation dynamique :

$$\left\{ \begin{array}{l} \delta J_a(x,\gamma) = h(x) + \lambda J_a(x-1,\gamma) + \mu J_a(x+1,\gamma) - (\lambda+\mu) J_a(x,\gamma) \\ \\ \delta J_p(x,\gamma) = h(x) + \lambda J_p(x-1,\gamma) - \lambda J_(x,\gamma) + \gamma. \end{array} \right. \qquad [9]$$

La linéarité de ce système implique que sa solution est de la forme :

$$\begin{array}{l} J_a(x,\gamma) = C_{a^+}(w_+)^x + C_{a^-}(w_-)^x + S_a(x,\gamma) \\ J_p(x,\gamma) = C_p(w_0)^x + S_p(x,\gamma) \end{array} \qquad [10]$$

où C_{a^+}, C_{a^-}, C_p sont des constantes d'intégration,

$$w_+ = \frac{(\delta+\lambda+\mu) + \sqrt{(\delta+\lambda+\mu)^2 - 4\lambda\mu}}{2\mu}$$

$$w_- = \frac{(\delta+\lambda+\mu) - \sqrt{(\delta+\lambda+\mu)^2 - 4\lambda\mu}}{2\mu}$$

$$w_0 = \frac{\lambda}{\lambda+\delta}. \qquad [11]$$

Les fonctions $S_a(x)$ et $S_p(x)$ sont les solutions particulières correspondant aux situations respectives : produire en permanence ou ne jamais produire.

En observant que $0 \leq w_- \leq w_0 \leq 1 \leq w_+$ le comportement à l'infini des solutions implique que :

$$\lim_{x \to -\infty} J_a(x,\gamma) = S_a(x,\gamma) \Rightarrow C_{a^-} = 0.$$

En appliquant un principe de « collage lisse » (« smooth fitting ») à la frontière active/passive x_0 nous obtenons que :

$$\begin{aligned} J_a(x_0,\gamma) &= J_p(x_0,\gamma) + \gamma \\ J'_a(x_0,\gamma) &= J'_p(x_0,\gamma) \\ J''_a(x_0,\gamma) &= J''_p(x_0,\gamma). \end{aligned} \qquad [12]$$

Ces conditions nous permettent d'exprimer, après des calculs simples mais fastidieux (c.f. annexe B), l'indice $\nu(x)$ sous la forme compacte :

Quand $x \geq 0$:

$$\nu(x) = \frac{2A\mu - \mu(A+B)(w_+ + w_-)(w_-)^x + \mu(A+B)(w_+ - w_-)(w_-)^x}{2\delta(\delta+1)} \qquad [13]$$

Quand $x < 0$:

$$\nu(x) = -\frac{B\mu}{\delta(\delta+1)} \qquad [14]$$

Pour une production flexible de N produits satisfaisant à une dynamique de la forme de l'équation [9] et caractérisée par des coûts du type de l'équation [8)], nous obtenons donc N indices de la forme des équations [13] et [14] dont une représentation graphique typique se trouve en Figure 1. L'ordonnancement dynamique découlant de ces indices de priorité est donc la suivante : *« A chaque instant, engager la production du type de produit présentant la plus petite valeur d'indice. »*

Des résultats des équations [13] et [14] on tire les observations suivantes :

– Pour $x < 0$ on produit la variante ayant la plus petite valeur de $B\mu$ parmi toutes les variantes ayant leur stock $x < 0$, ce qui est en accord avec la politique optimale déja connue pour cette situation [HA 94].

– Pour $x \to \infty$ on a que $\lim\limits_{x\to\infty} \nu(x) = \frac{A\mu}{\delta(\delta+1)}$ ce qui est également en accord avec la politique optimale (voir [VEA 96]).

– La position d^* donnée par $\nu_k(d^*) = 0$ correspond au stock de couverture de la variante k. Nous l'avons dérivé dans [DUS 00] et nous trouvons :

$$d^* = \left\lfloor \frac{1}{\ln(\beta)} \ln\left(\frac{A}{A+B}\right) \right\rfloor$$

avec $\beta = \frac{(\delta+\lambda+\mu)-K}{2\mu}$ et $K = \sqrt{(\delta+\lambda+\mu)^2 - 4\lambda\mu}$.

– Quand $\delta \to 0$, l'indice $\nu(x) \to -\infty$. Il n'existe donc pas d'indices de priorité dans cette limite comme déja remarqué dans [VEA 96].

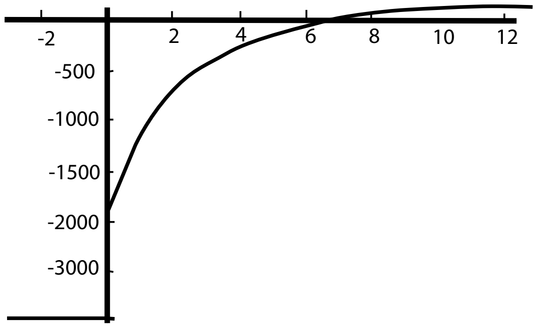

Figure 1. *Indice pour une variante ayant les paramètres suivant* $A = 1$, $B = 40$, $\mu = 1$, $\lambda = 0.5$

4. Illustration numérique et conclusions

Dans cette section nous testons la qualité de la politique d'indices de priorité obtenue par les Restless Bandits sur le problème de l'ordonnancement dynamique d'une machine flexible produisant sur stock. Plus précisément la politique d'indice de priorité va être comparée à la politique optimale, calculée numériquement par Y. Ha ([HA 97]), dans le cas ou la machine ne fabrique que deux sortes de produits. Pour le cas plus général d'une machine fabriquant N produits différents, nous la comparerons à quelques heuristiques classiques. Pour nos tests, nous avons utilisé un taux de production μ identique pour chaque produit. Nous avons choisi un facteur d'actualisation $\delta = 0.01$ et nous avons supposé que les stocks étaient vides initialement.

4.1. *Heuristique classique pour une machine flexible produisant sur stock*

La première heuristique est connue sous le nom d'heuristique $h_k\mu/b_k\mu$ (ici b_k représente le coûts engendré par les demandes en retard (backorder) de type k, h_k représente le coût de stockage d'un produit de type k et μ_k correspond au taux de production du produit k). On peut la définir comme suit :

a) **Si il y a des demandes en retard :** Fabriquer le produits ayant la plus grande valeur de $b_k\mu_k$ parmi tous les produits en backorder.

b) **Si il n'y a pas de demande en retard :** Fabriquer le produit ayant la plus petite valeur de $h_k\mu_k$ parmi tous les produits qui ont un niveau de stock inférieur à leur stock de couverture.

La deuxième heuristique est connue sous le nom *d'heuristique d'enclenchement* (« *switching rule* »). Elle est définie en modifiant l'heuristique $h_k\mu/b_k\mu$ comme suit :

a) **Si il y a des demandes en retard :** l'heuristique d'enclenchement est identique à l'heuristique $h\mu/b\mu$.

b) **Si il n'y a pas de demandes en retard :** l'heuristique d'aiguillage ordonne de fabriquer le produit ayant la plus grande valeur $b_k\mu_k(1 - x_k/d_k^*)$ si cette valeur est positive, sinon de ne rien produire (c.f. [HA 94] ou [VEA 98] pour plus de détails).

4.2. *Résultats numériques*

Nous avons utilisé le simulateur Taylor ED pour réaliser nos tests. Taylor ED est un simulateur événementiel en temps discret qui possède une bibliothèque de routines spécifiques aux problèmes de production. Chaque résultat présenté est la moyenne du coût total mesuré sur plus de 4000 simulations. Etant donné que nous ne pouvons pas simuler un coût de production sur un horizon infini nous avons choisi un horizon $\mathcal{H}$ suffisamment grand tel que tout choix d'un horizon $\mathcal{H}' > \mathcal{H}$ n'engendre pas de changement significatif du coût de production (ceci est possible grâce à la présence du facteur d'actualisation). Pour la simulation de l'heuristique $h\mu/b\mu$ et de l'heuristique d'enclenchement nous avons déterminer numériquement la position des stocks de couverture qui minimisent le coût total de production.

Les résultats.

– Une machine produisant deux variantes avec les paramètres suivants, $\lambda_1 = 0.4$; $\lambda_2 = 0.5$; $\mu = 1$; $B_1 = 30$; $B_2 = 40$; $A_1 = 1$; $A_2 = 1$:

	Hedging	Cost
Ha	$(9, 11)$	7401
Restless	(5,8)	7437
Switching	$(4, 7)$	7639
$h\mu/b\mu$	$(1, 8)$	7838

	Pourcentage supplémentaire
Ha	optimal
Restless	0.5 % de plus que l'optimal
Switching	3.2 % de plus que l'optimal
$h\mu/b\mu$	5.9 % de plus que l'optimal

Tableau 1. *Coût moyen de production pour une machine produisant 2 variantes*

– Une machine produisant quatre variantes avec les paramètres suivants, $\lambda_1 = \frac{1}{4}$; $\lambda_2 = \frac{1}{4.1}$; $\lambda_3 = \frac{1}{4.2}$; $\lambda_4 = \frac{1}{4.3}$; $\mu = 1$; $B_1 = 40$; $B_2 = 30$; $B_3 = 20$; $B_4 = 10$; $A_k = 1$, $k = 1, 2, 3, 4$:

	Hedging	Coût	Pourcentage supplémentaire
Restless	(2,3,3,3)	7745	Best
Switching	$(2, 3, 3, 4)$	8366	10.4 % de plus que le Restless
$h\mu/b\mu$	$(1, 1, 3, 3)$	8983	18.6 % de plus que le Restless

Tableau 2. *Coût moyen de production pour une machine produisant 4 variantes*

Observations :

– L'ordonnancement dérivé à l'aide des indices de priorités pour une machine fabriquant deux produits donne des résultats très proches de l'optimum.

– Pour plus de variantes, les simulations numériques effectuées montrent que la politique d'indices de priorité est meilleure, sinon au moins équivalente à la politique d'aiguillage.

A. La relaxation de Whittle

Définissons les fonctions d'action :

$$I_k(t) = \begin{cases} 1 & \text{si le projet } k \text{ est activé au temps t} \\ 0 & \text{sinon} \end{cases}$$

$$\bar{I}_k(t) = 1 - I_k(t).$$

A l'aide de $I_k(t)$, le coût total engendré pour le problème relaxé de Whittle $J^W(\vec{x})$ s'écrit :

$$J^W(\vec{x}) = \begin{cases} \min\limits_{u \in \mathcal{U}} E_{(u,\vec{x})} \Bigg[\int_0^\infty \sum\limits_{k=1}^N h_k^a(X_k(t)) I_k(t) e^{-\delta t} dt + \\ \qquad\qquad \int_0^\infty \sum\limits_{k=1}^N h_k^p(X_k(t)) \bar{I}_k(t) e^{-\delta t} dt \Bigg] \\ \\ \text{sous la contrainte} \quad \int_0^\infty \sum\limits_{i=1}^N I_i(t) e^{-\delta t} dt = \frac{M}{\delta}. \end{cases} \qquad [15]$$

De la définition du temps d'occupation $I_k(t)$, nous avons que :

$$\int_0^\infty \sum_{k=1}^{N} (I_k(t) + \bar{I}_k(t))e^{-\delta t} dt = \frac{N}{\delta}.$$

Puis, en utilisant le multiplicateur de Lagrange, l'équation [15] se réécrit comme :

$$J^W(\vec{x}, \gamma) = \sum_{k=1}^{N} J^k(x_k) - (N - M)\frac{\gamma}{\delta}, \qquad [16]$$

avec

$$J^k(x_k, \gamma) = \min_{u \in \mathcal{U}} E_{(u,x_k)} \Big[\int_0^\infty h_k^a(X_k(t)) I_k(t) e^{-\delta t} dt$$
$$+ \int_0^\infty h_k^p(X_k(t)) + \gamma) \bar{I}_k(t) e^{-\delta t} dt \Big].$$

Whittle interprète le multiplicateur γ comme étant une *taxe que l'on doit payer lorsque nous n'engageons pas le projet* k. Clairement, le problème d'optimisation donné par l'équation [16] peut être découplé en N sous-problèmes de Bandits Restless à un seul projet $J^k(x_k, \gamma)$. Dans la suite nous nous concentrons sur un de ces problèmes $J^k(x_k, \gamma)$ et nous n'écrivons donc plus l'argument k. En utilisant la théorie de la programmation dynamique, nous pouvons montrer que le coût optimal $J^*(X(t), \gamma)$ satisfait la propriété suivante : Si au temps t, nous décidons d'activer ou de désengager le projet pendant un temps (infinitésimal) ξ, alors le coût encouru de t à $t + \xi$ plus le coût optimal $J^*(X(t + \xi), \gamma)$ reçu à partir du temps $t + \xi$ est plus élevé ou égal que le coût optimal $J^*(X(t), \gamma)$.

De cette observation, nous tirons l'inégalité :

$$J^*(X(t), \gamma) \leq \min \Big(\int_t^{t+\xi} h^a(X(t)) I(t) e^{-\delta(s-t)} ds + e^{-\delta\xi} J^*(X(t + \xi), \gamma);$$
$$\int_t^{t+\xi} (h^p(X(t)) + \gamma) \bar{I}(t) e^{-\delta(s-t)} ds + e^{-\delta\xi} J^*(X(t + \xi), \gamma) \Big). \qquad [17]$$

Si la politique optimale est d'être actif ou passif pendant l'intervalle de temps ξ alors l'équation [17] est satisfaite avec l'égalité. Supposons que le ξ est bien choisi et que l'équation [17] est satisfaite avec l'égalité. Nous pouvons alors définir le coût $J_a(X(t), \gamma)$ [respectivement $J_p(X(t), \gamma)$] engendré par la décision d'être actif [respectivement passif] pendant l'intervalle de temps ξ par :

$$J^*(X(t)) = \min \left(J_a(X(t), \gamma), J_p(X(t), \gamma) \right),$$

Nous pouvons maintenant définir *l'indice* $\nu(x)$ pour un projet dans l'état $X(t) = x$ comme étant la valeur minimale de γ tel que :

$$J_a(X(t), \gamma) = J_p(X(t), \gamma). \qquad [18]$$

Autrement dit, $\nu(x)$ est la plus petite valeur de la taxe qui rend équivalant la décision d'engager le projet ou de le laisser passif quand il est dans l'état x. Remarquons que si nous fixons le valeur de γ, il est optimal d'engager le projet dans tous les états x pour lesquels $\nu(x) \leq \gamma$ et inversement, quand $\nu(x) > \gamma$, il est optimal de le laisser passif. Pour pouvoir utiliser l'indice $\nu(x)$ en tant qu'indice de priorité, le problème équation [7] doit satisfaire une condition d'indexabilité :

Définition (Indexabilité). On dit que le problème équation [7] est indexable si pour chaque projet k, l'ensemble des états dans lesquels il est optimal activer le projet k croît de façons monotone de l'ensemble vide à tout l'espace quand la taxe γ augmente de $-\infty$ à $+\infty$.

B. L'indice de priorité pour une machine produisant sur stock.

Les calculs de l'indice pour une machine flexible dont la demande et la production sont décrites par des chaînes de Markov en temps discret sont assez fastidieux et nous n'indiquons ici que les grandes étapes (voir [DUS 00] pour tous les détails).

Puisque l'équation [9] est une équations linéaires, une solution explicite peut être dérivée pour $J_a(x,\gamma)$ et de $J_p(x,\gamma)$. Nous obtenons une solution générale de la forme de l'équation [10] avec les constantes de l'équation [11]. Les fonctions $S_a(x,\gamma)$ et $S_p(x,\gamma)$ sont les solutions particulières qui correspondent respectivement à produire en permanence ou à ne jamais produire ($\theta \in \{a,p\}$) :

$$\begin{aligned}
&S_\theta(x,\gamma) = \int_0^\infty e^{-\delta s} h(X(s))ds = \\
&\int_0^\infty e^{-\delta s} \sum_{k=-\infty}^{\infty} h(k) P_\theta\{X(s)=k|X(0)=x\}ds = \\
&\int_0^\infty e^{-\delta s} \sum_{k=-\infty}^{\infty} h(k) P_\theta\{X(s)=k-x|X(0)=0\}ds.
\end{aligned} \qquad [19]$$

Concentrons nous d'abord sur $S_a(x,\gamma)$. Pour ce cas, le processus $X(t)$ est une marche aléatoire de paramètre λ et μ. En utilisant les résultats connus pour des marche aléatoire en temps continu, nous pouvons calculer explicitement

$$P_n(t) = P_a\{X(t)=n|X(0)=0\}.$$

L'équation [19] devient (voir [DUS 00]) :

$$S_a(x,\gamma) = \frac{1}{\mu(w_+ - w_-)} \Bigg\{ h(x) + (w_-)^x \sum_{k=-\infty}^{x-1} h(k)(w_-)^{-k} + (w_+)^x \sum_{k=x+1}^{\infty} h(k)(w_+)^{-k} \Bigg\}.$$

Pour calculer $S_p(x, \gamma)$, il suffit de remarquer que $P_p\{X(s) = k - x|X(0) = 0\}$ est un processus de Poisson et nous obtenons :

$$S_p(x, \gamma) = (w_0)^{x+1} \sum_{k=-\infty}^{x} \frac{(h(k) + \gamma)(w_0)^{-k}}{\lambda}.$$

La deuxième étape consiste à résoudre le système d'équation [12] pour γ. Tous calculs faits, nous obtenons l'indice sous la forme donnée par les équations [13] et [14].

5. Bibliographie

[DUS 00] DUSONCHET F., HONGLER M.-O. « Continuous Time Restless Bandits for Make-to-stock Productions », *Preprint EPFL* 2001.

[HA 97] HA A. « Optimal dynamic scheduling policy for a make-to-stock production system », *Operation Res.*, **45**, 1997, 42-54.

[NIN 99] NINO-MORA J. « Restless Bandit, partial conservation law and index ability », *Adv. Appl. Prob.*, **33**, 1, 2001, 76-98.

[NIN 00] NINO-MORA J. « On certain greedoid polyhedra, partially indexable scheduling problems, and extended restless Bandit allocation indices », Submitted to *mathematical programming*, 2000.

[PAP 94] PAPADIMITRIOU C. H., TSITSIKLIS J. N. « The complexity of optimal queueing network control », *Math. Oper. Res.*, **24**, 1999, 293-305.

[PEN 97] PENA A., ZIPKIN P. « Dynamic scheduling rules for a multiproduct make-to-stock queue », *Operation Res.*, **15**, 1997, 919-930.

[VEA 96] VEATCH M.H, WEIN L. M. « Scheduling a make-to-stock queue : index policies and hedging points », *Operation Res.*, **44**, 1996, 634-647.

[VER 99] de VERICOURT F., KARAESMEN F., DALLERY Y. « Dynamic scheduling in a make-to-stock system : A partial Characterization of optimal policies », *Oper. Res.*, **48**, 5, 2000, 811-819.

[WHI 82] WHITTLE. P. *« Optimization over Time. Dynamic Programming and Stochastic Control », J. Wiley, New-York*, 1982.

[WHI 88] WHITTLE P. « Restless Bandit : Activity in a changing world », *J. Appl. Prob.*, **25A**, 1988, 287-298.