






Open Archive Toulouse Archive Ouverte

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is an author's version published in: <http://oatao.univ-toulouse.fr/20927>

To cite this version:

Duquenne, Philippe  and Vidal, Emmanuel  and Pingaud, Hervé 
Optimisation des plans de charge pour un flow-shop dans le cadre d'une production en juste temps : 1 - Convergences entre besoins et capacités. (1999) In: 3ème Congrès Franco-Quebequois de Génie Industriel, 1999 - 1999 (Montréal, Canada). (Unpublished)

Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr

**OPTIMISATION DES PLANS DE CHARGE POUR UN FLOW-SHOP
DANS LE CADRE D'UNE PRODUCTION EN JUSTE A TEMPS :
1- CONVERGENCE ENTRE BESOINS ET CAPACITES**

Philippe DUQUENNE, Emmanuel VIDAL, Hervé PINGAUD
E.N.S.I.G.C., Département Génie Industriel,
Institut National Polytechnique de Toulouse
Laboratoire de Génie Chimique UMR CNRS 5503, INPT/ENSIGC
Chemin de la Loge, 31078 Toulouse Cedex, France

Mots clefs : juste à temps, ordonnancement, gestion des effectifs, optimisation, convergence

Résumé court :

Les politiques de production en « Juste A Temps », contrairement aux fabrications à « débit constant », satisfont aux impératifs de réactivité et de réduction des en-cours, mais engendrent d'importantes perturbations au niveau de l'affectation des ressources lorsque la demande est variable. La caractérisation et l'utilisation de méthodes d'organisation du travail permettent d'assurer la convergence entre besoins et capacités de production.

Introduction

La gestion des ateliers de production consiste à rechercher le meilleur compromis, ou du moins un compromis économiquement acceptable entre des impératifs commerciaux (le respect de dates de livraison des produits à leurs acquéreurs), financiers (la minimisation des stocks et en-cours, mobilisateurs de trésorerie) et industriels (maximisation du taux d'utilisation des investissements matériels et humains, respect de la qualité). Les critères permettant de considérer un tel compromis comme acceptable font partie des décisions portant sur la politique de fabrication adoptée par l'entreprise, qui fixera les différentes priorités respectives attribuées à ces trois critères.

D'une manière très schématique, on peut classer les politiques de fabrication en deux grandes familles, en fonction des poids relatifs accordés à chaque critère :

- historiquement les plus anciennes, les méthodes visant en premier lieu à optimiser l'utilisation de l'outil industriel. Elles se caractérisent par la recherche d'un lissage maximal de l'emploi des ressources de fabrication (matérielles ou humaines) au cours du temps : il en résulte une production à débit constant. La prise en compte des impératifs commerciaux vient alors en seconde position dans le processus de planification des activités de production. Elle a pour effet correcteur, lorsqu'on prévoit ou constate des fluctuations de la demande, d'anticiper les périodes de fortes demandes par la constitution de stocks de produits finis lors des "périodes creuses" précédentes. L'aspect financier, lorsqu'il est pris en compte, vient finalement examiner une solution quasiment figée, et voit limiter son rôle à celui de mesure du niveau des immobilisations, disposant d'assez peu de degrés de liberté pour influencer de manière significative comme agent correcteur - par exemple, la définition de bornes maximales sur le niveau des stocks.

- la transition d'une économie de marché où l'offre était inférieure à la demande vers une situation inverse met le client dans une position dominante, réclamant outre des produits irréprochables, des coûts réduits et des délais de livraison courts. Cette situation justifie la mise en œuvre de politiques différentes visant à la production en juste-à-temps : les impératifs commerciaux et financiers prennent le pas sur les critères purement industriels (Rosenblatt et Hau, 1996 ; Agrawal *et al.*, 1996). Cette fois, partant d'un calendrier de travail calqué sur les dates de mise à disposition des produits finis à leurs clients, la priorité impose une planification visant à la minimisation des attentes des produits finis comme des en-cours de production. On en retiendra essentiellement qu'elles restent souvent focalisées sur les immobilisations sous forme de produits finis au détriment des en-cours. D'autre part, et ce point est plus significatif, si la disponibilité des machines est prise en compte comme contrainte incontournable dans le problème d'ordonnancement, la gestion des ressources humaines est rarement abordée, ou alors d'une manière encore très schématique (Benhouiba-Zelfani *et al.*, 1997) : les conséquences d'une politique de juste-à-temps sur les plans de charge des opérateurs restent souvent peu explorées quant à leur contribution à l'évaluation d'une solution calculée.

L'objet de cette étude est de présenter une démarche intermédiaire entre ces deux approches, consistant à résoudre un problème d'ordonnancement sans privilégier *a priori* l'un ou l'autre des impératifs de la gestion de production : partant du principe de la satisfaction des attentes, en termes de délai, des acheteurs de produits, l'ordonnancement d'un programme de fabrication sera élaboré avec la prise en compte permanente à la fois des immobilisations de trésorerie et des perturbations de plans de charge que chacune des solutions adoptées entraînera.

Nous présenterons dans un premier temps le principe de l'adaptation du travail aux contraintes imposées par le calendrier de disponibilité des produits finis. Les décisions envisageables favorisant cette adaptation seront ensuite énumérées et détaillées. Nous présenterons enfin les critères permettant d'évaluer (voire hiérarchiser) des solutions obtenues.

Plans de charge : nivellement, lissage et convergence

L'utilisation des ressources nécessaires à la réalisation de toute activité industrielle est, assez logiquement, un des premiers critères permettant de juger de la faisabilité même de cette réalisation. Cette affirmation, assez générale pour paraître évidente, se traduit toutefois par des disparités importantes quand il s'agit de la mettre en pratique.

Dans le domaine de la gestion de la production tout d'abord (Blondel, 1997), l'examen se porte en premier lieu sur les machines mobilisées par un programme de fabrication. L'état binaire de ces machines (disponibles ou non) a introduit la notion de contraintes disjonctives, qui pilotent les algorithmes d'ordonnancement. Quand aux ressources humaines, leur état de service est déduit de celui des machines ou des postes qu'elles supervisent. De ce fait, culturellement, le responsable de la production a tendance à juger comme acceptable un programme de fabrication dont les plans de charge sont lissés - c'est-à-dire réguliers et bien entendu proches du fatidique 100% d'utilisation pendant les heures ouvrées de l'atelier. Avec la principale conséquence que l'on a déjà citée d'un rythme de production lui aussi régulier, pas nécessairement adapté aux besoins du marché d'où l'obligation de constituer de coûteux stocks - tampons de produits finis afin d'absorber d'éventuelles fluctuations de la demande.

Plus récemment, la gestion de projet a repris les fondements méthodologiques de la gestion de la production, mais avec cette fois des impératifs davantage axés sur la rapidité d'exécution des réalisations industrielles. La notion de plan de charge subsiste, et reste

susceptible de remettre en cause une première politique de planification (Westney, 1991). Mais le chef de projet doit conserver l'arbitrage entre la vitesse d'avancement du projet et la gestion des éventuelles sur - utilisations de ressources : ces dernières ne sont plus le critère essentiel de faisabilité (Chvidchenko et Chevallier, 1996 ; Esquirol et Lopez, 1999). Le terme de lissage suppose alors l'arrêt des algorithmes de résolution du problème d'ordonnancement contraint par les ressources dès lors que la durée de réalisation pourrait être remise en cause. Libre au chef de projet de trouver des solutions alternatives - s'il n'en trouve pas ou que le délai n'est pas crucial, on accepte alors une augmentation de la durée du projet, et on parle alors de nivellement.

Ce regard apporté par la gestion de projet sur des méthodes issues de la gestion de production a fait admettre l'existence, voire la nécessité économique, de solutions alternatives à la régularité des plans de charge, au nom de la réactivité de l'entreprise (Jagou, 1993 ; Maurino, 1993). Dans cette optique, un algorithme d'ordonnancement ne peut plus se limiter à l'examen des seules ressources disjonctives : d'une part le principe commence à être généralement admis d'une certaine "souplesse" dans la disponibilité des ressources humaines, de part et d'autre d'un quota fixe d'heures de travail par jour et par personne ... Quand aux ressources matérielles, hors problèmes de maintenance, leur service n'a jamais été entravé par la réglementation du travail, et le nivellement de leur plan de charge a toujours été la conséquence de la disponibilité de leurs opérateurs ...

La démarche présentée ici consiste donc à rechercher une méthode d'ordonnancement qui ne soit plus pilotée par une disponibilité figée des ressources, mais qui prenne en compte leur flexibilité. L'objectif est d'arriver à une concordance la plus étroite possible entre deux courbes représentant, au cours du temps, l'une les besoins en ressources d'un programme de fabrication donné et l'autre la disponibilité de ces ressources. Mais plutôt que de chercher à systématiquement ré - échelonner le calendrier de travail pour niveler la courbe des besoins, on cherchera à trouver une convergence entre ces deux courbes, c'est-à-dire un compromis entre ce réagencement et une perturbation jugée acceptable des plans de charge.

Deux types de critères pilotent la recherche d'un tel compromis. D'une part des critères économiques : le réagencement d'un programme de réalisations supposera la mise en attente de produits intermédiaires : on peut en estimer le coût. Des retards pris ont des répercussions en terme de pénalités financières. Le gonflement ponctuel des effectifs qui entraîne le recours à de la main d'œuvre externe , ou encore l'inactivité engendrée par des périodes de sous-utilisation des opérateurs peuvent également être évalués selon un critère économique.

Plus délicates à interpréter, les perturbations induites dans les plans de charge des ateliers peuvent se répercuter sur la disponibilité, les tâches, voire les activités des opérateurs.

Dans un cas comme dans l'autre, l'objet de ce travail est de présenter les paramètres de mesure des perturbations induites, et de proposer des indicateurs facilitant leur appréciation.

Limites de l'étude et hypothèses de base

Soit un atelier mono - produit à cheminement unique, chargé de réaliser un programme de fabrication. Nous supposons bien entendu que la charge de travail induite par ce programme est *globalement* compatible avec les capacités de l'atelier, et ce, compétence par compétence - à ceci près que, bien entendu, les dates de disponibilité attendues pour les produits finis ne correspondent pas du tout à une production régulière (figure 1).

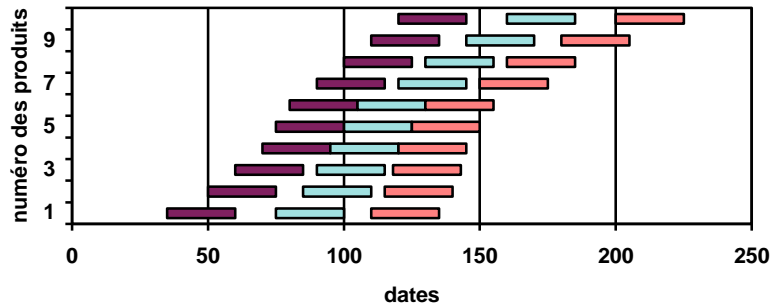


figure 1 : les produits et leurs dates de livraison

Supposons une fabrication initialement basée sur une date d'achèvement des produits identique à leur date de livraison (fabrication en juste-à-temps). Le premier problème qui se pose est justement de déterminer quel est le lieu de convergence des deux courbes besoins / capacités. En laissant de côté pour l'instant les périodes où la capacité excède les besoins, donc en nous focalisant sur les périodes de surcharge, nous pouvons identifier quels sont les postes concernés par chacune de ces surcharges. Avec deux attitudes extrêmes :

- l'anticipation du travail sur le produit d'ordre le moins élevé dans le programme - et pour conséquence la création d'une attente sur le produit fini ou sur un produit intermédiaire, conséquence dont on peut évaluer le coût en terme d'immobilisation de trésorerie dans un premier temps. On pourra également majorer ce coût de frais financiers liés à cette immobilisation.
- Accepter la surcharge, en supposant que la flexibilité dont est capable notre atelier est suffisante : il est dans ce cas moins facile d'estimer les répercussions économiques de cette appréciation. D'autre part, accepter la surcharge suppose donc que l'on soit à même de quantifier les possibilités de notre atelier d'absorber des surcharges.

Enfin, la recherche d'un juste milieu entre la quantité de charge excédentaire que peut absorber l'atelier et la remise en cause de la planification au plus tard implique la connaissance fine, par le décideur, des commandes dont il dispose en terme de flexibilité du travail.

La flexibilité et ses leviers de commande

* *Le choix des dates de livraison :*

Pour chacun des produits du programme de fabrication, ce qu'on appelle "date de livraison" est en fait caractérisé par :

- un ensemble de dates jalonnant le processus de négociation de livraison du produit à un client, qui peut être à l'origine (prévisions à moyen terme) le service commercial de l'entreprise avant d'être un client identifié,

- pour chacune de ces dates - jalon, une fenêtre de livraison d'étendue de plus en plus restreinte. Il n'est pas exclu dans ce contexte que l'avis du responsable de l'atelier ait une influence déterminante sur la date finale retenue. Pour chacun des produits 'i' à fabriquer, la livraison contractuelle doit donc avoir lieu entre deux dates L_i^{\min} et L_i^{\max} . La date de livraison réelle L_i^f doit donc vérifier :

$$L_i^{\min} \leq L_i^f \leq L_i^{\max}$$

Pour la dernière date ou plage de livraison retenue, tout écart est sanctionné : en avance

par des frais d'immobilisation du produit fini, et en retard par des pénalités dont une partie au moins est relativement aisée à déterminer.

*** La modulation des cycles :**

Chacun des postes successifs 'j' de la gamme de fabrication du produit est caractérisé par une charge de travail à réaliser par les opérateurs - voire par plusieurs charges de travail $\chi_{j,k}$ si plusieurs compétences différentes 'k' peuvent intervenir sur le même poste. Le cycle du poste $c_{i,j}$ est donc lié à l'effectif présent $E_{i,j,k}$ (compétence par compétence), et à cette charge caractéristique par la relation :

$$c_{i,j} = \frac{\chi_{j,k}}{E_{i,j,k}}$$

Dans cette optique, le produit à fabriquer n'est plus vu comme une succession d'opérations de durées fixes à ordonnancer : chacune de ces opérations peut se voir affecter une durée fluctuante de part et d'autre d'une valeur indicative, en fonction du nombre des opérateurs affectés : il n'y a plus indépendance entre ressources et durée d'opération.

Première caractéristique de cette modulation des cycles : les effectifs sont par nature des nombres entiers, ce qui implique que les cycles correspondants prendront des valeurs discrètes. D'autre part, pour des raisons de synchronisation des postes entre eux ou avec les horaires de travail, certaines de ces valeurs pourront être soit arrondies soit tout simplement inapplicables. Des soucis de rentabilité du poste en opération d'un côté, et de saturation en effectif de l'autre viendront borner les valeurs admissibles pour le cycle $c_{i,j}$:

$$c_j^{\min} \leq c_{i,j} \leq c_j^{\max}$$

*** La modulation d'horaires :**

Cette commande de flexibilité porte sur la plage quotidienne ou hebdomadaire d'ouverture de l'atelier, dont on suppose qu'elle peut fluctuer entre des valeurs extrêmes - les maxima étant régis par la réglementation du travail.

Cette variable de décision peut également inclure des choix sur les périodes de fermeture des atelier (fermeture générale en été, ou lors de fêtes caractéristiques, ou étalement contrôlé des congés, etc.)

Là encore, les paramètres caractéristiques de cette commande seront un choix à faire parmi un nombre limité de valeurs discrètes, assorties d'un délai de préavis et/ou d'une certaine "inertie" dans la périodicité des modifications. En revanche, elle ne remet pas en cause les affectations des opérateurs sur les postes ni la distribution du travail, donc n'entraîne aucune conséquence en terme d'efficacité : elle correspond juste à un changement d'échelle entre les vitesses de défilement relatives du temps calendaire et du temps ouvré des ateliers.

*** Les compétences multiples :**

Il s'agit ici de considérer qu'un opérateur donné n'est plus caractérisé par une compétence unique, mais par un ensemble de compétences différentes qui peuvent régir son affectation sur divers postes de travail. On peut identifier trois familles de paramètres :

* D'une part, ceux qui décrivent l'éventuelle polyvalence de l'opérateur : dans chacune des compétences représentées au sein de l'atelier, on peut définir son efficacité - 100% dans son métier d'origine, et des valeurs non nulles en regard de

compétences qu'il aura acquises.

* On peut également décrire la facilité qu'aura un opérateur d'un métier donné à acquérir un autre métier. Sans aller jusqu'à prendre en compte les profils psychologiques individu par individu, on peut évaluer la facilité de se tourner d'une qualification vers une seconde au moyen d'une "matrice de passage" indiquant par exemple un coût ou une durée de formation - une valeur prohibitive traduisant la difficulté d'envisager une réorientation excessive.

* Enfin, on peut tenir compte de la dégradation des compétences acquises et non entretenues au moyen d'une réévaluation périodique des coefficients d'efficacité en fonction de l'activité vécue.

On note au passage que la gestion des compétences multiples oriente très rapidement vers un suivi quasiment individuel des opérateurs : il est difficile de la prendre en compte, même à un niveau agrégé, sans augmenter considérablement la taille de notre problème. Par exemple, en ne considérant que les éventuelles doubles compétences et sans nous occuper de la notion d'efficacité, l'existence de m compétences dans l'atelier nous amène à gérer m^2 sous-groupes d'opérateurs ...

*** La main d'œuvre extérieure :**

Dernier outil aidant à la flexibilité, le recours à la main d'œuvre extérieure à l'entreprise (intérim, sous-traitance, ...) permet de tolérer un certain nombre de surcharges non absorbées par les autres méthodes, sans nécessairement recourir au ré - échelonnement du travail - et sans avoir à aligner l'effectif permanent de l'entreprise sur des périodes de suractivité ponctuelles.

Caractéristiques de ce recours : un taux maximal d'utilisation de cette main d'œuvre, imposé par le souci de préserver le climat social de l'entreprise, ou la maîtrise de ses réalisations. On peut prendre en compte une efficacité des personnels extérieurs moindre que celle des opérateurs permanents, ainsi qu'un coût horaire d'utilisation différent.

Indicateurs de performance

*** Les critères économiques :**

Ils ont l'avantage d'exprimer l'estimation résultante dans une seule unité de mesure commune, une unité monétaire : dans ce domaine, les différentes conséquences des solutions étudiées seront aisément comparables, au moins qu'une automatisation de l'algorithme de résolution du problème posé passera nécessairement par ces critères : la solution visera à minimiser une fonction d'évaluation du coût total de réalisation du programme de fabrication.

Ce coût total peut se décomposer en :

F_1 : frais financiers liés aux stockages des composants avant entrée en atelier ;

F_2 : frais financiers liés aux en-cours de production ;

F_3 : frais financiers liés au stockage des produits finis ;

F_4 : pénalités de retard ;

F_5 : coût du travail réellement effectué;

F_6 : coût de l'inactivité (coût des salaires à effectif maximum sur toute la période – F_5);

Par ailleurs, il convient de noter que l'adoption d'un plan de charge non lissé en terme d'effectif revient à admettre que contrairement au cas d'une production à débit fixé (production cadencée), une partie seulement du coût salarial correspond à du travail effectif : ce coût salarial, en fonction du taux d'occupation moyen des effectifs pendant les périodes ouvrées, se

distribuera entre un coût "utile" à la production, et un coût accepté d'inactivité des opérateurs.

*** Les indicateurs physiques :**

Citons le taux d'utilisation moyen des effectifs. Il sert tout d'abord, on l'a vu, à distribuer la masse salariale sur le programme de fabrication envisagé entre une dépense "utile" et une dépense liée à « l'inactivité acceptée ». Derrière l'interprétation économique de ce taux, se cache l'hypothèse que nous savons maîtriser les conséquences sociales de l'inactivité engendrée ...

$$\tau_{ut} = \frac{\sum_{t=1}^T E(t)}{T \cdot E^{MAX}}$$

Deuxième critère, dont il est également difficile de prévoir la portée mais qui a le mérite de pouvoir être quantifié, le taux moyen de variation des effectifs : il correspond à la variation moyenne des effectifs présents dans l'atelier d'un jour à l'autre. En ce sens, il est représentatif de la perturbation induite dans le travail par rapport à un plan de charge rigoureusement lissé (pour lequel ce taux est nul).

$$\tau_{var} = \frac{\sum_{t=2}^T ABS[E(t) - E(t-1)]}{E^{MOY}}$$

Récapitulatif

commande de flexibilité	fréquence d'application	amplitude	grandeurs caractéristiques	coût de mise en œuvre	impact social
dates de livraison	grande à moyen terme	grande à moyen terme	différents jalons, différentes plages pénalités de retard	faible ou nul	sans objet
modulation de cycles	faible	grande	préavis valeurs discrètes efficacité moyenne	important	négociation
modulation d'horaires	faible	moyenne	préavis valeurs discrètes	faible ou nul	négociation ?
compétences multiples	moyenne à grande	moyenne ?	préavis efficacité acquisition maintien	moyen à important	négociation ?
main d'œuvre extérieure	faible	limitée	taux limite efficacité	faible à moyen	négociation

Mise en œuvre

Deux voies d'approfondissement s'offrent à la démarche présentée ici. Le premier souci est la recherche d'une solution optimale, et s'appuie donc sur une formulation complète et exacte.

Par ailleurs, on a vu que les indicateurs de performance des solutions obtenues ne sont pas nécessairement compatibles entre eux. La question du temps de réponse se pose

également, qui fait hésiter entre l'identification d'une solution optimale et l'émergence plus rapide de solutions satisfaisantes. D'autant plus que se pose ensuite le problème de la robustesse de ces solutions.

C'est pourquoi cette communication sera complétée dans le cadre de ce congrès par deux autres présentations sur le même thème ("*Optimisation des plans de charge pour un flow-shop dans le cadre d'une production en Juste A Temps*"), mais avec des objectifs différents : l'une, ("*Formulation mathématique*"), vise à la recherche d'une solution optimale au problème posé, au moyen d'une formulation exacte ; la seconde, ("*Algorithmes génétiques et approche multi-critères*"), présente la possibilité d'étude de cas plus complexes par le biais de méthodes heuristiques.

Nomenclature

c_j^{\max}	:	cycle maximum pour le poste j
c_j^{\min}	:	cycle minimum pour le poste j
$c_{i,j}$:	cycle pour le produit i sur le poste j
$E_{i,j,k}$:	effectif de la compétence k pour le produit i sur le poste j
$E(t)$:	effectif total de l'atelier pour le jour t
E^{MAX}	:	effectif maximal de l'atelier sur la période considérée
E^{MOY}	:	effectif moyen de l'atelier sur la période considérée
F_1	:	frais financiers liés aux stockages des composants avant entrée en atelier ;
F_2	:	frais financiers liés aux en-cours de production ;
F_3	:	frais financiers liés au stockage des produits finis ;
F_4	:	pénalités de retard ;
F_5	:	coût du travail réellement effectué;
F_6	:	coût de l'inactivité
L_i^f	:	date de livraison effective du produit i
L_i^{\max}	:	date de livraison au plus tard du produit i
L_i^{\min}	:	date de livraison au plus tôt du produit i
T	:	nombre de jours de la période considérée
τ_{ut}	:	taux d'utilisation moyen des effectifs
τ_{var}	:	taux de variation moyen des effectifs
$\chi_{j,k}$:	charge de travail pour la compétence k sur le poste j

Références bibliographiques

- Agrawal A., Harhalkis G., Minis I., Nagis R., "*Just-in-time production of large assemblies*", IEEE Transactions (1996), 28, pp 653-667
- Benhouba-Zelfani F. Z., Zelfani M., Caillaud E., Duquenne P., Lacoste G., "*Applications intégrées et outils de simulation et d'aide à la décision*", deuxième congrès international franco-québécois de Génie Industriel, Albi (1997)
- Blondel, F., "*Gestion de la production*", Dunod, 1997
- Chvidchenko I., Chevalier J., "*Conduite et gestion de projets, principes et pratique pour petits et grands projets*", Cepadues, 1993
- Esquirol P., Lopez P., "*L'ordonnancement*", Economica, 1999
- Jagou P., "*Concurrent Engineering, la maîtrise des coûts, des délais et de la qualité*", Hermès, 1993
- Maurino M., "*La gestion des données techniques, technologie du concurrent engineering*", Masson, 1993
- Rosenblatt M. J., Hau L. L., "*The effects of work-in-progress inventory costs on the design and the scheduling of assembly lines with low throughput and high component costs*", IEEE Transactions (1996), 28, pp 405-414

Westney R. E., "*Gestion de petits projets, techniques de planification, d'estimation et de contrôle*", AFNOR
Gestion,, 1991

de chaque article de produit fini correspond à une succession d'opérations réalisées sur les postes, et enchaînées de manière linéaire. Chaque opération d'un article fait appel à des ressources sur le poste de travail. Nous nous focalisons sur la gestion des ressources humaines.

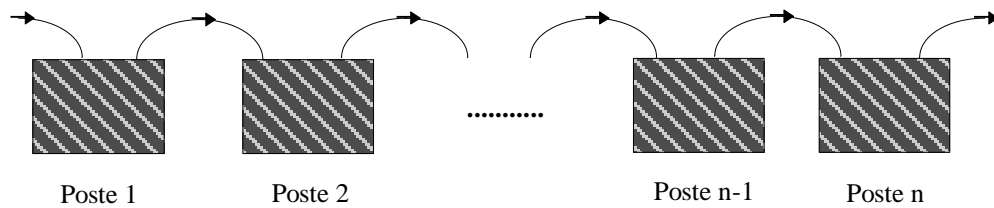


Figure 1 - Atelier de fabrication à cheminement unique

Par définition, la partie de la gamme opératoire attribuée à un poste correspond une charge de travail déterminée (en hommes.heures). Pour chaque article et opération, la durée opératoire, ou cycle, est considérée comme une fonction inversement proportionnelle au nombre d'opérateurs travaillant sur le poste au moment de l'opération. Par exemple, plus l'effectif du poste est élevé, plus le cycle est court. Cette variabilité est limitée dans un intervalle déterminé a priori. Par variations adaptées des cycles, il est donc possible de jouer sur les flux de production pour essayer de répondre à une demande non régulière en bout de chaîne. Corrélativement, cette adaptation induit des variations du rythme de travail des individus que l'on cherchera à minimiser dans la mesure du possible.

Les variables de décision caractérisant la gestion des ressources humaines dans un contexte réel est le sujet d'un article des mêmes auteurs dans ce cycle de conférences [2]. Les différentes formes de la flexibilité y sont énumérées, ainsi que les critères de performances permettant de juger de la pertinence d'une solution. Dans la suite, le problème sera simplifié. L'affectation des ressources sera considérée sous une forme agrégée: la composition des équipes sur les postes et les compétences ne seront pas détaillées.

Nous étudierons la modélisation du système en vue d'une aide à la décision en formulant un problème d'optimisation sous contraintes. La fonction critère est de nature technico-économique. Elle évalue le coût de revient réel d'un nombre déterminé d'articles (horizon fini de planification), incluant la main d'œuvre et les coûts de stockage. Ainsi, par exemple, pour rendre compte de façon réaliste des perturbations inhérentes à la variation des cycles, elle augmente sous l'effet de l'inactivité éventuelle des opérateurs ou de la non qualité.

Le lien tissé entre les durées opératoires et le niveau de ressources est une spécificité et une source de complexité. Les modélisations classiques abordent généralement la modélisation sous l'hypothèse de durées opératoires indépendantes [3]. Il convient donc de développer cette partie de l'étude et nous proposons, au paragraphe suivant, une forme de modélisation appropriée.

3- Modélisation du processus de production

3.1 - Représentation de l'échelle de temps

Le temps peut être modélisé sous forme discrète par un nombre fini d'intervalles de durées connues. Ce nombre sera d'autant plus grand que la précision souhaitée sera fine. Chaque variable de décision est définie par une valeur sur un intervalle de temps. Les intervalles étant numérotés par des entiers, le temps apparaît directement comme un indice de la variable. Une décision se traduit par le changement de valeur d'au moins une variable de décision entre deux intervalles consécutifs.

modèle par la superposition de deux diagrammes de Gantt : produit/temps et ressource/temps. Chaque opération est représentée par une barre copiée en deux exemplaires qui est étiquetée par le numéro de l'opération. Ces copies sont placées dans la partie supérieure et dans la partie inférieure et vont évoluer de concert. La partie supérieure du diagramme (produit/temps) représente les contraintes temporelles pour chaque article : l'opération 1 précède l'opération 2. La partie inférieure du diagramme (ressource/temps) traduit les contraintes disjonctives : une ressource n'accepte qu'une activité à un temps donné.

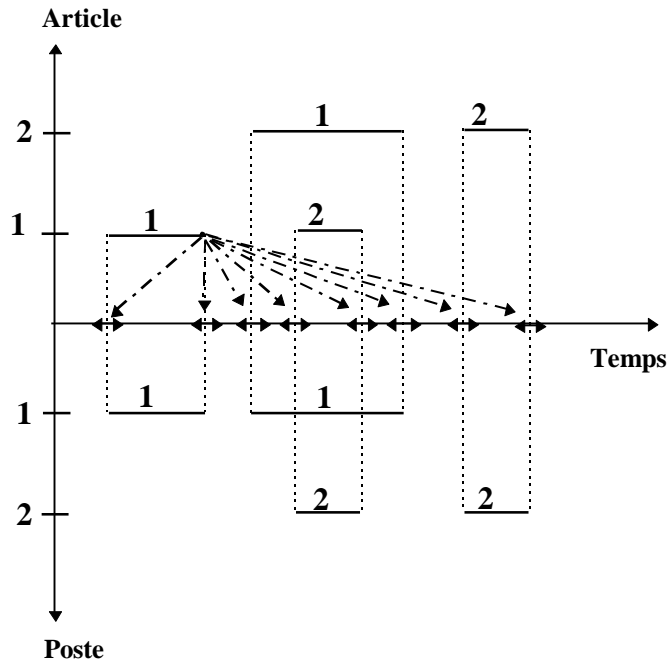


Figure 3 - Diagrammes de Gantt du plan de production

Les flèches qui partent de la fin de l'opération 1 pour l'article 1 indiquent le rôle des variables binaires ($ae(1,1,k)$, $k=1,8$), c'est à dire l'association effective avec un des huit événements possibles. Les événements portés sur l'axe de temps sont accompagnés de flèches à double sens pour rappeler que les événements ne sont pas datés, mais peuvent se déplacer.

3.3 Généralisation du modèle

Dans ce paragraphe, nous établissons la forme générale du modèle qui sera soumis à l'optimiseur. Les concepts introduits précédemment au moyen d'un exemple simple sont maintenant repris sous une forme générique. Nous présentons d'abord les variables du modèle, puis la forme des contraintes et, enfin, la fonction objectif.

3.3.1 Les variables du modèle

Soient

i indice pour l'article
 j indice pour le poste
 k,o indices pour l'événement

Les ensembles de variables entières sont:

$n(j,k)$ variable indiquant la disponibilité du poste j entre les événements k et $k+1$
 $ae(i,j,k)$ variable binaire désignant l'affectation de la fin d'opération à l'événement k
 $as(i,j,k)$ variable binaire désignant l'affectation du début d'opération à l'événement k
 E^{tot} effectif présent sur l'atelier
 $E(k)$ effectif présent sur l'atelier entre les événements k et $k+1$ (état k)
 $e(i,j)$ effectif utilisé pour le poste j et l'article i

$y(k)$ variable d'écart pour un événement k possible et non affecté

Les ensembles de variables continues sont:

$t(k)$ dates d'occurrence de l'événement k

$c(i,j)$ durée opératoire ou cycle

Les données sont :

m nombre d'articles
 n nombre de postes et d'opérations
 v nombre d'événements possibles
 $c_{j,\min}$ valeur minimale du cycle pour le poste j
 $c_{j,\max}$ valeur maximale du cycle pour le poste j
 $l(i)$ date de livraison planifiée de l'article i
 w_j charge de travail pour le poste j

3.3.2 Les contraintes

L'écriture des contraintes se traduit par un ensemble d'équations de type égalité ou inégalité. Chaque groupe de relations correspond à l'expression d'une contrainte physique particulière du système de production. Puisque le nombre d'événements possibles est un nombre maximal, il faut prévoir le cas où des événements de cette liste ne sont pas affectés. Dès lors, il n'y aura pas de date, ni de fait, correspondant à la réservation. Ces événements seront gérés au moyen des variables d'écart $y[k]$, et rangés en début de liste grâce à des contraintes spécifiques. Les autres événements seront classés par ordre chronologique.

- relation de précédence pour un article i entre deux opérations (postes) de la gamme

$$\forall i, \forall j > 1, \forall k > 1, \quad as(i, j, k) \leq \sum_{o=1}^{k-1} ae(i, j, o) \quad (2)$$

Pour qu'une opération j débute sur l'événement k , l'événement déclenchant la fin de l'opération précédente ($j-1$) doit être affectée à un événement antérieur (de 1 à $k-1$).

- cohérence des événements de début et de fin de chaque opération

$$\forall i, \forall j, \forall k > 1, \quad ae(i, j, k) \leq \sum_{o=1}^{k-1} as(i, o, j) \quad (3)$$

Pour qu'une opération se termine à l'événement k , elle doit avoir débutée par une affectation sur un événement antérieur (de 1 à $k-1$).

- respect de la date de livraison échue de chaque article telle que consignée par la planification

$$\forall i, \quad l(i) \geq \sum_{k=1}^n ae(i, n, k) \cdot t(k) \quad (4)$$

La mise à disposition de l'article se fait à la fin de la dernière opération, date définie par la partie droite de l'inégalité ci dessus.

- affectation cohérente des postes

$$\begin{aligned}
& \forall j, \forall k > 1, & n(j,0) = 1 \\
& \text{et} & n(j, k) = n(j, k - 1) - \sum_{i=1}^m as(i, j, k) + \sum_{i=1}^m ae(i, j, k) \\
& \text{et} & 0 \leq n(j, k) \leq 1 \\
(5)
\end{aligned}$$

Le poste, initialement disponible, fait l'objet d'un bilan des actions qui débutent et qui se terminent à la fin de chaque période. La variable désignant l'état de disponibilité est contrainte puisqu'il ne peut y avoir au plus qu'une opération par poste sur chaque période.

- définition de la durée opératoire et des dates de réalisation

$$\begin{aligned}
& \forall i, \forall j, & c(i, j) - \sum_{k=1}^v as(i, j, k) \cdot t(k) + \sum_{k=1}^v ae(i, j, k) \cdot t(k) = 0 \\
& & c_{j,\min} \leq c(i, j) \leq c_{j,\max} \\
(6)
\end{aligned}$$

$$\forall k > 1, \quad t(k - 1) \leq t(k)$$

La durée de chaque opération est calculée comme la différence entre les dates de fin et de début d'opération. Chaque durée est contrainte entre deux bornes. Les événements sont classés par ordre chronologique.

- relation entre l'effectif et le cycle sur chaque poste et article

$$\forall i, \forall j, \quad e(i, j) \cdot c(i, j) - w_j = 0 \quad (7)$$

Le produit de l'effectif du poste pour chaque article par la durée de l'opération correspond à la charge de travail programmée sur le poste.

- Détermination de l'effectif de l'atelier

(8)

L'effectif total de l'atelier à chaque état est calculé comme la somme des effectifs des postes entre les événements qui délimitent l'état. L'effectif maximal sur l'horizon de calcul sera le plus grand effectif calculé des différents états.

- Assignation effective des événements possibles et classement par ordre chronologique

(9)

Les variables d'écart sont forcées à 1 s'il n'y a aucun démarrage et aucune fin d'opération associés à cet événement. Les variables d'écart sont classées par valeurs décroissantes de manière à regrouper les événements non affectés en début de liste. Ainsi, dès qu'une opération commence ou se termine, la variable d'écart est forcée à zéro.

3.3.3 La fonction objectif

La fonction objectif chiffre le coût de revient d'une solution en ne prenant en compte que les frais financiers sur les sommes engagées. Si le taux d'opportunité a pour valeur τ , le surcoût sur un objet de valeur V pour une durée d petite est calculé par l'approximation suivante :

$$V(1 + \tau)^d - V \approx V\tau d \quad (10)$$

Elle est formée d'une somme de cinq contributions correspondant aux activités suivantes:

- les stocks de matière première avant chaque poste

$$F_1 = \sum_{i=1}^m \sum_{j=1}^n V_{i,j} \tau (d_{i,j} - \delta_{i,j})$$

avec

$V_{i,j}$ valeur ajoutée de l'article i au début du poste j

$d_{i,j}$ date de début de l'opération j pour l'article $i = \sum_{k=1}^v as(i, j, k) \cdot t(k)$

$\delta_{i,j}$ date de disponibilité de la matière pour l'opération j de l'article i

- les articles en cours

$$F_2 = \sum_{i=1}^m \sum_{j=1}^n W_{i,j} \tau (f_{i,j} - d_{i,j})$$

avec

$W_{i,j}$ valeur de l'article i pendant la durée du poste $j = \sum_{l=1}^j V_{i,l}$

$f_{i,j}$ date de fin de l'opération j pour l'article $i = \sum_{k=1}^v ae(i, j, k) \cdot t(k)$

- le stock final avant livraison

$$F_3 = \sum_{i=1}^m W_{i,n} \tau (l(i) - f_{i,n})$$

- le travail

$$F_5 = (f_{m,n} - d_{1,1}) \cdot E^{\text{tot}} U$$

avec

U coût horaire du travail

- l'inactivité

$$F_6 = \sum_{k=1}^{v-1} (t(k+1) - t(k)) \cdot (E^{\text{tot}} - E(k)) \cdot S$$

avec

S coût horaire de l'inactivité

4- Conclusions et perspectives

Nous avons présenté une programmation mathématique du problème en variables mixtes (entières et continues). La fonction et certaines des contraintes sont de type non linéaire exclusivement sous la forme d'un produit de deux variables. Il est possible de reprendre la formulation pour les rendre linéaires au prix d'une augmentation du nombre de contraintes. Par ailleurs, la présence de variables à trois indices (i, j, k) augmente rapidement la taille du problème qui deviendra nécessairement limité par la complexité dans une utilisation à l'échelle de l'atelier industriel.

C'est pour cette raison que la démarche adoptée dans le travail de thèse d'E.Vidal [8] est composée de

deux étapes. La première consiste à résoudre la formulation proposée dans cet article sur un cas d'étude représentatif de la problématique industrielle, mais de taille modeste de manière à rester dans le domaine d'obtention de la solution dans un temps raisonnable. Cette étape conduit alors à un optimum. Dans une seconde partie du travail, il reprend le problème en le résolvant par des heuristiques basées sur une séquence de sous problèmes qui autorise le traitement de tailles importantes. Les solutions des deux parties sont comparées sur le cas d'étude pour estimer le bien fondé des heuristiques [7]. Puis les heuristiques sont employées sur les problèmes de taille plus importante (plusieurs dizaine d'articles et une dizaine de postes)

Références :

- [1] Gousty Y. (1998) Le Génie Industriel, Coll. « Que sais je ? », PUF Ed.
- [2] Vidal E., Duquenne P., Pingaud H. (1999) Optimisation des plans de charge pour un flow shop dans le cadre d'une production en juste à temps : 1-convergence entre besoins et capacités, 3^{ème} colloque international de Génie Industriel, Montréal, Canada
- [3] Lopez P., Esquirol P. (1998) L'ordonnancement, Ed. Economica
- [4] Blazewicz J., Dror M., Weglarz J. (1991) Mathematical programming formulations for machine scheduling : A survey, European Journal of Operational Research, 51, 283-300
- [5] Pinto J. M. and I. E. Grossmann (1995) A continuous time mixed integer linear programming model for short term scheduling of multistage batch plants, Ind. Eng. Chem. Res., **34**
- [6] Pinto J. M. and I. E. Grossmann (1996) An alternate MINLP model for short-term scheduling of batch plants with preordering constraints, Ind. Eng. Chem. Res., **35**
- [7] Vidal E., Duquenne P., Pingaud H. (1999) Optimisation des plans de charge pour un flow shop dans le cadre d'une production en juste à temps : 3-algorithmes génétiques et approche multi-critères, 3^{ème} colloque international de Génie Industriel, Montréal, Canada
- [8] Vidal E., (1999) thèse de doctorat en préparation à l'Institut National Polytechnique de Toulouse, Département Génie des Systèmes Industriels

Optimisation des plans de charge pour un flow-shop dans le cadre d'une production en Juste A Temps :
2- Algorithmes génétiques et approche multi - critères

Emmanuel VIDAL, Philippe DUQUENNE, Hervé PINGAUD*
Département Génie des Systèmes Industriels, Institut National Polytechnique de Toulouse
6, allée Émile Monso, BP 4038, 31029 Toulouse Cedex, FRANCE
E-mail: vidale@univ-inpt.fr duquenne@univ-inpt.fr
* Laboratoire de Génie Chimique UMR CNRS 5503, INPT/ENSIGC
Chemin de la Loge, 31078 Toulouse Cedex, France
E-mail: Herve.Pingaud@ensigct.fr

Mots clefs : juste à temps, ordonnancement, gestion des effectifs, optimisation, algorithmes génétiques

Résumé court :

L'ordonnancement d'un atelier dans lequel on cherche simultanément à réduire les frais financiers engendrés par les stockages et les en-cours tout en conservant une affectation satisfaisante des effectifs est un problème d'optimisation complexe. Que la fonction objectif soit mono ou multi – critères, les algorithmes génétiques permettent d'obtenir rapidement des solutions efficaces.

Introduction

L'évolution des relations entre les entreprises et leurs clients a amené les premières à s'adapter au mieux aux exigences des seconds. Parmi les impératifs majeurs, on notera, outre les impératifs en termes de qualité et de personnalisation des produits, la réduction des délais et la satisfaction des dates de livraison.

De telles contraintes ont donc conduit à envisager la transition de politiques de production « à cadence régulière » vers des politiques de production en « Juste à Temps ». Il s'agit alors de passer d'un système dans lequel la planification et l'ordonnancement sont établis de manière à obtenir les moindres perturbations des plans de charge (et par conséquent en produisant selon le rythme le plus régulier possible) à un système dans lequel la date de livraison au client sert de référence à toute la planification des tâches et des activités selon le mode « au plus tard » ; un tel mode de fonctionnement permet alors de réduire les coûts liés aux stockages (immobilisation du capital) mais entraîne alors des perturbations importantes du plan de charge. L'industriel peut cependant choisir un compromis entre les deux principes afin de limiter les perturbations dans l'organisation du travail et les coûts afférents (baisse de performance, désorganisation des équipes, etc.). Il est alors intéressant de considérer le problème d'optimisation multi-critères correspondant à la prise en compte de ces différents objectifs.

La problématique générale et les méthodes d'organisation permettant de formuler le problème sont présentées dans la communication : « Optimisation des plans de charge pour un flow shop dans les cadre d'une production en Juste A Temps : Convergence entre besoins et capacité » (Philippe Duquenne, Hervé Pingaud, Emmanuel Vidal).

La recherche et l'évaluation des solutions de telles questions sont des problèmes d'optimisation complexes portant sur des systèmes de grandes dimensions et prenant en compte des critères tels que :

- l'évaluation financière des frais financiers dus aux stocks et en-cours ;
- la quantification des perturbations des plans de charge ;
- la satisfaction du client en terme de respect des dates de livraison.

Ces caractéristiques projettent le problème dans le domaine de l'optimisation multi-critères.

De plus, les différentes fonctions prises en compte ne présentent pas nécessairement les propriétés mathématiques nécessaires à l'application de méthodes de programmation mathématique, notamment en termes de continuité, de convexité et de dérivabilité.

Dans un tel contexte, l'utilisation des algorithmes génétiques apparaît comme un bon compromis entre la qualité des solutions (obtention de « bonnes » solutions) et le temps de calcul, particulièrement dans le cas de l'optimisation multi – critères.

Présentation du problème

Description du système :

Le but est de calculer l'ordonnancement d'un atelier de type flow-shop monoproduit, dans lequel chaque station de travail possède les caractéristiques suivantes :

- un ensemble d'opérateurs, correspondant le cas échéant à plusieurs types de compétence, travaille simultanément sur la station de travail ;
- chaque station est caractérisée par la charge de travail (= volume d'heures de travail) par compétence correspondant au passage d'un produit ;
- selon les cas, une station peut effectuer simultanément un ou plusieurs traitements parallèles ;
- la relation entre la charge de travail et l'effectif nécessaire est la suivante :
$$\text{charge} = \text{cycle} * \text{effectif}$$

où le cycle désigne la durée opératoire affectée à la station ;
- le cycle de chaque poste doit être compris dans un intervalle dont la borne inférieure est déterminée par le chemin critique des tâches à effectuer et dont la borne supérieure est établie selon des considérations organisationnelles (taille minimale d'une équipe, par exemple).

Explicitation de la demande :

Le calendrier de livraison est connu *a priori* et peut être caractérisé par :

- l'intervalle de tolérance à l'intérieur duquel peut avoir lieu la livraison ;
- les pénalités de retards encourues en cas de dépassement de la borne supérieure du créneau de livraison ;
- les frais de stockage du produit fini, si celui-ci est terminé avant le début de l'intervalle de livraison.

Objectifs du calcul :

On cherche donc à déterminer les valeurs optimales pour les variables suivantes :

- dates de livraison (à l'intérieur de l'intervalle de tolérance) ;
- cycles de production par poste et par produit ;
- la durée des éventuelles périodes de stockage ;
- horaires de travail (dans le cadre de la modulation d'horaires).

Dans un premier temps, la résolution du problème a suivi un processus séquentiel, issue de la décomposition du problème en une série de sous-problèmes, pour lesquels chaque optimisation porte sur un type homogène de variables de décisions liées à l'organisation (par exemple, valeurs des cycles de production, horaires de travail, etc.).

Afin d'étudier la position des solutions ainsi obtenues par rapport à la solution optimale globale, une formulation s'appuyant sur une représentation discrète et non-uniforme du temps a été établie, permettant d'obtenir la solution optimale (cf. communication : « Optimisation des plans de charge pour un flow shop dans les cadre d'une production en Juste A Temps : Formulation mathématique », Philippe Duquenne, Hervé Pingaud, Emmanuel Vidal).

L'utilisation d'algorithmes génétiques mono - critère permet de résoudre de manière séquentielle les différents sous - problèmes, chaque solution d'un sous problème permettant de figer les valeurs d'un groupe homogène de variables de décision pour la résolution du sous problème suivant. De plus, le recours à des algorithmes génétiques multi - critères rend possible l'obtention de solutions satisfaisantes au regard des différents critères, ainsi que la détermination de la surface de Pareto (dont la définition est donnée ci-après).

Les algorithmes génétiques

Principe de la méthode

Les algorithmes génétiques reposent sur des concepts inspirés de la biologie et visent à reproduire l'évolution d'une population dans un milieu hostile ([2], [3]). Ils constituent un outil intéressant dans le cadre de la résolution de problèmes d'optimisation combinatoire, bien que la théorie mathématique associée (théorie des schémas) soit encore limitée face à la complexité inhérente aux problèmes traités.

Alors que la plupart des méthodes d'optimisation combinatoire s'attachent à rechercher la solution

optimale à un problème donné en faisant évoluer un seul point à chaque itération, les algorithmes génétiques travaillent sur une population de chromosomes qui évoluent au cours des itérations. Un chromosome représente en fait un point de l'espace d'état, auquel est associée une valeur (également désignée sous le nom d'adaptation) permettant de situer ce chromosome (ou individu) dans l'espace objectif. Le codage d'un chromosome dépend de la modélisation retenue ; chaque chromosome est composé d'un nombre déterminé de gènes, dont chacun correspond à une variable et auquel on associe une valeur qui évoluera au cours du calcul.

Lorsque le principe de codage des chromosomes est défini, un algorithme génétique se déroule selon le schéma suivant :

1. génération aléatoire de la population initiale, afin d'assurer la plus grande homogénéité des points dans l'espace d'état ;
2. évaluation des chromosomes afin d'associer à chacun d'eux une valeur d'adaptation (ou plusieurs dans le cas de l'optimisation multi-objectifs) ;
3. sélection des individus pour la reproduction ;
4. croisement des individus sélectionnés selon une certaine probabilité (probabilité de croisement) ;
5. mutation des chromosomes (selon la probabilité de mutation) ;
6. retour à l'étape 2 tant que le critère d'arrêt n'est pas atteint.

Chaque séquence constituée des étapes 2 à 5 est appelée une génération. Le nombre de générations à effectuer est un critère d'arrêt fréquemment utilisé.

Par la suite, on supposera que l'on considère un problème de maximisation dans un ensemble de solution à valeurs positives, sachant que l'on peut ramener toute fonction à ce type de caractéristique.

Le croisement a pour objectif d'augmenter la diversité de la population en manipulant la structure des chromosomes. Nous nous cantonnons ici au principe usuel de croisement à découpage de chromosomes (*slicing over*) appliqué à deux parents et permettant d'obtenir deux enfants.

L'opérateur de mutation confère aux algorithmes génétiques la propriété d'ergodicité de parcours d'espace : l'algorithme est ainsi susceptible d'atteindre tous les points de l'espace d'état. De même que pour le croisement, une probabilité est associée à cet opérateur ; les valeurs généralement conseillées sont de l'ordre de 0,5 à 0,8 pour la probabilité de croisement, et de $1/N$ ($N =$ nombre de chromosomes) à 0,1 pour la probabilité de mutation.

Les algorithmes génétiques dans les problèmes d'ordonnancement [1]

La complexité élevée des problèmes d'ordonnancement a conduit à utiliser les algorithmes génétiques pour leur résolution ; cependant, ces problèmes présentent des caractéristiques et des contraintes particulières en ce qui concerne l'ordre des permutations et le respect des contraintes associées lors d'opérations sur les chromosomes.

Pour mémoire, on peut citer les croisements suivants, développés spécifiquement pour les problèmes d'ordonnancement :

- croisement PMX (*Partially Mapped Crossover*) : application aux problèmes de job-shop et de voyageur de commerce ;
- croisement MPX (*Maximal Preservative Crossover*) : application au problème du voyageur de commerce ;
- croisement par recombinaison d'arêtes : application au problème du voyageur de commerce symétrique ;
- croisement d'ordre (à deux points, à deux points linéaires, uniforme) ;
- croisement de cycles : application au problème du voyageur de commerce ;
- croisement simple à un l point (à k points) pour les permutations.

Les applications pratiques portent sur plusieurs types de problèmes, et notamment :

- problème de voyageur de commerce, qui existent dans le monde industriel sous la forme de minimisation du déplacement d'un outil ou du temps de changement d'un outil ;
- problèmes de job-shop
- ordonnancement de lots dans un flow shop (int J of Prod res C. Jordan).

Cependant, les développements théoriques sur la convergence des algorithmes génétiques restent limités ; ils permettent toutefois d'obtenir des bonnes performances dans un grand nombre de cas, et notamment lorsque le

codage et les opérateurs génétiques adaptés aux connaissances du problème sont utilisés.

Codage du problème

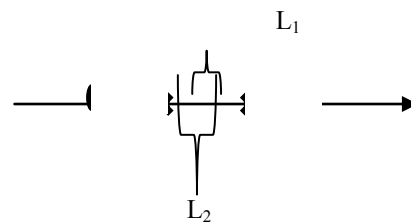
Dans le cadre du problème étudié, les gènes des chromosomes représentent les valeurs :

- des dates de livraison, dans le cas de la détermination de la date optimale ;
- des cycles de chaque poste, dans le cadre de la modulation de cycle ;

Il existe pour chacune de ces valeurs des contraintes sur lesquelles nous reviendrons.

La date de livraison d'un OF L_i est comprise entre des bornes que l'on peut définir comme suit (chaque intervalle L_1 et L_2 étant associé à un OF i , l'indice étant omis par souci de clarté) :

- intervalle de livraison L_1 souhaité par le client, à l'intérieur duquel devra avoir lieu la livraison sous peine de pénalités ;
- intervalle L_2 ($L_1 \subseteq L_2$) souhaité par l'industriel.



L'existence de l'intervalle L_2 peut être interprétée comme suit :

- si $L_i \in L_{2A}$, l'OF sera terminé avant que le client n'en prenne livraison, il y aura donc une période de stockage du produit fini, et par conséquent des frais financiers liés à l'immobilisation du capital ;
- si $L_i \in L_1$, l'industriel aura satisfait les souhaits du client ;
- si $L_i \in L_{2B}$, la livraison aura lieu en retard et des pénalités proportionnelles au retard seront dues au client.

Pour toute solution au problème de détermination des dates de livraisons, on devra donc avoir $L_i \in L_2$.

Le cycle d'un OF i sur un poste j est désigné par $c_{i,j}$. Le problème de détermination des cycles peut être envisagé selon deux possibilités :

- choix entre bornes : les cycles autorisés peuvent être choisis de manière continue entre deux bornes minimum et maximum ;
- choix dans liste : les cycles autorisés sont discrétisés et doivent donc être choisis dans une liste prédéterminée.

A partir des valeurs des cycles et des dates de livraison, on calcule :

1. la planification au plus tard des ordres de fabrication, en pénalisant (au niveau de la fonction objectif) les solutions qui ne respectent pas les capacités des postes ou en créant des tampons de stockage pour respecter ces capacités selon les cas (voir ci-dessous : AG mono-objectifs) ;
2. le plan de charge des effectifs, c'est à dire la courbe de besoin des effectifs en fonction du temps ;
3. les différents indicateurs relatifs à la solution, dont certains peuvent tenir le rôle de fonction objectif.

La prise en compte de contraintes, notamment celles liées aux intervalles de tolérance des dates de livraison, aux cycles autorisés ainsi qu'aux liens logiques entre tâches, est assurée par l'intermédiaire de pénalisations associées à la violation de ces contraintes : la fonction objectif est pénalisée (càd majorée dans le cas d'un problème de minimisation, et réciproquement) d'une quantité proportionnelle à la violation de la contrainte. La valeur du coefficient de pénalisation a dans ces conditions des conséquences notables sur la qualité des solutions obtenues (cf. [1]).

On utilise également des algorithmes de réparation (cf. [1]) qui rendent admissibles une solution non admissible générée par un opérateur génétique (par exemple, rétablissement des contraintes de précédence).

Approche séquentielle – Algorithmes génétiques mono – objectifs

1

Les indices i , j et k désignent respectivement les OF, les postes et les jours (unités de temps).

Présentation des critères

Les différents critères utilisés pour évaluer les solutions sont les suivants (Cf. :« Optimisation des plans de charge pour un flow shop dans les cadre d'une production en Juste A Temps : Convergence entre besoins et capacité », Philippe Duquenne, Hervé Pingaud, Emmanuel Vidal) :

- perturbation du plan de charge : à minimiser.
- activité des opérateurs : à maximiser.
- coût global : à minimiser.

Leur expression mathématique sera présentée ci-après.

La résolution du problème est effectuée dans un premier temps selon une approche séquentielle, en décomposant le problème général en sous-problèmes de tailles réduites qui seront traités les uns après les autres. L'ordre des séquences dans cette démarche a été déterminé expérimentalement, et donne les meilleurs résultats. On procède alors comme suit :

- détermination des dates de livraison optimales ;
- détermination des cycles de production adaptés ;
- application de la modulation d'horaires ;
- application des compétences multiples.

Détermination des dates de livraison

La première étape de la résolution est la détermination de la date de livraison optimale pour chaque ordre de fabrication, à l'intérieur de l'intervalle de tolérance défini avec le client. Les valeurs des cycles sont alors des données du problème (égales pour chaque poste à la valeur minimale du cycle autorisé), elles ne seront donc pas modifiées au cours de cette étape. Le nombre de variables est alors égal au nombre d'ordre de fabrications à traiter.

On peut ainsi calculer la planification au plus tard en fonction des dates obtenues, tout en respectant les capacités de production, et en déduire le plan de charge des effectifs. A ce stade-là, on cherche à minimiser les perturbations du plan de charge, sans se préoccuper des valeurs de la fonction coût et du taux d'activité.

La démarche de calcul de la planification et du plan de charge des effectifs suit alors le pseudo-code suivant :

```
m = nombre d'OF
n = nombre de postes
Eff(i,j) = effectif sur le poste j pour l'OF i
E(k) = effectif total pour le jour k
T = durée de l'horizon de temps (en jours)

de i = m à 1 par pas de -1
  f(i,n) = L(i)
  de j = n à 2 par pas de -1 faire
    d(i,j) = f(i,j) - c(i,j)
    f(i,j-1) = d(i,j)
  fin faire
  d(i,1) = f(i,1) - c(i,1)
fin faire

de i = 1 à m faire
  de j = 1 à n faire
    de k = d(i,j) à f(i,j) faire
      E(k) = E(k) + Eff(i,j)
    fin faire
  fin faire
fin faire
```

On peut alors déterminer la valeur de la fonction d'évaluation F correspondant aux perturbations du plan de charge, puis la pénaliser si la solution proposée entraîne la violation de contraintes d'occupation des postes

(càd si deux livraisons L_i et L_{i+1} sont prévues à un intervalle inférieur au plus long des cycles de l'OF i , M_i) :

```

de k=1, T-1 faire
  F = F + |E(k+1) - E(k)|
fin faire

de i = 1 à m-1
  si  $L_{i+1} - L_i < M_i$  alors
    F = F + K * ( $L_i - L_{i+1} + M_i$ )
  fin si
fin de

```

Le coefficient K mentionné ci-dessus est le coefficient de pénalisation ; il permet de dégrader les solutions violant cette contrainte, sans toutefois les éliminer totalement, car il est essentiel lors de l'utilisation d'algorithmes génétiques de préserver, au moins durant les premières générations, des solutions « décevantes » afin de préserver la diversité de la population.

La solution obtenue est alors utilisée pour le calcul des cycles associés à chaque poste et à chaque ordre de fabrication.

Modulation de cycles

On considère maintenant que les dates de livraison obtenues lors de l'étape précédente sont figées, et l'on fait varier la durée opératoire sur chaque poste et pour chaque ordre de fabrication. Le nombre de variables est alors égal au produit du nombre d'OF par le nombre de postes de travail.

On calcule tout d'abord la planification, en respectant cette fois les contraintes de capacité :

```

f(m, n) = L(n)
de j = n à 2 par pas de -1 faire
  d(m, j) = f(m, j) - c(m, j)
  f(m, j-1) = d(m, j)
fin faire
d(m, 1) = f(m, 1) - c(m, 1)
de i = m-1 à 1 par pas de -1
  f(i, n) = min( $L(i)$ ,  $L_{i+1} - M_i$ )
  de j = n à 2 par pas de -1 faire
    d(i, j) = f(i, j) - c(i, j)
    f(i, j-1) = d(i, j)
  fin faire
  d(i, 1) = f(i, 1) - c(i, 1)
fin faire

de i = 1 à m faire
  de j = 1 à n faire
    de k = d(i, j) à f(i, j) faire
      E(k) = E(k) + Eff(i, j)
    fin faire
  fin faire
fin faire

```

On cherche alors à minimiser la fonction coût F, critère économique prenant en compte les frais liés aux stockages et aux en-cours, ainsi que le coût du travail, de l'inactivité et des perturbations liées à la modulation de cycles.

Ces deux étapes peuvent éventuellement être répétées plusieurs fois de suite afin d'obtenir une solution satisfaisante, voire stable, c'est à dire ne variant pas par l'application de l'algorithme génétique.

Résultats expérimentaux

Les graphiques ci-dessous présentent les résultats obtenus dans le cadre de la détermination des dates de livraison optimales, en prenant comme critère les perturbations du plan de charge. Ce critère doit bien sûr être minimisé ; on utilise alors la formule suivante afin de se situer dans le cadre d'un problème de maximisation :

$$F(x) = C_{Max} - C(x)$$

où est la valeur de la perturbation et C_{Max} un majorant de $C(x)$.

Les courbes ci-dessous représentent, en fonction du nombre de générations et du mode de sélection retenu :

- l'adaptation du meilleur chromosome (1A, 2A, 3A) ;
- la moyenne de l'adaptation de tous les chromosomes de la génération courante (1A, 2A, 3A) ;
- le nombre de chromosomes satisfaisant les contraintes (1B, 2B, 3B : échelle de gauche du graphique) ;
- la valeur moyenne de violation des contraintes (1B, 2B, 3B : échelle de droite du graphique).

Evolution de l'adaptation du meilleur chromosome et de la moyenne des adaptations en fonction de la génération

3A : sélection par crowding

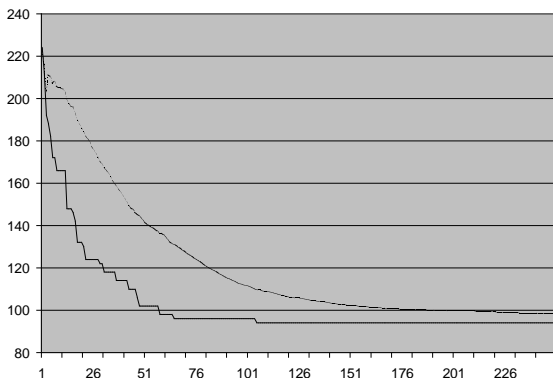
Meilleur chromosome : ———
Moyenne des chromosomes : - - - - -



1A: sélection par roulette de casino

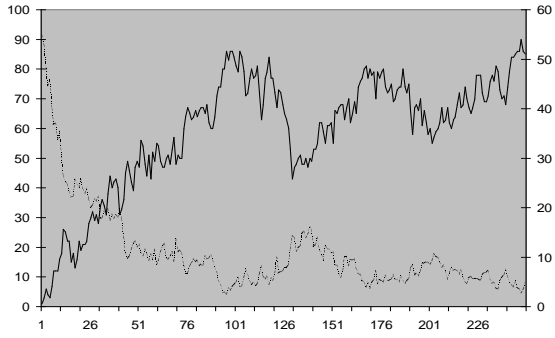


2A : sélection par division stochastique

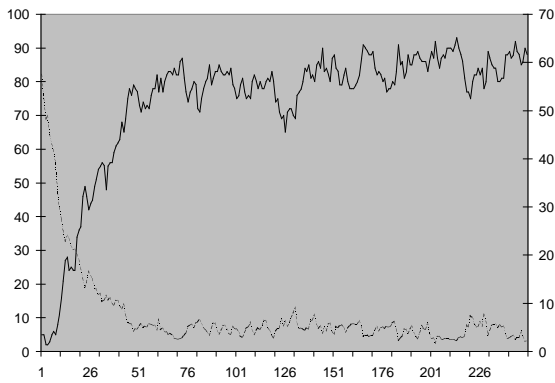


Nombre de chromosomes satisfaisant les contraintes
et valeur moyenne de la violation de contrainte
Nombre de chromosomes satisfaisant les contraintes
et valeur moyenne de la violation de contrainte

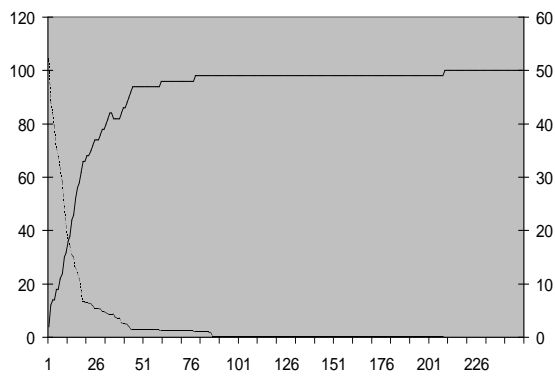
Nombre de chromosomes satisfaisant la contrainte : ———
Valeur moyenne de la violation de contrainte : - - - -



1B: sélection par roulette de casino



2B : sélection par division stochastique



3B : sélection par crowding

Les méthodes de sélection suivantes ont été testées (cf. [2], [3], [5]) :

- roulette de casino ;
- division stochastique ;
- *crowding* (comparaison entre parents et enfants à chaque génération, afin de préserver les meilleurs individus entre deux générations successives).

Les meilleurs résultats sont obtenus avec la division stochastique et le *crowding*, qui permettent de préserver les bons individus à chaque génération : ces deux principes de sélection sont ceux qui permettent d'obtenir le plus souvent la meilleure solution, en gardant un nombre raisonnable de chromosomes et de générations.

Pendant, en examinant les résultats obtenus pour un grand nombre d'exécutions d'une même occurrence du problème, on constate qu'un nombre important de solutions conduisent à des valeurs d'adaptation peu différentes : si l'on ne s'intéresse qu'aux chromosomes dont l'adaptation est au moins égale à 95% de la solution optimale, on obtient onze chromosomes différents.

Dans le cas de la modulation de cycles, les courbes ci-dessous montrent que la convergence est lente, aussi bien pour la courbe de la meilleure valeur que pour la courbe de la valeur moyenne ; en examinant en détail les résultats, on remarque qu'un nombre important de solutions correspondent à des valeurs de la fonction coût peu différentes, mais qu'elles présentent une grande diversité en ce qui concerne les valeurs de taux d'occupation et de taux de variation des effectifs. Cette remarque nous conduit alors à envisager une approche multi-critères qui permettrait de déterminer les solutions pour lesquelles les trois critères présentent simultanément des valeurs satisfaisantes. Le développement de cette démarche est décrit dans le paragraphe suivant.

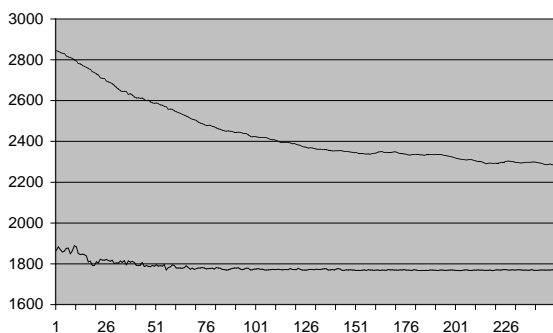


Fig 4 : Evolution du meilleur individu et de la moyenne dans la modulation de cycles

Approche globale – Algorithmes génétiques multi – critères

Les algorithmes génétiques multi – objectifs ([3], [4]) permettent d'aborder des problèmes pour lesquels on cherche à améliorer simultanément plusieurs critères, qui peuvent parfois être contradictoires ; on peut ainsi chercher à maximiser un gain tout en minimisant un coût. Une méthode consiste à pondérer les différents critères, mais le résultat est alors très sensible aux valeurs des coefficients. Pour une approche globale, indépendante de ce type de pondération, on utilise alors la notion d'optimum de Pareto.

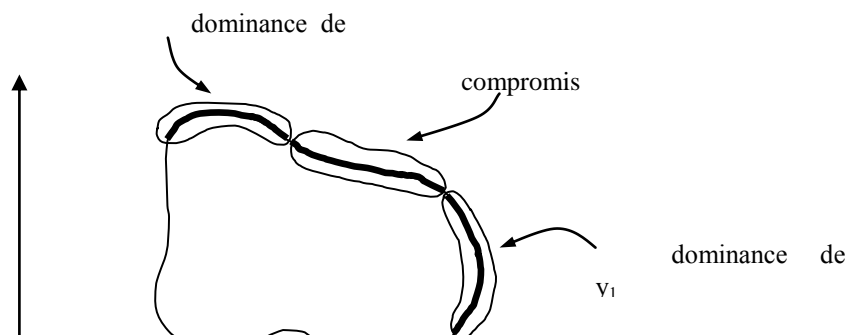
Notion d'optimum de Pareto

Soient y_i , les différents critères de la fonction multi – objectifs à optimiser ; on appelle surface de Pareto la limite du domaine telle qu'aucune solution ne permette d'améliorer simultanément tous les critères de la fonction.

Dans le cas d'une fonction à deux objectifs, on obtient alors un arc séparant le plan objectif en deux sous-domaines :

- l'ensemble des points qui ne correspondent à aucune solution ;
- l'ensemble des points correspondants à une solution du problème.

Les points de cet arc sont donc des solutions au regard de cette notion d'optimum de Pareto. On remarquera que cet arc correspond à la trajectoire décrite par $\alpha*y_1+(1-\alpha)*y_2$ lorsque α varie entre 0 et 1.



Dans le cas de deux objectifs, on retient généralement trois zones (cf graphique ci-dessous) :

- dominance de y_1 (α proche de 1) ;
- compromis (aucun critère n'émerge par rapport à l'autre) ;
- dominance de y_2 (α proche de 0).

Principe de sélection dans les algorithmes génétiques multi-objectifs

Les principes de sélection présentés ci-dessus ne peuvent plus s'appliquer puisque l'on prend maintenant en compte plusieurs valeurs correspondant à un seul chromosome. On applique donc un nouveau principe de sélection, mettant en avant la notion de dominance.

On tire tout d'abord au hasard deux individus dans la population I_1 et I_2 ; puis on tire au hasard une sous-population SP appelée à servir de référence, dont la taille est déterminée par un paramètre nommée t_{dom} . trois cas sont alors possibles :

- I_1 est dominant par rapport à SP, c'est à dire qu'il domine tous les éléments de SP, et I_2 n'est pas dominant par rapport à SP : I_1 est sélectionné ;
- I_2 est dominant par rapport à SP, c'est à dire qu'il domine tous les éléments de SP, et I_1 n'est pas dominant par rapport à SP : I_2 est sélectionné ;
- I_1 et I_2 sont tous les deux dominants ou tous les deux non-dominants ; on pourrait alors choisir aléatoirement l'un des deux individus, mais on risque alors de limiter la répartition des individus à une portion réduite de la surface de Paréto. On cherche donc à sélectionner l'individu situé dans la région la moins peuplée de l'espace objectif (partage, ou *sharing* dans l'espace objectif).

Ce partage ne modifie pas les critères, mais sélectionne les individus en fonction de la répartition dans l'espace objectif. Le but est de favoriser la reproduction d'individus appartenant à des régions où la densité de population est faible. On peut également envisager de réaliser ce partage dans l'espace d'état, mais on observe généralement de moins résultats.

Le pseudo-code ci-dessous explique schématiquement le processus suivi (en prenant comme critère d'arrêt le nombre de générations, qui est un paramètre du problème) :

Génération de la population initiale
Evaluation de la population initiale

Tant que (critère d'arrêt) est Vrai Faire
Sélection des chromosomes (selon principe ci-dessus)
Croisement et mutations des individus sélectionnés
Evaluation des individus
Comparaison des parents et de leurs enfants :
 Détermination des paires parents enfants les plus proches
 Dans chaque paire, sélection de l'individu dominant
 A défaut d'individu dominant, sélection de l'individu situé dans la région la moins dense.

Fin tant que

Les graphiques (Figures 5A à 5C) ci-dessous représentent l'évolution des individus entre la population initiale et la génération finale (sur 50 générations), évalués selon les trois objectifs :

- Minimisation de la fonction coût ($F1$) ;
- Maximisation du taux d'activité ($F2$) ;

- Minimisation du taux de variation (F3).

Chaque graphique représente les points dans un projection de l'espace objectif, où chaque axe représente un des objectifs. Les résultats intéressants sont donc situés dans les zones suivantes :

- F2 en fonction de F1 (Fig 5A) : portion supérieure gauche ;
- F3 en fonction de F1 (Fig 5B) : portion inférieure gauche ;
- FF3 en fonction de F2 (Fig 5C) : portion inférieure droite.

Les résultats obtenus confirment les capacités de convergence de l'algorithme vers les zones où les trois critères atteignent simultanément des valeurs satisfaisantes. De plus, on observe donc une convergence rapide des chromosomes : alors que la population initiale occupe l'espace objectif de manière homogène, la population se concentre au fur à mesure vers une zone limite connue sous le nom de surface de Pareto, et séparant les points correspondants à des solutions des points auxquels aucune solution ne peut être associée.

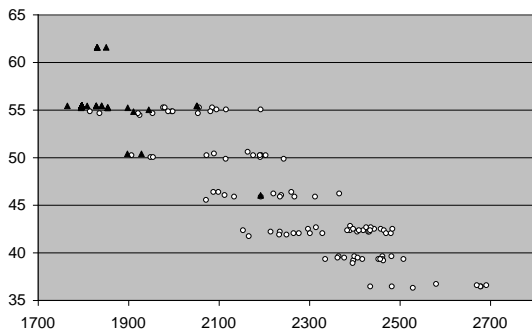


Fig 5A : F2 en fonction de F1

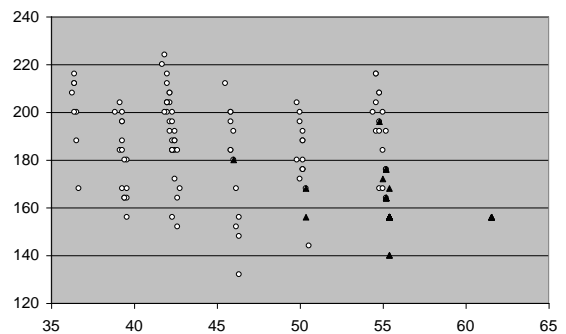


Fig 5B : F3 en fonction de F2

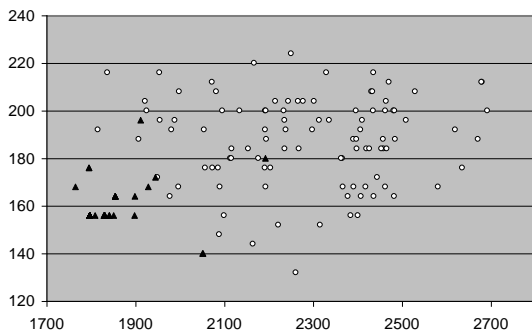


Fig 5C : F3 en fonction de F1

Conclusion:

Bien que les algorithmes génétiques ne présentent aucune garantie quant à l'obtention d'une solution optimale, ils permettent de trouver des solutions proche de l'optimum dans des temps de calcul réduits ; de plus, ils se montrent particulièrement adaptés pour des problèmes caractérisés par leur aspect hautement combinatoire ainsi que le manque de propriétés mathématiques de leur fonction objectif, nécessaire pour appliquer les outils classiques de programmation mathématique.

Dans le présent exemple, leur aptitude à résoudre des problèmes d'optimisation multi-critères sont mis en évidence. En outre, la résolution d'un tel problème par ces méthodes ne requiert ni l'augmentation du nombre de chromosomes traités à chaque génération, ni l'accroissement du nombre de génération à effectuer ; l'obtention d'un ensemble de solutions, proches de la surface de Pareto, n'a donc pas d'influence sensible sur le temps de calcul et permet de proposer aux décideurs un plus large éventail de solutions, desquelles pourront être éliminées celles qui ne respectent pas d'éventuelles contraintes supplémentaires.

Références bibliographiques

Agrawal A., Harhalkis G., Minis I., Nagis R., "*Just-in-time production of large assemblies*", IEEE Transactions (1996), 28, pp 653-667

- Benhouba-Zelfani F. Z., Zelfani M., Caillaud E., Duquenne P., Lacoste G., "*Applications intégrées et outils de simulation et d'aide à la décision*", deuxième congrès international franco-québécois de Génie Industriel, Albi (1997)
- Blondel, F., "*Gestion de la production*", Dunod, 1997
- Chvidchenko I., Chevalier J., "*Conduite et gestion de projets, principes et pratique pour petits et grands projets*", Cepadues, 1993
- Esquirol P., Lopez P., "*L'ordonnancement*", Economica, 1999
- Jagou P., "*Concurrent Engineering, la maîtrise des coûts, des délais et de la qualité*", Hermès, 1993
- Maurino M., "*La gestion des données techniques, technologie du concurrent engineering*", Masson, 1993
- Rosenblatt M. J., Hau L. L., "*The effects of work-in-progress inventory costs on the design and the scheduling of assembly lines with low throughput and high component costs*", IEEE Transactions (1996), 28, pp 405-414
- Westney R. E., "*Gestion de petits projets, techniques de planification, d'estimation et de contrôle*", AFNOR Gestion,, 1991

Bibliographie :

- [1] Caux C., Pierreval H., Portmann M. C., « Les algorithmes génétiques et leur application aux problèmes d'ordonnancement », *Journées d'étude « Ordonnancement et entreprise : Applications Concrètes et Outils pour le Futur » Preselction Revisited, Toulouse, Juin 1994*
- [2] Davis, L., « Handbook of Genetic Algorithms », Van Nostrand Reinhold, 1991
- [3] Delahaye D., « Optimisation de la sectorisation de l'espace aérien par algorithmes génétiques », Thèse 1995
- [4] Horn, J., Nafpliotis, N., Goldberg, D.E., « A niched Pareto genetic algorithm for multiobjective optimization », In *Proceedings of the First IEEE Conference on Evolutionary Computation, IEEE World Congress on Computational Intelligence, Volume 1, 1994*
- [5] Mahfoud, S.W., « Crowding and Preselection Revisited », In *Parallel Problem Solving From Nature, 2 (PPSN 2)*, 27-36, 1992