



Modélisation et décomposition des problèmes de transbordement dynamiques : application a la répartition des wagons Fret-SNCF

Marie-Hélène Morin

► To cite this version:

Marie-Hélène Morin. Modélisation et décomposition des problèmes de transbordement dynamiques : application a la répartition des wagons Fret-SNCF. Modélisation et simulation. Université Joseph-Fourier - Grenoble I, 1993. Français. NNT : . tel-00343876

HAL Id: tel-00343876

<https://theses.hal.science/tel-00343876>

Submitted on 3 Dec 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

70 80186

THESE

Présentée par

Marie-Hélène MORIN

Pour obtenir le titre de

DOCTEUR

(arrêtés ministériels du 5 Juillet 1984 et du 30 Mars 1992)

de

L'UNIVERSITE J. FOURIER - GRENOBLE I

Spécialité Recherche Opérationnelle

Modélisation et décomposition des problèmes de transbordement dynamiques : application à la répartition des wagons Fret/SNCF

Soutenance le 09 Avril 1993

devant la commission d'Examen

Mr **FINKE**
Mr **DEJAX**
 FONLUPT
 MAHEY
 MAISONNEUVE

Président
Rapporteurs

Examineurs

*Thèse préparée au sein du laboratoire ARTEMIS. Unité de recherche INPG UJF, associée au
CNRS (URA 396)*

. A Nicolas et Dominique

***. A mes parents en témoignage de
mon affection***

Avant propos

- :: -

Le travail présenté dans ce mémoire a bénéficié d'une Convention Industrielle de Formation par la Recherche et, à ce titre, a été réalisé au sein du Département Informatique Fret de la société SNCF, en collaboration directe avec le groupe "Recherche Opérationnelle" du laboratoire ARTEMIS de l'Institut de Mathématiques Appliquées de Grenoble.

Je tiens à remercier Madame le Professeur D. Borionne, Directeur de ce laboratoire, de m'avoir accueillie dans son équipe de recherche.

Je remercie également Monsieur le chef du Département Informatique Fret J.L. Martin, de m'avoir confié ce travail de recherche et de développement.

Monsieur G. Finke, Professeur à l'Université J. Fourier - Grenoble I, a accepté de s'intéresser à mes travaux et d'être Président du jury de thèse. Je l'en remercie très sincèrement.

Je suis particulièrement honorée par la présence à ce jury de Monsieur le Professeur P. Dejax, de l'Ecole Centrale - Paris. Je lui suis reconnaissante de l'intérêt qu'il a bien voulu porter à ce travail en tant que rapporteur.

Je tiens à exprimer toute ma gratitude à Monsieur le Professeur J. Fonlupt pour l'attention qu'il a accordée à ce travail et pour sa présence en tant que rapporteur, dans ce jury.

Je voudrais que Monsieur A. Maisonneuve, chef de la Division IFM, veuille bien trouver ici l'expression de ma profonde reconnaissance pour n'avoir cessé de me prodiguer conseils et encouragements tout au long de cette étude.

Ce travail doit beaucoup à Monsieur P. Mahey, Maître de Conférences à l'Université J. Fourier Grenoble I et Directeur de recherche de ce travail. Sa collaboration et ses encouragements constants m'ont permis de mener à bien

cette recherche. Je souhaite que ce mémoire soit l'expression de mon amitié et de ma profonde reconnaissance.

Je voudrais enfin exprimer toute ma reconnaissance à Monsieur J. Daniel, de la Division IFM, pour toute l'aide et les précieux conseils qu'il m'a apportés dans la mise au point du logiciel.

Table des matières

<u>INTRODUCTION</u>	13
<u>CHAPITRE I - PRESENTATION DU PROBLEME</u>	19
I.1 - L'acheminement d'un wagon	21
I.1.1 - Le cycle de vie d'un wagon	21
I.1.2 - Les principes de l'acheminement	23
I.2 - MARS	27
I.2.1 - MARS ... "qu'est ce que c'est ?"	28
I.2.2 - MARS ... " pourquoi ? "	28
I.3 - La Répartition	29
I.3.1 - Les limites de l'organisation actuelle	29
I.3.2- La Répartition automatisée	31
I.4 - Les données et les règles du problème	36
I.4.1 - Les données	36
I.4.2 - Les règles de gestion	40
I.4.3 - Respect des règles par catégorie	41
<u>CHAPITRE II - APPROCHE GLOBALE</u>	43
II.1 - Le problème de transport	45
II.1.1 - Historique du problème	46
II.1.2 - Classification de différentes méthodes	49
II.1.3 - Choix d'une méthode	52
II.2 - Planification et prévisions	59
II.2.1 - Gestion à court terme	59
II.2.2 - Planification sur un horizon glissant	61
II.2.3 - Prévision des besoins et des ressources	64
II.3 - Le modèle SNCF	70
II.3.1 - Modélisation du noyau	70

II.3.2 - Algorithme Répartition (n° 1)	74
II.3.3 - Résultats numériques	76
II.4 - Flexibilité de l'horizon	81
II.4.1 - Algorithme Répartition (n° 2)	81
II.4.2 - Résultats numériques	83
II.5 - Extension du modèle SNCF	87
II.5.1 - Modélisation du problème complet	88
II.5.2 - Algorithme Répartition (n° 3)	89
II.5.3 - Les limites de l'approche globale	90
 <u>CHAPITRE III - APPROCHE PAR DECOMPOSITION</u>	 91
III.1 - Introduction	93
III.1.1 - Objectifs des méthodes de décomposition	93
III.1.2 - Approche conceptuelle du problème	94
III.1.3 - Transformation du problème initial	95
III.1.4 - Classification de différentes méthodes	98
III 1.5 - Structure du système SNCF	102
III 1.6 - Notations	104
III.2 - Décomposition duale	106
III.2.1 - Relaxation des contraintes de couplage	107
III.2.2 - Algorithme de sous-gradient	109
III.3 - Décomposition primale	111
III.3.1 - Variables de couplage	112
III.3.2 - Algorithme de sous-gradient	114
III.4 - Décomposition mixte	117
III.4.1 - Dégénérescence des sous-problèmes	118
III.4.2 - Algorithme de décomposition mixte	125
III.5 - Résultats numériques	131

<u>CONCLUSION</u>	137
<u>ANNEXES</u>	141
Annexe I : Extrait du rapport Eureco	143
Annexe II : Langage CHARME	147
Annexe III : Programmation Linéaire	151
Annexe IV : Données numériques	155
<u>BIBLIOGRAPHIE</u>	157

Introduction

Introduction

Dans la réalisation technico-commerciale du transport, *la répartition des wagons vides* participe fortement aux coûts de production, à la qualité et à la consistance du service offert. Cette notion, par ses intérêts économiques et fonctionnels, permet à l'entreprise, de progresser dans la qualité du service offert et au personnel concerné, de bénéficier d'un outil performant et adapté à leur travail. C'est pourquoi les enjeux sont considérables dans l'environnement extrêmement concurrentiel du transport de marchandises.

La répartition est bâtie une gestion dynamique du parc de wagons. Son *objectif fondamental* est d'assurer à la clientèle la mise à disposition de wagons vides, sur les lieux et aux dates demandés, en fonction de critères économiques (délais et coûts d'acheminement) et de ressources disponibles.

Une durée de rotation trop longue entre deux chargements successifs est due à diverses causes telles que les immobilisations chez les clients, la technique ferroviaire, les effets saisonniers et les temps de déplacement à vide. Cette notion donne une idée de l'enjeu économique lié à une gestion performante des répartitions : celle-ci permet, en effet, de réduire les coûts d'immobilisation et les frais d'exploitation liés aux parcours à vide.

Quelques chiffres, proposés par la SNCF en 1991, permettent de situer le problème:

- nombre de wagons de marchandises du parc : 85.000
- valeur du parc de wagons : 14 Milliard de francs.
- nombre de chargements moyens d'un wagon par an : 19
- quantité de marchandises transportées par an : 141 Millions de tonnes.
- pourcentage de wagons vides dans les trains : 39 %.
- parcours quotidiens des trains de marchandises : 420.000 km.

La nouvelle application répartition de la SNCF, s'intègre de façon naturelle dans le prolongement de la refonte de l'informatique Fret. Les économies porteront notamment sur le matériel. Le nombre de wagons du parc devrait se stabiliser autour de 50.000, vers 2010, grâce à une meilleure utilisation des wagons.

Le modèle répartition, des wagons de marchandises sur le réseau ferroviaire français, doit tenir compte de la satisfaction des demandes, en chaque point (gares) et à chaque période de temps, tout en minimisant les coûts de transport. La complexité de ce problème vient de sa taille, de ses aspects dynamiques et stochastiques et du coût élevé du calcul des délais d'acheminement. Désireux de se placer à un niveau de décision à moyen terme prenant en compte l'aspect prévisionnel de la demande (typiquement sur un horizon de 7 jours) et contraints de limiter les temps de réponse du programme (destiné à être actionné deux fois par jour), nous avons choisi de modéliser le problème comme un problème de transbordement dynamique.

Trois aspects principaux conditionnent l'explosion numérique du modèle :

- l'aspect dynamique, qui multiplie les données du réseau sur chaque période. Un modèle stochastique est nécessaire si on veut prendre en compte les demandes prévisionnelles sur tout l'horizon. On se reportera aux travaux de Jordan et Turnquist [Jor,1983], Powell [Pow,1987] ou Crainic et al [Cra,1993] pour de plus amples détails.

- l'aspect multiproduits : l'ensemble des 85.000 wagons sont répartis en 210 catégories qui, par le jeu des substitutions autorisées sous conditions, ne sont pas indépendantes (cf. [Cra,1993]).

- l'aspect géographique : le réseau français est un réseau complexe comportant de l'ordre de 2.500 gares avec une structure hiérarchisée (300 centres de répartition, 32 sous-zones régionales et 3 superzones de trafic, la région parisienne jouant le rôle de pôle attracteur) et hétérogène (certains produits, par exemple, circulant du nord vers le sud impliquent de retours à vide).

L'objectif principal de ce travail est d'adapter une *méthode de décomposition* au modèle global de transbordement dynamique. Nous proposons une décomposition temporelle du problème, puis une décomposition par catégorie de matériel et enfin une décomposition spatiale :

- le choix de la taille et du découpage de l'horizon de travail est un facteur déterminant pour résoudre approximativement le problème par étape. Une approche heuristique exploitant le découpage temporel est décrite dans cette étude. Des

conditions de consistance peuvent être rajoutées pour ne résoudre le problème que sur la première période (cf. Lasserre [Las,1989]).

- un modèle de type multiflots est nécessaire pour traiter les substitutions entre catégories (cf. Gondran et Minoux [Gon,1979] et Crainic et al [Cra,1993]).

- un découpage en zones géographiques permet de ne pas traiter directement l'ensemble des données. On résout alors des sous-problèmes locaux de taille plus raisonnable et on gère itérativement le trafic inter-zone jusqu'à satisfaire certaines conditions d'optimalité globales. La méthode choisie est une méthode de décomposition mixte (Mahey [Mah,1986]) qui utilise des allocations à la fois primales et duales en chaque point frontières par où s'effectue le trafic inter-zone.

La première partie de ce rapport présente les objectifs de la nouvelle application répartition, l'organisation projetée et l'architecture du système.

La deuxième partie est centrée sur la description et la validation d'un modèle destiné à la SNCF et mis en service depuis le mois de Juin 1992.

Nous présentons un historique des problèmes de transport, ainsi qu'une comparaison des différents modèles existants. Le modèle de base qui est proposé s'appuie sur un problème de transbordement et l'ensemble du produit réalisé prend en compte, sur cette base, toutes les contraintes du problème de la répartition, notamment les aspects dynamiques et prévisionnels.

Nous proposons ensuite une méthode pour traiter des problèmes de grandes dimensions en décomposant l'horizon de planification en n sous-horizons suivant les cas. Cette approche heuristique est efficace en terme de temps d'exécution d'un traitement.

Nous proposons enfin un modèle très général pour traiter toutes les demandes de toutes les catégories de wagons en une seule étape. Seulement, le nombre de contraintes et de variables générées ne permet plus de traiter le problème dans son ensemble avec le même algorithme.

La troisième partie a pour objet l'étude de problèmes de grandes dimensions. Le problème de la répartition constitue un cadre privilégié pour l'application de méthodes de décomposition. L'intérêt de la structure du système SNCF réside dans le fait que chaque sous-système se présente comme un problème de transbordement étudié dans la deuxième partie de ce rapport.

Nous introduisons cette partie en présentant des problèmes décomposables et une classification de différentes méthodes.

Nous proposons ensuite deux méthodes de résolution, classiques et duales l'une de l'autre, adaptées aux Programmes Linéaires de grandes dimensions. Ces méthodes sont nécessaires à la conception d'un l'algorithme Primal-Dual. Cette dernière méthode conduit à une décentralisation complète des décisions, sans le concours d'un niveau supérieur de coordination.

Différentes études ont été menées à la SNCF pour définir la méthode la plus adaptée au problème de la répartition. Dans un premier temps, la société EURECO a réalisé une étude diagnostique de la répartition et mené une étude comparative entre deux approches, l'Intelligence Artificielle et la Recherche Opérationnelle (*annexe 1*). Dans un deuxième temps, la société BULL a proposé un langage déclaratif d'expression de contraintes, le langage CHARME (*annexe 2*).

Présentation du problème

Chapitre I

I.1 - L'acheminement d'un wagon

I.1.1 - Le cycle de vie d'un wagon

Ce premier point concerne la vie d'un wagon FRET à la SNCF, pour situer fonctionnellement l'acheminement et le séjour des wagons.

Deux types d'acheminement

En vue de répondre à la clientèle, deux modes d'acheminement sont prévus. On distingue l'acheminement par *trains entiers* pour les transports massifs, sans remaniement en cours de route et ne posant pas de problèmes de gestion, et l'acheminement par *wagons isolés* pour les envois diffus.

Cycle de vie d'un wagon

Le cycle de vie d'un wagon est ponctué par un certain nombre d'évènements définissant ses principales fonctions qui sont l'acheminement et le séjour.

Lorsqu'un client désire effectuer un transport de FRET avec la SNCF, il commande un wagon *vide*. C'est l'application *répartition* qui est chargée de le lui fournir.

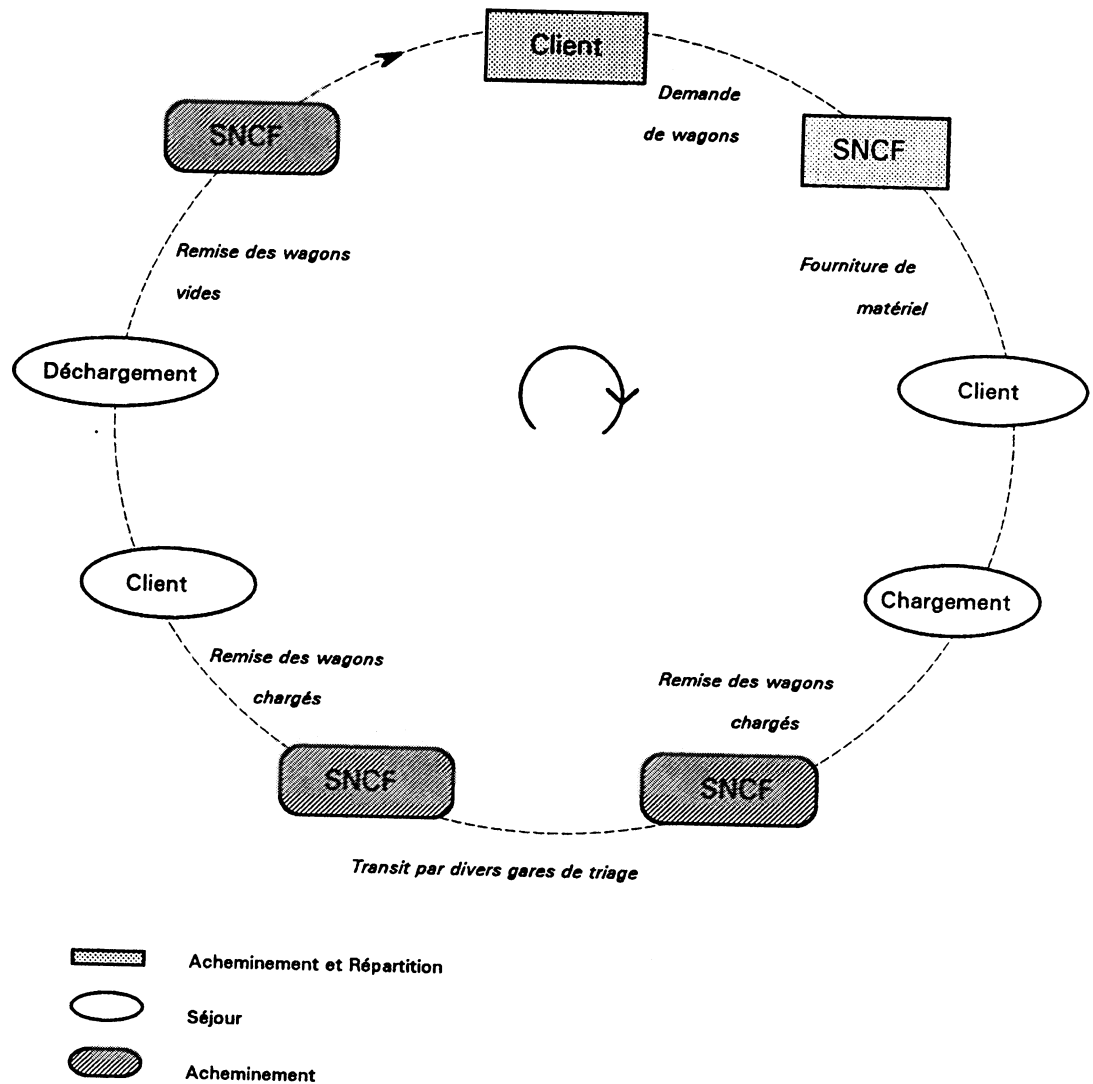
Le chargement du wagon par le client est géré par l'*application séjour*.

Dès que le wagon est remis chargé à la SNCF, l'application *acheminement* gère le wagon jusqu'à la livraison au client destinataire.

Le déchargement par ce dernier jusqu'à la remise à disposition du wagon à la SNCF est géré par l'*application séjour*.

La répartition attend alors qu'un autre client fasse une demande de transport pour fournir à nouveau ce wagon. Cette fourniture fera l'objet d'un acheminement à *vide*.

Tout mouvement commercial génère presque automatiquement un mouvement technique à vide entraînant des déséquilibres régionaux entre les demandes de wagons vides d'un type donné et les ressources du même type. Ces mouvements à vide occupent une place déterminante dans les opérations de transport de marchandises, par l'impact considérable qu'ils peuvent avoir sur la qualité du service et sur les ressources qu'ils mobilisent; c'est pourquoi il est nécessaire de les contrôler par une planification des opérations. C'est une des caractéristiques des modèles que nous développons.



- cycle de vie d'un wagon

I.1.2 - Les principes de l'acheminement [Sncf, 1990]

Dans ce deuxième point, nous énonçons les principes généraux de l'acheminement des wagons FRET SNCF.

Le trajet d'un wagon

Le trajet d'un wagon, d'une gare-lieu à une autre gare-lieu, s'effectue en trois temps.

Un parcours de ramassage :

Les wagons sont enlevés au départ des lieux (clients ou SNCF) par des dessertes et sont acheminés vers la première gare de triage.

Un parcours inter-triage :

En fonction des destinations et des critères d'incorporation d'un wagon dans un train, les wagons sont regroupés en lots (ensemble indissociable de wagons). Ces lots empruntent une succession de trains jusqu'à la dernière gare de triage.

Un parcours de distribution :

Les wagons reçus du triage sont distribués sur les lieux (clients ou SNCF) destinataires par des trains de dessertes.

Le Plan de Transport

Le Plan de Transport est un ensemble de dispositions destinées à organiser l'acheminement des wagons.

Pour élaborer un Plan de Transport, il faut avoir une connaissance fine de l'importance et de la nature du trafic des services passés. Il faut déterminer le nombre, les charges et le régime de circulation des trains par relation. Enfin, il faut déterminer le contenu des trains en lots, les correspondances dans les gares d'escales et les critères d'incorporation d'un wagon dans un lot (critères techniques : vitesse et masse par essieux; critères commerciaux : nature de la marchandise et client destinataire; critères de transport : code gare destinataire, etc).

ATHEO, marge et niveau d'acheminement

Le Plan de Transport permet de déterminer un Acheminement THEOrique (ATHEO) entre deux gares à une date donnée.

Malheureusement, la réalité est parfois différente. En effet, on observe des fluctuations de trafic qui étaient imprévisibles lors de l'élaboration du Plan de Transport. Pour réguler de manière économique ces fluctuations, la SNCF a mis au point un système de marges et de niveaux d'acheminement ainsi qu'un produit commercial correspondant à une prestation. Cette prestation garantit au client une date-heure d'arrivée des wagons. On distingue actuellement trois niveaux d'acheminement :

Acheminement en niveau 1 :

Le wagon n'a pas de marge, car la date d'arrivée garantie à destination est égale à la date d'arrivée théorique calculée par l'ATHEO.

Acheminement en niveau 2 :

Le wagon a 24 h. de marge, car la date d'arrivée garantie à destination est égale à la date d'arrivée théorique calculée par l'ATHEO plus 24 h.. Le coût d'acheminement est moins élevé qu'en niveau 1.

Acheminement en niveau 3 :

Le wagon a 72 h. de marge. Le coût d'acheminement est moins élevé qu'en niveau 2.

Remarque :

Le modèle répartition de la SNCF est intimement lié au Plan de Transport. Ce modèle ne comprend pas la recherche du meilleur acheminement entre deux gares du réseau. Cette étape, très importante pour l'étude, est remplacée par des calculs d'ATHEO, regroupés dans une table de délais, et des niveaux d'acheminement (cf chapitre II).

Le système informatique en place [Mar,1987]

L'application acheminement repose sur un système informatique, composé d'un *ordinateur central* (BULL - DPS 9000) localisé à Lyon, nommé Base d'Intérêt Général (BIG) et 57 *ordinateurs* (BULL - DPS 6000) répartis sur le territoire, nommés Base d'Intérêt Local (BIL).

Ces ordinateurs dialoguent entre eux via le réseau de téléinformatique RETIPAC. La BIG coordonne et enregistre les acheminements et leurs exécutions. Elle gère les domaines relatifs aux *données d'acheminement* (géographie du réseau, codes gares

etc), aux *courants de transports* (flux de trafic stables etc), au *Plan de Transport*, au *suivi des trains* en composition et en circulation et au *suivi des wagons*.

Les BIL gèrent le travail d'un triage. Elles assurent le calcul des *correspondances*, la saisie des *courants de transport* pour la BIG et la saisie des *incorporations* des wagons dans les trains.

I.2 - MARS

Le problème de la répartition se caractérise par la satisfaction de demandes de matériel vide formulées par la clientèle, dans les délais, au moindre coût et en fonction des ressources disponibles.

Toutefois, la gestion des wagons de marchandises se révèle complexe, car le réseau SNCF est très dense (2.500 gares), le parc de wagons est important (environ 85.000 wagons répartis) et diversifié (210 catégories différentes). De plus, 20.000 clients font appel plus ou moins régulièrement à la SNCF pour leur transport. On compte actuellement 6.000 demandes de wagons vides en moyenne par semaine. La fonction répartition est chargée de satisfaire ces 6.000 demandes.

L'objet de ce mémoire est d'étudier les différents aspects du problème de la répartition qui, comme nous allons le voir, est un problème difficile par sa dimension, ses multiples contraintes et son environnement dynamique.

Nous allons voir, dans ce paragraphe, où se situe la Répartition dans la gestion du transport de FRET à la SNCF et dans quelle mesure notre étude a été influencée par les contraintes de l'organisation actuelle.

I.2.1 - MARS ... "qu'est ce que c'est ?"

Le projet MARS est le dernier volet de la refonte de l'informatique FRET. Ce projet a pour but d'améliorer la mise à disposition du matériel (objectif de la REPARTITION), d'assurer un suivi plus précis et plus fiable des wagons (objectif du SEJOUR) et de permettre aux différents intervenants de mesurer les conséquences de leurs décisions en terme de gestion.

I.2.2 - MARS ... " pourquoi ? "

Le projet MARS porte sur la mesure optimale du parc de véhicules. A cet égard, quelques chiffres sont évocateurs :

- les 85.000 wagons du parc représentent une valeur de 14 milliards de francs;
- les économies prévues, sont de 1200 wagons ;
- les trains sont composés de 39 % de wagons vides ;
- un wagon prend, en moyenne, 19 chargements par an.

Ce projet porte également sur la satisfaction optimale des demandes de transport (par une meilleure adaptation des ressources aux besoins), sur la gestion optimale du parc de wagons (grâce à une meilleure rotation) et sur l'ouverture de notre système d'information à la clientèle.

I.3 - La Répartition

I.3.1 - Les limites de l'organisation actuelle

La structure actuelle

C'est une structure hiérarchisée à 3 niveaux :

Au niveau supérieur de cette structure, on trouve le "Central Répartition" localisé à Paris.

Au niveau médian, on trouve 32 zones correspondant grossièrement au domaine d'action d'un triage.

Au niveau inférieur, on trouve 307 sous-zones correspondant à un ensemble de gares.

L'organisation actuelle

Au niveau inférieur, les sous-zones sont chargées de coordonner la répartition des wagons de gare à gare puis de faire part des excédents ou des déficits de wagons constatés par catégorie, à la zone dont elles dépendent.

Au niveau médian, les zones sont alors chargées de coordonner la répartition des wagons (excédents et déficits) de sous-zone à sous-zone, puis de faire part des excédents ou déficits de wagons constatés par catégorie, au Central Répartition.

Au niveau supérieur, le "Central Répartition" est alors chargé de coordonner la répartition des wagons (en excédent et en déficit) de zone à zone.

Les limites de l'organisation actuelle

Une gestion en temps différé, dans un contexte dynamique, entraîne des anomalies et fait que l'état des ressources est parfois inexact. En effet, les traitements se font pour la plupart manuellement et les différents intervenants se contactent par téléphone. Conséquence : Un seul traitement par jour n'est pas suffisant pour prendre en compte les nouvelles ressources.

La partie informatique de la répartition a plus de 10 ans et demande à être réactualisée en accord avec la refonte de l'informatique Fret.

Les demandes de wagons vides par la clientèle à très court terme (95 % d'entre elles sont pour le lendemain) sont satisfaites grâce à des stocks de wagons disponibles à proximité des points de demande. Conséquence : elles nécessitent un parc très important pour un taux de rotation très faible, 19 chargements/an (cf. annexe).

I.3.2 - La Répartition automatisée

La Répartition ... " les objectifs "

La Répartition des wagons vides a pour mission d'assurer à la clientèle la mise à disposition de wagons vides (sur les lieux et aux dates demandés), connaissant la position géographique des ressources (susceptibles d'être envoyées en temps utile vers les lieux de chargement), en fonction de critères économiques et techniques (tels que les coûts d'acheminement et les délais d'acheminement).

La Répartition ... "l'organisation ?"

Les objectifs de MARS sont d'optimiser l'utilisation du matériel et d'offrir une qualité de prestation accrue à la clientèle. Cette finalité repose sur la mise en oeuvre d'une répartition automatisée, qui elle-même s'appuie sur une organisation à deux niveaux :

- *Au niveau supérieur* - le Central Répartition - où un traitement informatique satisfait les demandes en fonction des ressources, des acheminements et des coûts.
- *Au niveau inférieur* - les Centres de Répartition (CR) - où les tâches fastidieuses sont informatisées (détermination des ressources, amortissement des vides,...) et l'initiative demeure (réutilisation de wagons, satisfaction d'une demande tardive,...).

Répartir ... "quoi, où ?"

Sont répartissables :

- les wagons du parc commercial,
- les wagons étrangers, utilisés sans conditions particulières,
- les wagons étrangers, respectant le Règlement International (RIV).

Notons que le parc commercial SNCF est important et diversifié, car il comptabilise 85.000 wagons pour 210 catégories différentes.

La géographie :

La répartition se fait de lieux de gare à d'autres lieux de gare. Un client se trouve sur un lieu d'une gare, qui se trouve elle-même sur la zone d'action géographique d'un Centre de Répartition.

Notons que le réseau SNCF est très dense, car il comptabilise 20.000 clients pour 2.500 gares et 300 Centres de Répartition.

La répartition ... "comment ça marche ? "

Les ordinateurs :

La partie automatisée du traitement est implantée sur des ordinateurs (les BIR), installés au Central Répartition de la direction du FRET et reliés à la BIG.

Les transactions et la production des résultats sont développées aux niveaux BIL et BIG,

Le principe de fonctionnement :

Les BIL, puis la BIG, drainent vers les BIR les informations, telles que des demandes et des ressources, qui sont nécessaires au traitement de la Répartition.

Les demandes de matériels, formulées par les utilisateurs, s'expriment en nombre de wagons par catégorie, avec possibilité de substitution (entre catégories). Ce sont des demandes fermes ou des prévisions de fourniture (demandes à l'avance, faites à la suite d'informations recueillies auprès de la clientèle).

Les ressources sont déterminées par le système ou saisies par les Centres de Répartition. En retour, les BIL reçoivent, via la BIG qui les a mises en forme, les

informations, produites par le traitement centralisé et nécessaires à l'exécution de la répartition. Ces informations comprennent les ordres d'expédition de wagons vides, pour les gares expéditrices et les avis de réception, pour les gares destinataires.

Remarque :

Lorsqu'au terme du traitement, des ressources sont inutilisées, le système les "évacue" des CR où elles se trouvent en excédent, vers les CR qui en sont déficitaires. Excédents et déficits sont évalués en comparaison de seuils maxima et minima d'utilisation habituelle, définis par catégorie de matériel au niveau du CR.

Caractéristique du traitement :

La Répartition traite les demandes de matériel exprimées dans l'amplitude de temps J à J+5, et s'efforce de les satisfaire le plus tard possible par rapport aux dates qu'elles comportent, dans le respect des acheminements et des coûts optima.

Le traitement n'a pas lieu les samedis, dimanches et fêtes, mais les demandes du samedi et du dimanche sont traitées. Il est prévu de le lancer deux fois par jour, à la mi journée (12h40) et le soir (18h).

La répartition ... "avec qui ? "

Rôle du central Répartition :

Les hommes du Central Répartition ne peuvent pas intervenir sur le traitement lui-même, mais ont la possibilité d'agir avant ou après celui-ci, avec les tâches suivantes:

- surveillance du traitement et application des mesures de secours en cas de défaillance de tout ou partie du système ;
- saisie des demandes et ressources particulières ;
- attribution de priorités, en cas de pénurie de wagons ;
- intervention sur la chronologie des traitements (traiter une catégorie avant une autre) ;
- définition des équivalences entre catégories ;
- suivi de listes de wagons ;
- validation du traitement ou modification de ses paramètres pour le relancer.

Rôle des Centres de Répartition locaux :

Les Centres de Répartition assurent les relations technico-commerciales avec la clientèle, en ce qui concerne la fourniture et l'utilisation du matériel FRET.

- Les demandes des clients :

Les Centres de Répartition doivent avoir un contact étroit avec les clients, pour connaître leurs demandes à l'avance, les encourager à la programmation et les consulter en cas de non satisfaction. A défaut d'informations de leur part, les Centres de Répartition peuvent également faire des demandes prévisionnelles. Toutes les demandes sont saisies (même celles pour le jour J).

- Les ressources :

Les ressources sont saisies par les Centres de Répartition, lorsqu'elles résultent d'une entente avec la clientèle, ou déterminées par le système.

- A l'issue du traitement Répartition :

Les Centres de Répartition exécutent les ordres d'expédition de wagons vides et assurent un suivi de cette exécution. Ils sont renseignés, par les avis de réception, sur le degré de satisfaction des demandes. Ils amortissent les demandes au fur et à mesure de la livraison, ou de la mise à disposition, des wagons à la clientèle.

- Les seuils de rééquilibrage :

En outre, le CR gère les seuils minima et maxima d'utilisation habituelle des matériels sensibles pour leurs activités.

La Répartition ... "quand ? "

La répartition a été mise en service en deux temps :

- La phase 1.2 a démarré au mois de juin 1992 avec une catégorie de matériel;
- La phase 2 démarrera en avril 1993, avec une généralisation progressive à toutes les catégories de matériels.

Les utilisateurs

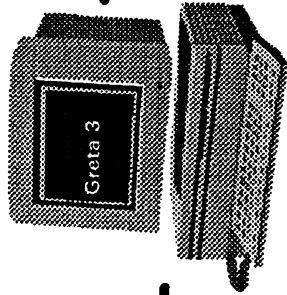
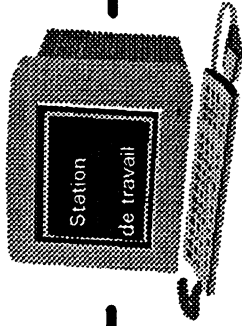
le central répartition

crée, puis met à jour
les tables des données permanentes
(triages, C.R., catégories, ...)
modifie, en cas de besoin :
les périodicités de traitement;
les seuils "mini" prioritaires.
exploite les résultats
suit les alertes éventuelles

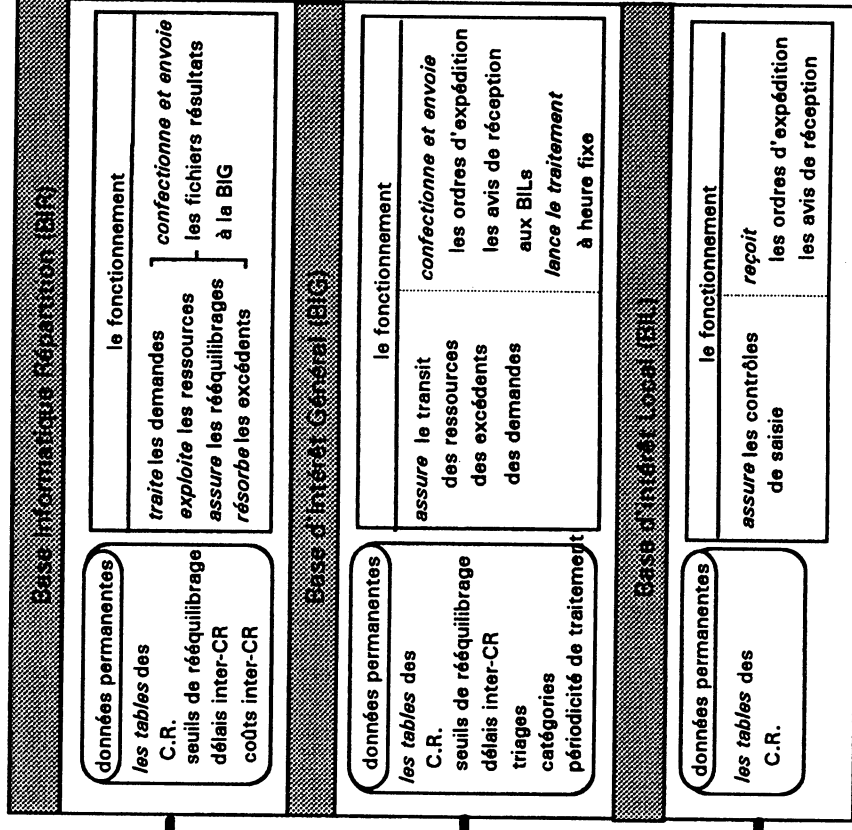
le répartiteur d'un C.R.

crée puis met à jour
les données permanentes de son C.R.
(seuils mini et maxi, ...)
seisit
ses demandes, ses ressources
exploite
les ordres d'expédition
les avis de réception

Les outils de saisie/calcul



Les ordinateurs de traitement



I.4 - Données et règles du problème

I.4.1 - Les données

Les demandes et les ressources

Prévisions :

Le fait que la plupart des clients commandent du matériel vide pour le lendemain, génère de gros problèmes de gestion. C'est pourquoi la SNCF se propose d'inciter les clients à exprimer leurs demandes à l'avance, ou de les prévoir partiellement, de manière à planifier au mieux la répartition de son parc de wagons.

Priorité :

Certains clients sont désignés prioritaires sous certaines conditions et pour une certaine période. Leurs demandes sont alors satisfaites avant les autres dans le cas où il n'y aurait pas suffisamment de ressources pour couvrir toutes les demandes.

Substitution :

Certains clients n'ont pas d'exigence particulière quant au type de matériel dont ils ont besoin. Le client a la possibilité de demander une catégorie de wagon en priorité tout en acceptant une ou plusieurs autres catégories de wagons compatibles avec le type de marchandise transportée. Grâce à cette formule le client a de plus grandes chances d'obtenir satisfaction. Nous estimons à environ 2% le nombre de demandes

avec substitutions. Les catégories de wagons sont regroupées et traitées par classe. Les substitutions n'existent qu'entre catégories d'une même classe.

Planification :

Pour planifier une répartition il faut connaître les ressources à court terme et déterminer des ressources prévisionnelles à partir de flux réels se dirigeant vers un point de déchargement en tenant compte des prévisions d'arrivée et des délais habituels de déchargement.

Réutilisation RIV :

Les wagons étrangers, présents sur le réseau, sont utilisés suivant la Règlementation Internationale (RIV).

Volume des demandes et des ressources :

Nous estimons que 6.000 demandes sont effectuées par semaine, pour toutes les catégories. Pour le traitement, nous regroupons les demandes des clients, affectés à un même CR, et nous regroupons les ressources disponibles, sur une zone d'action d'un CR, en tenant compte des délais d'acheminement.

Les éléments constitutifs d'une demande :

- catégorie;
- réseau propriétaire (fac);
- réseau destinataire (fac);
- période de la demande;
- point de répartition destination;
- nombre de wagons par type de demande.

remarque : les points suivis de la mention "fac" sont facultatifs.

Les types de demandes :

- 1 - prioritaires sans substitution, entre J et J+5;
- 2 - prioritaires avec substitution, entre J et J+5;
- 3 - fermes sans substitution, entre J et J+5;
- 4 - fermes avec substitution, entre J et J+5;
- 5 - prévisionnelles sans substitution, entre J+3 et J+5;
- 6 - prévisionnelles avec substitution, entre J+3 et J+5.

Les éléments constitutifs d'une ressource :

- catégorie;
- réseau propriétaire(fac);
- Ti (période de la ressource);
- point de répartition origine;
- sous-ensemble (gare/lieu) origine;
- nombre de wagons par type(s) de ressource.

Les types de ressources :

- . Effectives : wagons vides sains disponibles en J;
- 3 - R1 : ressources en dessous du seuil minimum;
- 2 - R2 : ressources entre le seuil minimum et le seuil maximum;
- 1 - R3 : ressources au dessus du seuil maximum;
- . Potentielles: wagons prévus disponibles entre J+1 et J+3;
- 6 - R1 : ressources en dessous du seuil minimum;
- 5 - R2 : ressources entre le seuil minimum et le seuil maximum;
- 4 - R3 : ressources au dessus du seuil maximum.

Les seuils de rééquilibrage

Actuellement les clients expriment la plupart de leurs demandes pour le lendemain. Or, pour pouvoir exploiter au mieux un système de gestion basé sur un horizon d'une semaine, il faut inciter les clients à exprimer leurs demandes plus longtemps à l'avance. Par défaut, il faut définir, pour chaque catégorie, un seuil minimum et un seuil maximum correspondant à la fourchette habituelle des demandes journalières pour chaque Centre de Répartition. Le traitement du rééquilibrage consiste alors à déplacer les ressources restantes du jour J, après satisfaction des demandes connues sur l'horizon, vers les gares dont les seuils minima et maxima ne sont pas couverts. Après l'étude du rééquilibrage, s'il reste en J des ressources inutilisées, il faut répartir judicieusement ces excédents de wagons vers d'autres CR ayant une capacité de garage suffisante.

Les types de seuil :

- 1 - Seuils Minima prioritaires;
- 2 - Seuils Minima non prioritaires;

3 - Seuils Maxima.

Les délais et les coûts d'acheminement

Nous utilisons des coûts et des délais d'acheminement théoriques, pour le calcul de la solution. Le calcul des délais d'acheminement est une donnée très importante, car c'est le seul élément qui nous permette de savoir si un acheminement est réalisable entre deux points du réseau. Pour le traitement répartition, une table est constituée à partir d'un million d'ATHEO. Une approche intuitive du problème aurait été d'inclure la recherche du meilleur acheminement dans le calcul de la solution. C'est à dire de faire l'équivalent d'un ATHEO pour chaque affectation d'une ressource à une demande. Seulement comme nous l'avons vu précédemment, le calcul d'un ATHEO est complexe du fait des nombreuses contraintes liées à la gestion des wagons de marchandise. Nous avons envisagé une étude simplifiée propre à l'acheminement des wagons vides. Mais nous nous sommes heurtés rapidement à des problèmes techniques, nous montrant qu'il n'était pas encore possible actuellement de faire abstraction du Plan de Transport des wagons vides et chargés.

Remarques :

- . Ces valeurs sont indépendantes du temps de chargement et de déchargement, ainsi que de la catégorie traitée.
- . L'unité de temps d'un délai d'acheminement est la même que celle des demandes et des ressources.
- . Nous supposons que le coût est linéaire par rapport au nombre de wagons acheminés et qu'il est non négatif.

Les éléments constitutifs d'un délai d'acheminement :

- point de répartition origine;
- date de disponibilité du wagon;
- point de répartition destination;
- date théorique d'arrivée à destination;
- niveau d'acheminement (1 ou 2).

Les éléments constitutifs d'un coût d'acheminement :

- point de répartition origine;
- point de répartition destination;

- niveau d'acheminement (1 ou 2).

I.4.2 - Les règles de gestion

Parmi toutes les solutions réalisables, respectant les règles de gestion suivantes, on cherche celle de moindre coût.

Les demandes et les ressources

Les demandes doivent être satisfaites dans l'ordre chronologique des périodes de l'horizon de travail. Si le volume des ressources est supérieur au volume des demandes, alors le traitement ne tient pas compte du types des demandes.

Par définition, *les ressources* sont utilisées par niveau de priorité (cf. § précédent).

Certaines catégories sont désignées par le Central Répartition, comme étant *équivalentes*. Pour cela, il désigne un ordre de traitement des catégories de wagons et une demande non satisfaite est retraitée avec les demandes de la catégorie équivalente.

Certaines catégories de référence sont réutilisables suivant les normes *RIV*. Dans ce cas un *réseau étranger* est mentionné. Deux cas se présentent, en cas de pénurie de wagons :

- les wagons du parc SNCF sont utilisés pour satisfaire des demandes à destination du réseau mentionné;
- les wagons du réseau mentionné sont utilisés pour satisfaire des demandes du réseau SNCF.

Les seuils de rééquilibrage

Nous rééquilibrons les ressources qui n'ont pas été utilisées pour satisfaire les demandes connues sur l'horizon de répartition. Ces ressources sont du type R2 ou R3. Nous comblons, dans la mesure du possible, les seuils en suivant l'ordre de priorité défini précédemment. Nous ne tenons pas compte des délais, mais des coûts d'acheminement.

Les délais et les coûts d'acheminement

Une ressource est affectée à une demande si la ressource peut être acheminée vers la demande dans les délais (en niveau 2 ou en niveau 1). Cette affectation est minimale en terme de coût d'acheminement.

Le *délai d'attente à destination* consiste à ne pas autoriser l'arrivée du wagon dans une gare plus de x heures avant la livraison de celui-ci au client, x étant paramétrable par l'utilisateur.

I.4.3 - Respect des règles par catégorie

Le traitement répartition se déroule en trois temps, d'après les différents types de catégories que nous avons définis :

- T1 est le traitement des demandes pour les catégories fines et réseaux.
- T2 est le traitement des demandes pour les catégories de références avec réutilisation des wagons étrangers selon les normes RIV.
- T3 est le traitement du rééquilibrage et des excédents de wagons.

Les règles de gestion précédentes sont alors utilisées dans les conditions suivantes:

	T1	T2	T3
gestion des ressources	*	*	*
affectation d'une ressource à une demande	*	*	
concordance entre la catégorie demandée et proposée	*	*	*
délai d'attente à destination		*	
catégories équivalentes	*	*	
gestion des RIV		*	
gestion du rééquilibrage et des excédents			*

Approche globale

Chapitre II

II.1 - Le problème de transport

Des limites de l'organisation actuelle, il se dessine une nouvelle organisation. Elle repose sur une structure hiérarchisée à deux niveaux et sur un horizon glissant d'une semaine.

L'objet de ce chapitre est de présenter le modèle-répartition, opérationnel à la SNCF depuis le mois de Juin 1992.

Le calcul de la répartition est, à la base, un problème de transport déjà très étudié en Recherche Opérationnelle. Il existe de nombreux algorithmes qui donnent la solution optimale aux problèmes théoriques, mais il est souvent difficile d'y insérer l'ensemble des contraintes issues des problèmes industriels. L'objet de cette étude est de mettre au point un modèle efficace regroupant les diverses contraintes de la Répartition.

II.1.1 - Historique du problème

Un système de transport porte en général sur un parc de véhicules transportant des biens ou des personnes sur un réseau. Un véhicule étant défini comme un moyen de transport, tel que le camion, l'avion, le bateau, le wagon, le conteneur, etc..

De nombreuses recherches ont été faites sur les problèmes de distribution de véhicules vides. Historiquement, l'allocation de matériels vides par rail a retenu davantage l'attention que les problèmes correspondant à d'autres modes de transport (par route, par bateau, par avion, etc.). Plus récemment, des problèmes de transport par camion et conteneur ont été étudiés. Ces modèles consistent à rechercher un déplacement optimal ou une mesure optimale d'un parc de véhicules et génèrent un bénéfice relatif à une réduction du coût d'acheminement (le kilométrage par exemple).

La plupart des systèmes de transport ont deux caractéristiques principales : ils sont *dynamiques*, car les demandes du système changent avec le temps, et *non prévisibles*, car les demandes ne sont pas connues à l'avance avec certitude.

Il y a environ 40 ans [Fee,1957], l'industrie du chemin de fer a pris conscience de l'importance des mouvements à vide. Les premières modélisations de ce problème supposaient que les demandes et les mouvements à charge des véhicules étaient donnés, indépendants du temps et déterministes. Ces informations permettaient aussi de déterminer les ressources. Le système reposait sur un problème de transport, où les arcs, munis d'un coût, représentaient l'acheminement de wagons vides entre points terminaux. Durant les années 60 et 70, plusieurs modèles de ce type ont été mis en application sur différents réseaux de chemin de fer (cf. [Mis,72], [Her,73], [Bak,77]).

Par ailleurs, différentes recherches ont permis de modéliser les premiers problèmes dynamiques (cf. [Whi,69], [Whi,72], [Her,77]). Ces modèles ne traitaient que les mouvements à vide, en supposant connues les ressources et les demandes de wagons dans le temps.

Les années suivantes, le modèle de base s'est enrichi de quelques fonctionnalités sur la gestion des wagons (cf. [Dec,78], [Gli,85]). Par ailleurs, des formulations similaires ont été proposées pour d'autres modes de transport (*par exemple* [Ede,77]).

Herren en 1977 ([Her,77]) décrit un modèle de distribution pour les Chemins de Fer Suisse, consistant à maximiser la satisfaction de la demande tout en minimisant les coûts de manoeuvre et de transport. Le modèle Suisse considère un parc hétérogène de wagons avec des possibilités de substitution. Il représente la distribution des wagons vides, comme un problème de flot minimum dans un réseau espace-temps, où chaque noeud correspond à un type de wagon spécifique sur un train au départ. Les arcs représentent les opérations de triage, les substitutions de wagons et les mouvements des trains. Du fait de la petite dimension du réseau Suisse, le problème est résolu par un algorithme du type "out-of-kilter". Les résultats indiquent la possibilité de réduire le parc de wagons, tout en satisfaisant les demandes. Ils obtiennent une réduction du kilométrage et des coûts associés et répartissent 12000 wagons de marchandises vides, de 70 types différents, entre les 850 gares du réseau.

Cooper et Leblanc en 1977 ([Coo,77]) ont montré comment introduire des demandes stochastiques dans un problème de transport. Puis Jordan et Turnquist en 1983 ([Jor,83]) ont incorporé des éléments dynamiques et stochastiques, dans un problème de distribution de véhicules vides.

Plus récemment, les travaux de Powel, Sheffi et Thiriez en 1984 ([Pow,84]), sur l'allocation dynamique de véhicules dans l'industrie du transport routier, ont beaucoup contribué à l'évolution du modèle : ils ne supposaient plus que les mouvements à charge étaient donnés, mais les déterminaient avec les mouvements à vide en les intégrant comme des variables de décision dans le modèle. En 1987, Powell [Pow,1987], traite le problème d'allocation dynamique de véhicule, pour minimiser les profits totaux. Le problème est revu dans le contexte du transport par camion, avec une attention spéciale sur la répartition des camions et l'anticipation des demandes. Cinq méthodes différentes sont proposées pour résoudre ce problème. En 1988, Powell [Pow,1988] propose une étude sur le transport routier Nord-Américain. Le problème est de répartir des centaines de camions entre clients, chaque semaine. Un nouveau type de modèle est alors développé. Il combine des informations, connues en temps réel, pour affecter des conducteurs aux chargements. Il prévoit des demandes de chargement et l'activité des camions. Il intègre aussi la gestion des opérations de chargement, la tarification et le marketing.

Roy en 1985 ([Roy,1985]) décrit le modèle afin de simuler les flots de marchandises sur le réseau ferroviaire Brésilien. Elle traite le problème d'allocation de wagons vides à un niveau de planification stratégique. Elle étudie différentes méthodes de prévision

des matrices origine/destination de marchandises, notamment un modèle de croissance simple. Il s'agit en effet d'un aspect important du problème, puisque ces matrices constituent les données de base du modèle.

Bigras en 1985 ([Big,1985]) propose un modèle pour estimer les matrices Offres/Demandes entre 8 régions du Canada et pour 64 catégories de biens. Le problème consiste à trouver les flux inter-régionaux les plus proches des flux à priori obtenus à partir d'un modèle de régression sur les flux observés et sur diverses variables socio-économiques. Le modèle est performant pour la prévision, mais présente certaines limites, notamment dans le choix modal et l'affectation.

Picard en 1987 ([Pic,1987]) développe un modèle de simulation des flux de marchandises inter-régionaux, en fonction du système de transport et d'un scénario donné. Dans sa version définitive, le modèle doit comporter une composante de choix modal, ainsi qu'une affectation des flux sur le réseau de transport Canadien.

Crainic, Dejax et Gendreau ([Cra,1989]) examinent la problématique et la planification des opérations de transport terrestre de conteneurs. Ils étudient, par exemple, la localisation de dépôts, la prévision de flux de conteneurs vides, les modes de transport des conteneurs, l'estimation des coûts et demandes. Ils présentent une approche de modélisation comprenant un modèle stratégique/tactique de localisation/distribution et un modèle stochastique et dynamique de répartition et d'acheminement (cf aussi [Crai,1989]).

Sanso & Co. ([San,1990]) mettent au point un important logiciel du nom de GENCOL. L'approche utilisée est la génération de colonnes et la classe des problèmes traités se formule comme un problème de recouvrement et de partitionnement : par exemple, l'ordonnancement des conducteurs de bus, le ramassage scolaire et le transport des handicapés.

Schlaepfer ([Sch,1991]) met au point un modèle pour répartir les wagons vides sur le réseau ferré allemand. L'algorithme utilisé est du type "out of kilter". Le réseau Allemand comprend 350 centres, 7 groupes de wagons, des demandes et ressources par catégorie et par groupe avec 6 niveaux de priorités. Les trains entiers sont traités manuellement, ainsi que quelques catégories particulières. Les critères principaux sont la ponctualité, l'optimisation des coûts, la satisfaction des demandes, la gestion des mouvements de wagons vides et le rééquilibrage prévisionnel des wagons disponibles vers certains points de demande.

Enfin, les derniers travaux ont porté sur la mesure optimale d'un parc de véhicules. Citons par exemple Beaujon et Turnquist [Bea,91], qui ont développé un modèle pour lequel les demandes fluctuent dans le temps (changements dynamiques et stochastiques) et les temps de trajets sont incertains.

II.1.2 - Classification de différentes méthodes

Nous définissons le noyau du modèle, comme *un problème de Transbordement Dynamique avec contraintes de capacité*, pour la répartition de wagons vides d'une certaine catégorie, sur un horizon de planification d'une semaine.

Nous verrons ensuite, comment ce noyau s'intègre au modèle-répartition, que nous définissons comme *un problème d'allocation dynamique et stochastique sur un horizon glissant*.

Définitions

Le *problème de transbordement* avec capacité est le plus général des problèmes de flots simples sur un réseau. Il regroupe les *problèmes de transport* (routiers, ferroviaires, aériens, etc.) avec et sans capacité, les *problèmes d'affectation*, les problèmes de *flots maxima* et les problèmes de *Plus Courts Chemins (PCC)*.

Ces modèles sont utilisés pour beaucoup d'applications. Un des meilleurs exemples pour illustrer la notion de flots sur un graphe est celui du courant électrique continu parcourant un réseau de dipôles.

Soit un réseau orienté $G = [N, E]$ déterminé par la donnée d'un ensemble N , dont les n éléments sont des sommets, et d'un ensemble E , dont les m éléments sont des arcs ou des couples ordonnés de sommets.

Un flot dans G est un vecteur à m composantes $x = (x_1, x_2, \dots, x_m)^T$ dans R^m tel que, en tout sommet i dans N de G , la première loi de Kirchhoff soit vérifiée, c'est à dire :

$$\sum_{\{j/(i,j) \in E\}} x_{ij} = \sum_{\{j/(j,i) \in E\}} x_{ji} \quad \forall i \in N$$

La composante x_e , du vecteur x , est la quantité de flot parcourant l'arc $e=(i,j)$ ou $e=(j,i)$. Cette relation exprime le fait que la somme des flux entrant sur un sommet est égale à la somme des flux sortants.

A chaque sommet i dans N , on associe un montant a_i :

- si $a_i > 0$, alors le sommet est un point de ressources;
- si $a_i < 0$, alors le sommet est un point de demandes;
- si $a_i = 0$, alors le sommet est un point intermédiaire sans ressources ni demandes.

A chaque arc (i,j) dans E , on associe un coût de transport c_{ij} par unité de flot et une capacité $k_{ij} > 0$.

Un *problème de transbordement* consiste à déterminer le flot maximum x sur le réseau, satisfaisant toutes les demandes.

Soit (P0) le problème de transbordement :

$$\left\{ \begin{array}{l} \text{Minimiser } z = \sum_{(i,j) \in E} c_{ij} \cdot x_{ij} \\ \text{s.c.} \\ \sum_{\{j/(ij) \in E\}} x_{ij} - \sum_{\{j/(ji) \in E\}} x_{ji} = a_i \quad \forall i \in N \quad (1) \\ 0 \leq x_{ij} \leq k_{ij} \quad \forall (i,j) \in E \quad (2) \end{array} \right.$$

L'équation (1) exprime la conservation du flot sur chaque noeud i de N .

L'équation (2) est une contrainte de capacité.

On suppose que le réseau donné est connexe et équilibré, c'est à dire :

$$\sum_{i=1}^n a_i = 0$$

Il est entendu qu'on autorise des arcs multiples $x_{ij}^1, x_{ij}^2, \dots$, mais pour simplifier les notations nous écrirons simplement x_{ij} .

Un *problème de Transport* est un problème de Transbordement où chaque sommet est soit une source, soit un puits. Dans ce cas, l'équation (1) devient :

$$\sum_{\{j/(ij) \in E\}} x_{ij} = a_i > 0 \quad (1')$$

$$\sum_{\{j/(ij) \in E\}} x_{ji} = a_i < 0 \quad (1'')$$

Un *problème d'affectation* est un problème de transport avec autant de sommets sources que de sommets puits. Les ressources et les demandes prennent respectivement les valeurs +1 et -1.

Un *problème de Transport complet* correspond à un graphe biparti des sommets sources et des sommets puits.

Un *problème de Flot Maximum* est un cas particulier, avec un seul puits et une seule source.

Classification de différentes méthodes

Par le passé un grand nombre de méthodes, pour résoudre les problèmes de Transport et de Transbordement, a été proposé. On trouve principalement une approche Primale du Simplexe [Dan,1951], une approche Duale du Simplexe [Bal,1962], une approche Primale-Duale du Simplexe, en particulier la méthode des arcs non conformes (en Anglais : out of kilter) [For,1957] et [Ful,1961], la méthode du cycle négatif [Ben,1973] et la méthode du Plus Court Chemin avec une approche Duale du problème de coût minimum avec bornes inférieures nulles [Bus,1961].

L'algorithme Dual de Busacker et Gowen [Bus,1961], appliqué au problème d'affectation, conduit à une méthode intéressante et efficace. Soit $n = 2.p$ le nombre de sommets du graphe. Les flots étant entiers à chaque itération, il y a au plus p itérations (p augmentations de flots). Chaque itération implique le calcul d'un ensemble de PCC d'origine s , les longueurs des arcs pouvant être négatives, on utilisera donc, par exemple, l'algorithme de Bellman et Dantzig ce qui nécessite $O(n^3)$ opérations. Globalement, l'algorithme est donc en $O(n^4)$. Cependant Edmonds et Karp [Edm,1972] ont montré qu'en adaptant convenablement l'algorithme Primal-Dual de la Programmation Linéaire, on pouvait se ramener, à chaque itération, à la recherche d'un ensemble de PCC fixés avec longueurs toutes négatives. L'algorithme est alors en

$O(n^3)$. Hopcroft et Karp [Hop,1973] ont encore amélioré la méthode et ont proposé un algorithme en $O(n^{2.5})$.

Le problème de transport Primal a été programmé dès 1952. Les expériences des années 1950, puis 1960 semblaient indiquer une supériorité des approches Primale-Duale [Dan,1963], [For1962], [Flo,1971]. Notons que l'algorithme Hongrois [Kuh,1955], pour le problème d'affectation, est un cas particulier de l'algorithme de Ford et Fulkerson [For,1962]. Un exemple typique est l'affectation de n tâches à n machines.

Un algorithme Primal de Klein [Kle,1967], pour le problème de Flot à coût Minimum avec bornes inférieures et supérieures de capacité, est apparu en 1967. Appliqué au problème de Transport, cet algorithme n'est autre que l'algorithme Primal de Dantzig [Dan,1963]. Sans être toujours le plus efficace, cet algorithme est un des plus connus.

Dans les années 1970, l'approche Primale s'est étendue au cas d'un graphe non complet [Glo,1974], au problème avec capacités [Lan,1974], puis au problème général de Transbordement [Glov,1974]. Le problème était alors de comparer l'algorithme des arcs non conformes à la méthode Primale.

Depuis il semble établi que l'approche Primale soit supérieure à l'approche Primale-Duale des arcs non conformes, grâce aux résultats de [Glov,1974] qui ont été vérifiés par ailleurs. L'approche Primale serait environ 30% plus rapide et nécessiterait moins d'espace mémoire que la méthode des arcs non conformes. Ces deux algorithmes permettent toutefois de résoudre des problèmes de très grande dimension (plusieurs milliers d'arcs) en des temps de calcul très réduits.

II.1.3 - Choix d'une méthode

Nous venons de justifier le choix d'une approche Primale pour le problème de Transbordement sur un réseau de Transport orienté.

En 1951, Dantzig [Dan,1951] spécialise une méthode Primale du Simplexe aux problèmes de Transport et de Transbordement, puis Charnes et Cooper [Cha,1961] définissent la méthode de "stepping stone". En 1977, des ouvrages décrivent une adaptation intéressante du Simplexe (cf. [Baz,1977]) ou traitent rigoureusement une

approche algébrique du problème (cf. [Bra,1977]). Enfin en 1980, Ahrens et Finke ([Ahr,1980]) donnent une très bonne implémentation (en Fortran) de la méthode de Bazarra. Ainsi, nos recherches nous ont naturellement porté vers ces travaux, que nous avons repris et adapté à notre problème.

Solution de base

L'algorithme Primal découle de la méthode du Simplexe avec variables bornées. On peut écrire l'équation (1), du problème (P0), sous forme matricielle $A.x = a$.

Une solution x est une *solution de base* si les colonnes de A (associées aux variables x_{ij} non nulles et non saturées) sont linéairement indépendantes. Les solutions de bases correspondent aux sous-matrices carrées régulières extraites de la matrice A . Chaque colonne de A est constituée d'un 1, d'un -1 et de 0.

Théorème :

Les solutions de base correspondent à un arbre issu d'un réseau connexe.

La démonstration (cf. [Ahr,1980]) découle des points suivants :

- (a) - les colonnes de A correspondant à un arbre sont linéairement indépendantes.
- (b) - le rang de la matrice A est plus petit que le nombre de sommets $|N|$.
- (c) - le graphe correspondant à une base de A ne contient pas de cycles.

Soit un arbre extrait d'un réseau connexe. Cet arbre a $|N|-1$ arcs par définition. D'après (a) et (b) on en déduit que c'est une solution en base. Chaque base a $|N|-1$ éléments. Son graphe a donc $|N|-1$ éléments et pas de cycle, c'est un arbre.

Critère d'optimalité

Le problème Dual (D0) de (P0) s'écrit de la manière suivante :

$$\left\{ \begin{array}{l} \text{Maximiser } h = \sum_{i \in N} a_i \cdot u_i - \sum_{(i,j) \in E} w_{ij} \cdot k_{ij} \\ \text{s.c.} \\ u_i - u_j \leq c_{ij} \quad \forall (ij) \in E \quad (3) \\ w_{ij} \geq 0 \quad \forall (ij) \in E \quad (4) \end{array} \right.$$

où les variables u_i ($i=1, \dots, |N|$) sont les variables duales des contraintes (1) du Primal et les variables w_{ij} ($\forall (i,j)$ dans E) sont les variables duales des contraintes (2) du Primal.

D'après le théorème des écarts complémentaires (cf. annexes), on sait que deux solutions (x,u) , respectivement primale et duale réalisables sont optimales si et seulement si

$$\forall (i,j) \in E \quad (c_{ij} - u_i + u_j) \cdot x_{ij} = 0 \quad (5)$$

$$\forall (i,j) \in E \quad (k_{ij} - x_{ij}) \cdot w_{ij} = 0 \quad (6)$$

et le coût réduit est égal à

$$\bar{c}_{ij} = c_{ij} - u_i + u_j \quad \forall (ij) \in E \quad (7)$$

Des relations (5), (6), (7), on en déduit les cas suivants :

cas 1 : $0 < x_{ij} < k_{ij} \quad \forall (i,j)$ dans E

$x_{ij} < k_{ij}$ et (6) impliquent $w_{ij} = 0$

$0 < x_{ij}$ et (5) et $w_{ij} = 0$ impliquent $c_{ij} = u_i - u_j$

cas 2 : $u_i - u_j < c_{ij}$ et $w_{ij} > 0$

$u_i - u_j < c_{ij}$ et $w_{ij} > 0$ et (5) impliquent $u_i - u_j - w_{ij} < c_{ij}$

$u_i - u_j - w_{ij} < c_{ij}$ et (5) impliquent $x_{ij} = 0$

cas 3 : $u_i - u_j > c_{ij}$

$u_i - u_j > c_{ij}$ et $w_{ij} > 0$ et (5) impliquent $w_{ij} > 0$

(6) et $w_{ij} > 0$ impliquent $x_{ij} = k_{ij}$

La condition d'optimalité s'exprime alors de la manière suivante : Une solution x , Primale réalisable, est une solution de base généralisée, s'il existe un arbre maximal T tel que :

$$(x_{ij} = 0, \bar{c}_{ij} \geq 0) \text{ ou } (x_{ij} = k_{ij}, \bar{c}_{ij} \leq 0) \quad \forall (i,j) \notin T \quad (8)$$

$$(0 < x_{ij} < k_{ij}, \bar{c}_{ij} = 0) \quad \forall (i,j) \in T \quad (9)$$

Changement de base

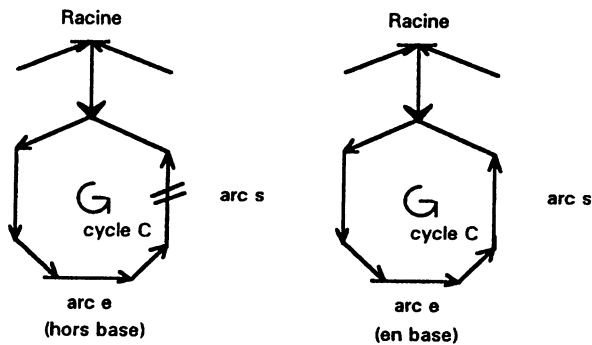
A chaque itération, on met en évidence une solution de base x réalisable. La relation (7) nous permet de déterminer de manière unique les valeurs des variables Duales associées à chaque sommet de l'arbre T . Si les conditions (8) et (9) sont vérifiées, alors la solution x est optimale, sinon on effectue le changement de base de la manière suivante :

l'arc $e = (l,k)$ non élément de T et candidat à entrer en base, est tel que :

$$x_{lk} = 0 \text{ et } u_l - u_k > c_{lk} \quad ((8) \text{ n'est pas vérifiée}) \quad (10)$$

$$x_{lk} = k_{lk} \text{ et } u_l - u_k < c_{lk} \quad ((9) \text{ n'est pas vérifiée}) \quad (11)$$

l'arc $s = (i,j)$ élément de T et candidat à sortir de base appartient au cycle C créé par l'ajout de l'arc $e = (k,l)$ à l'arbre T .



- L'arbre T avant et après le changement de base -

Deux cas se présentent :

Si (l,k) vérifie (10) alors on augmente le flot sur C d'une valeur Q , jusqu'à ce qu'un des x_{ij} , avec (i,j) dans C , atteigne une de ses bornes 0 ou k_{ij} . L'arc (i,j) est appelé *arc bloquant*. La nouvelle solution x' est la suivante :

$$x' = x_{ij} + Q \quad \forall (i,j) \text{ dans } C, \text{ tels que } (i,j) \text{ est un arc "avant".}$$

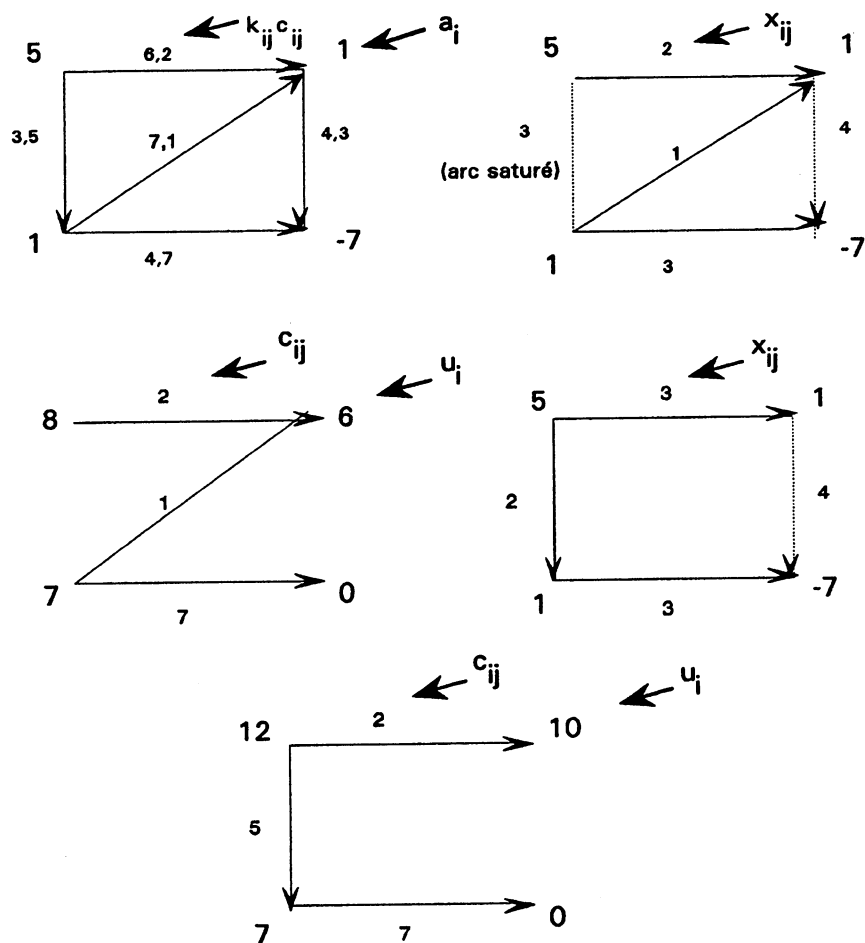
$$x' = x_{ij} - Q \quad \forall (i,j) \text{ dans } C, \text{ tels que } (i,j) \text{ est un arc "arrière".}$$

$$x' = x_{ij} \quad \forall (i,j) \text{ hors de } C.$$

Si (l,k) vérifie (11) alors on diminue le flot sur C d'une valeur Q , jusqu'à ce qu'un des x_{ij} , avec (i,j) dans C , atteigne une de ses bornes 0 ou k_{ij} . L'arc (i,j) est appelé *arc bloquant*. La nouvelle solution x' est la suivante :

$$\begin{aligned} x' &= x_{ij} - Q & \forall (i,j) \text{ dans } C \text{ tels que } (i,j) \text{ est un arc "avant".} \\ x' &= x_{ij} + Q & \forall (i,j) \text{ dans } C \text{ tels que } (i,j) \text{ est un arc "arrière".} \\ x' &= x_{ij} & \forall (i,j) \text{ hors de } C. \end{aligned}$$

Exemple



La solution optimale est trouvée, car les conditions 8 et 9 sont vérifiées.

Solution initiale

La première étape de l'algorithme Primal du Simplexe est de trouver une solution initiale de base qui soit réalisable. Trouver une bonne solution, c'est trouver un bon compromis entre la valeur de la solution initiale, le temps de calcul de celle-ci et le nombre de pivots (changements de base) nécessaire pour atteindre l'optimum.

Si le problème est non contraint en capacité, les arcs sélectionnés peuvent supporter un flot égal à la quantité de ressource ou de demande d'une des extrémités de l'arc. On sélectionne ainsi $|N|-1$ arcs :

La règle "*du coin Nord-ouest*" permet de trouver très rapidement une solution initiale sans se préoccuper du coût. Mais alors un grand nombre de pivotages est nécessaire pour atteindre l'optimum [Mod,1978].

La règle "*de la matrice minimum*" permet de trouver une "bonne" solution initiale. Mais cette étape est trop longue par rapport aux étapes suivantes de l'algorithme [Mod,1978].

Ces deux règles sont en fait des cas extrêmes. Srinivasen et Thompson [Sri,1973] proposent une alternative avec la règle "*de la colonne (ligne) minimum*". Ils obtiennent ainsi une bonne solution initiale par un calcul très rapide.

Si le problème est contraint en capacité, la recherche de la solution initiale est plus délicate :

Glover, Karney et Klingman [Glov,1974] ont appliqué la règle de la colonne minimum au problème de Transbordement. Seulement, les arcs manquants et les capacités n'assuraient plus l'existence d'une solution réalisable. Ils ont alors ajouté un sommet fictif et relié ce sommet à chaque point du graphe. Une heuristique leur permettait ensuite d'accommoder ce nouveau graphe.

Bradley, Brawn et Graves [Bra,1977] ont montré qu'il était plus simple et aussi plus efficace en général de construire une solution initiale totalement artificielle. Le principe est d'ajouter au graphe de départ un sommet fictif (la racine de l'arbre initial) et de relier ce sommet à chaque noeud. Ces arcs ont un coût et une capacité très élevés, pour être désactivés peu à peu au cours des itérations suivantes et ne plus apparaître à l'optimum.

Structure de données

Nous avons choisi cet algorithme pour son efficacité, la structure des données et le codage machine mis en oeuvre (cf. [Ahr,1980]). L'efficacité de la structure se situe au niveau de la recherche des données et de la mise à jour des solutions :

L'arbre T est défini par sa racine et pour chaque sommet i de T on connaît son prédécesseur (le noeud suivant i sur le chemin unique menant à la racine). L'orientation de l'arc entre i et son prédécesseur est conservée dans une liste et le flot courant sur cet arc est conservé dans une autre liste.

L'arc candidat à entrer en base crée un cycle sur l'arbre et la profondeur des sommets du cycle (longueur du chemin unique menant à la racine) permet de déterminer l'arc candidat à sortir de base. La mise à jour de la solution à l'itération courante peut être fastidieuse avec la recherche des différents successeurs (le noeud avant i sur le chemin unique menant à la racine). Knuth [Knu,1968] identifie le premier fils de i par un pointeur et le prochain frère de i par un autre pointeur, puis il définit récursivement les sommets par la *règle du successeur suivant le préordre*, en commençant par la racine de l'arbre.

D'autres algorithmes utilisent le nombre de successeurs ou la distance dans la relation de préordre. [Ali,1978] et [Bra,1977] font une étude comparative des différentes méthodes et désignent la *notion de profondeur* comme la plus rapide des différentes alternatives.

Dégénérescence de l'algorithme de transbordement

Un problème est dit dégénéré lorsque plusieurs variables en *base* prennent la valeur zéro en même temps. Une nouvelle solution de base, avec moins de $(m+n-1)$ valeurs strictement positives, est générée. On gère ce problème en choisissant la variable sortant de la base arbitrairement parmi les variables de base atteignant zéro en même temps et on traite les autres variables à zéro comme des variables de base. Ce point est abordé, en détail, dans la partie III de ce rapport.

II.2 - Planification et prévisions

II.2.1 - Gestion à court terme

Afin de mieux cerner la nature des décisions de gestion dans un système de transport, nous les regroupons en 3 niveaux [Ant,1965]. On distingue :

Les décisions stratégiques relatives à la politique générale de l'entreprise, qui portent en général sur un horizon de 2 à 5 ans. Ce niveau concerne le long terme, les investissements à engager et, à un niveau très général, l'équilibre coût/qualité à atteindre dans le transport. Citons, par exemple, l'amélioration du réseau, l'emplacement des gares de triage et la planification à très long terme des services.

Les décisions tactiques, qui assurent la gestion du travail et des moyens de production sur un horizon de 6 à 18 mois environ dans le cadre des décisions stratégiques. Le carnet de commande connu et la capacité des moyens de production, ayant été fixés par le niveau supérieur, il s'agit en général de définir un plan de production et un plan d'utilisation des moyens disponibles [Mer,1987]. Ce niveau concerne le moyen terme ainsi que l'allocation rationnelle et efficace des ressources. Les données doivent avoir un certain degré d'agrégation, les décisions prises n'étant sensibles qu'à des modifications importantes des paramètres du problème. Les décisions à prendre ne tiennent pas compte des variations journalières et ont comme objectif l'amélioration des performances de tout le réseau. Par exemple, on recherche

le Plan de Transport le plus économique et les services horaires les mieux adaptés au trafic attendu. C'est un des problèmes que doit résoudre une compagnie de chemin de fer pour offrir un service rentable, économique et de qualité.

Les décisions opérationnelles, qui régissent en détail la fabrication proprement dite pour les jours ou les semaines à venir. La répartition, des wagons vides et des locomotives classées au niveau opérationnel, font aussi l'objet de décisions au niveau tactique ou même stratégique. Les différences entre les niveaux sont cependant évidentes : chaque niveau prend ses grandes lignes de développement au niveau supérieur et se base sur les résultats du niveau inférieur.

Ces 3 catégories de décisions diffèrent entre autres choses par l'horizon qui les caractérise et par le niveau de détail des informations requises pour la prise de décisions [Gia,1981] à chaque niveau de la structure.

Dans la suite de ce travail, nous nous intéressons aux décisions opérationnelles. Notre objectif est d'assurer la gestion à court terme des wagons de marchandises vides, de réduire le niveau des stocks, d'améliorer le service rendu à la clientèle et de faire face à différents types d'aléas (variations de l'activité commerciale, etc.).

Complexité et dynamique

Le nombre et la variété des éléments nécessaires au traitement de la répartition sont à l'origine de la complexité de celui-ci. De plus, comme tout système dynamique, le système ferroviaire évolue constamment au cours du temps et doit faire face aux perturbations. De ce fait, un certain nombre de données, comme les demandes et les ressources, évoluent avec le temps. La complexité du système ferroviaire et son caractère dynamique compliquent grandement la tâche du répartiteur.

Nécessité de planifier et prévoir

Compte tenu de l'inertie du système, toute décision de répartition doit être anticipée pour pouvoir être mise en oeuvre. Une programmation prévisionnelle des décisions est donc absolument nécessaire, pour effectuer un arbitrage temporel entre les différents

éléments à prendre en compte. L'activité qui consiste à établir, sur la base de données disponibles, une programmation des décisions est appelée *planification*. Elle est essentielle pour les demandes à court terme et permet de mettre en oeuvre une gestion coordonnée, dans le temps, des acheminements et des wagons.

Or la planification de la répartition à court terme nécessite la connaissance des données relatives au système sur les quelques jours ou quelques semaines à venir et un certain nombre de ces données ne sont pas disponibles à l'avance. Leur évaluation relève de la *prévision*. La fonction planification doit aussi tenir compte de l'incertitude, sur certaines données, due aux erreurs de prévision. Pour le problème de la Répartition, une connaissance de la nature et de l'importance des services passés permet d'estimer la demande à moyen terme, par type de catégorie, par client et par période. Ces prévisions sont l'élément moteur du système, car elles permettent de rééquilibrer automatiquement les wagons sur le réseau. La connaissance des experts, capables de prévoir partiellement la demande sur un horizon d'une semaine, est utilisée, pour la prévision ou pour l'élaboration d'un modèle prévisionnel.

II.2.2 - Planification sur un horizon glissant

La planification de la répartition est l'activité qui consiste à analyser l'ensemble des informations disponibles en vue de définir un plan de répartition.

Le plan de répartition est une suite de décisions prévisionnelles conduisant à un fonctionnement satisfaisant du système de production sur un intervalle de temps appelé horizon de planification.

L'horizon de planification représente l'intervalle de temps sur lequel est établi le plan de répartition.

Cet horizon est généralement discrétisé en *périodes*, représentant l'unité de temps élémentaire. Les données (demandes, ressources,...) et les décisions font alors référence à cette unité de temps.

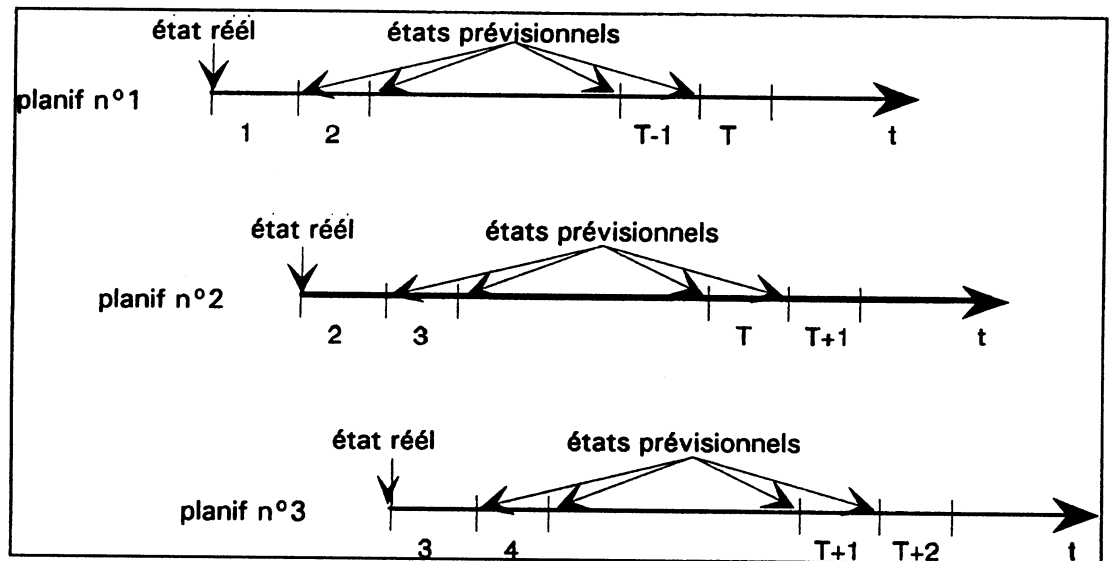
Horizon infini

En général, les problèmes d'optimisation sur un horizon infini sont des problèmes difficiles. Par contre, il est souvent possible de trouver une décision initiale qui soit optimale.

Le principe de la méthode est de discrétiser le temps en périodes élémentaires et de découpler le problème infini en un problème de T périodes et un problème infini débutant à la période T . En théorie, l'horizon de T périodes peut être étudié sur un ordinateur donné (donc T n'est "pas trop long"), et permet d'obtenir des décisions initiales optimales (donc T n'est "pas trop court"). En pratique, il n'est pas toujours possible de définir un tel horizon de planification, il est alors fixé de manière arbitraire en fonction du contexte: au niveau de décision opérationnel, il varie d'une semaine à un mois.

Horizon glissant

Les propositions ci-dessus ne sont malheureusement pas vérifiées pour le problème dynamique et stochastique de la Répartition. En effet, nous ne connaissons pas avec certitude toutes les données des T premières périodes ($T > 1$). Dans ce cas, pour assurer une plus grande fiabilité des données, nous proposons une approche du problème avec un horizon glissant de T périodes. Les problèmes liés à la détermination d'un horizon glissant ont été étudiés par [Las,1986] et [Ben,1983]. Par définition, l'horizon de planification délimite le seuil au-delà duquel les informations non prises en compte dans la planification n'ont pas d'influence. En pratique, nous définissons l'horizon de planification par rapport à la durée maximale d'un acheminement : en général une semaine.



- planification et prévision -

L'intérêt de cette procédure porte sur l'évolution de la planification. En effet, sur chaque horizon, on prend en compte les nouvelles données apparues depuis le début de la période précédente et les anciennes données réactualisées. Dans un contexte dynamique, on voit bien la nécessité des réactualisations périodiques sur le court terme, à l'exception des décisions relatives à la première période de l'horizon qui sont réellement prises en compte. En pratique, il est donc souhaitable d'attendre le dernier moment pour donner un ordre de répartition en gare, pour remettre en cause la satisfaction d'une demande le plus souvent possible.

Remarques (cf. [Gri,1977], [Gri,1983], [Bea,1984]) :

- *Les décisions initiales sont elles optimales ?*

D'une manière générale, le niveau d'incertitude augmente dans le futur, c'est pourquoi il ne faut pas que l'horizon de planification soit trop grand. Nous pouvons dire que les données sur les T premières périodes sont connues avec une certaine marge d'erreur relativement faible. Par conséquent, nous ne remettons pas en cause les décisions optimales de la première période. En pratique, cela revient à dire que l'unité de temps est faible (la journée ou la 1/2 journée), ce qui présente l'avantage de travailler avec des délais d'acheminement plus précis.

- *Quel horizon de planification ?*

En pratique, tout porte à croire que l'horizon de planification est d'une semaine. En effet, le degré d'incertitude au-delà de cet horizon n'est plus significatif et le délai maximal d'acheminement est d'une semaine. Cet horizon est ni trop long, ni trop court et il respecte en ce sens, les contraintes évoquées précédemment.

II.2.3 - Pr vision des besoins et des ressources

L'objet de ce paragraphe est d' tudier un mod le portant sur la pr vision des besoins et des disponibles,   partir d'un historique de donn es.

Pr vision des besoins

L'estimation de la demande dans un environnement concurrentiel et industriel est tr s difficile, voir impossible. On peut, en g n ral, d finir une tendance des mouvements saisonniers sur l'ann e, la semaine, la journ e ou l'heure. Seulement l' tude des pr visions est incompl te si on ne tient pas compte des mouvements irr guli rs ou al atoires, comme les intemp ries ou les gr ves.

Avant chaque traitement r partition, les besoins pr visionnels de l'horizon sont calcul s de mani re   tenir compte des informations les plus r centes. Des variations saisonni res n'ont pu  tre mises en  vidence, du fait notamment de l' volution de la client le (remplacement des march s importants par de petits march s tr s al atoires). La m thode choisie repose alors sur les donn es des 7 semaines pr c dentes, afin d' tre plus sensible aux variations, tout en disposant d'un historique suffisant.

Au d but d'un traitement, on effectue la r partition d'une cat gorie donn e, sur les 7 jours   venir. De cet horizon, on exclu les samedis et les dimanches. Par contre, les jours f ri s, qui ne sont ni samedi ni dimanche, sont conserv s afin de garder sur l'horizon les jours de la semaine allant du lundi au vendredi.

Les besoins d'une gare sont les wagons qu'elle est suppos e consommer   une p riode donn e. La pr vision des besoins a deux sources d'information : le traitement r partition et les donn es ci-dessous.

- le nombre WC de wagons charg s, par gare et par p riode, (ce nombre comprend les wagons exp di s   l' tranger).
- le nombre NF de wagons non fournis au client, par p riode.

L'analyse d'une *série temporelle* consiste à faire une description (mathématique en général) des mouvements qui la composent. Une série temporelle est basée sur une série d'observations effectuées sur des périodes ou des instants donnés et ordonnées selon leurs dates d'observations. Prévoir à partir de ces observations, c'est utiliser les informations du passé, pour fournir la valeur la plus probable d'une série temporelle sur une ou plusieurs périodes à venir.

La classification des *techniques de prévision* est, comme toute classification, en grande partie arbitraire, et celle que l'on propose ci-dessous n'échappe pas à la règle (cf. [Gia, 1981]).

Modèles explicatifs :

- Equation unique (régression linéaire).
- Equations simultanées (modèles économiques).

Modèles autoprojectifs :

- Historiques récents (filtres linéaires) :
 - Moyennes mobiles.
 - Lissage exponentiel.
 - Filtre différence.
- Historiques complets :
 - Moindres carrés.
 - Box et Jenkins, et prolongements.
 - Analyse spatiale.

Les modèles explicatifs se fondent en partie sur des valeurs prises par des variables autres que celles que l'on cherche à projeter. Ces modèles ne sont envisageables que pour la prévision à court terme portant sur des séries très agrégées.

Les modèles autoprojectifs considèrent que le futur se déduit tout naturellement du passé. C'est dans ce cas précis que nous nous situons. La participation retenue pour les modèles autoprojectifs s'explique par des considérations de coût : les techniques, opérant sur un historique récent et restreint, sont peu onéreuses mais peut être moins fiables. Pour notre étude nous avons choisi de conserver un historique récent, ce qui se traduit par un coût de stockage des données et de traitement de moindre importance.

Le *filtre linéaire* est le mode de transformation d'une suite chronologique $x(t)$ sur une autre suite $y(t)$ à partir d'une combinaison linéaire des termes de la chronique initiale :

$$y(t) = \sum_{i=1}^s a_i \cdot x(t - r + i)$$

où

a_i est un coefficient de pondération tel que $\sum_i a_i = 1$,

r est le décalage temporel et

s est le nombre de termes de la chronique initiale.

Les *moyennes mobiles* sont les plus anciennes des méthodes de filtrage et sont plus adaptées à l'estimation de composantes tendancielles, qu'à la prévision. Les objectifs poursuivis avec des techniques de *lissage exponentiel* sont, dans la pratique, exclusivement ceux de la prévision. Les filtres du lissage exponentiel ont des coefficients à croissance géométrique et portent implicitement sur l'intégrité de l'historique possédé. Les objectifs du *lissage exponentiel double* sont les prévisions par une évolution tendancielle linéaire et une absence de saisonnalité (ayant pu être éliminée par un filtre).

L'application du lissage exponentiel simple à une chronique linéaire $x(t)=at+b$, non perturbée par une constante aléatoire, génère une nouvelle chronique $y(t)$ décalée par rapport à $x(t)$.

Pour une chronique à $(s+1)$ termes, repérée par des indices variants de 0 à s , $(x(0), x(1), \dots, x(s))$, le filtre linéaire exponentiel définit la valeur $y(s)$ suivante :

$$y(s) = k \sum_{i=0}^{s-1} (1-k)^i x(s-i) + (1-k)^s \cdot x(0)$$

ou encore,

$$y(s) = k \cdot x(s) + (1-k) \cdot y(s-1)$$

où k est le coefficient de lissage de la série à lisser.

La somme des coefficients de pondération est égale à 1. En effet :

$$\sum_{i=0}^{s-1} (1-k)^i = \frac{1-(1-k)^s}{1-(1-k)} \Rightarrow k \sum_{i=0}^{s-1} (1-k)^i + (1-k)^s = 1$$

Le poids d'une information dans le calcul de $y(s)$ est d'autant plus faible que l'information est ancienne. C'est du reste l'un des buts recherchés par cette technique que de privilégier les informations les plus récentes au détriment des informations les plus anciennes, alors que la technique de la moyenne mobile centrée accorde le même poids à toutes les informations. Le poids du passé est d'autant plus fort que le coefficient k est faible.

Le lissage exponentiel double diffère par le fait que l'écart entre $x(t)$ et $y(t)$ n'est pas stable : il croît progressivement pour tendre assez rapidement (plus ou moins selon le coefficient de lissage k) vers une valeur limite qui est égale à :

$$a.(1-k)/k$$

Soit $z(t)$ un lissage exponentiel de $y(t)$. Cette double opération de lissage est à l'origine de l'appellation lissage exponentiel double :

$$z(t) = k.y(t) + (1-k).z(t-1)$$

L'écart entre $y(t)$ et $z(t)$ est le même que celui qui existe entre $x(t)$ et $y(t)$:

$$y(t) - z(t) = x(t) - y(t) = a.(1-k)/k$$

d'où on tire une estimation, à la date t , de la pente a :

$$\hat{a} = (y(t) - z(t)) k/(1-k)$$

Pour retrouver la valeur théorique de $x(t)$, il suffit d'ajouter cet écart à $y(t)$:

$$\hat{x}(t) = y(t) + (y(t) - z(t)) = 2.y(t) - z(t)$$

Pour trouver la valeur théorique de $x(t+r)$, il suffit d'ajouter $\hat{a}$ à cette valeur :

$$\hat{x}(t+r) = \hat{x}(t) + r.\hat{a} = y(t) + (y(t) - z(t))(1 + t.\frac{k}{1-k})$$

Au début d'un traitement, on désire effectuer la répartition d'une catégorie donnée sur l'horizon H . La première opération est d'effectuer la prévision des besoins sur cet horizon, en fonction d'un historique des données sur les 7 horizons précédents.

Algorithme :

Initialisation :

$$k = 0,1$$

Observations $x(s)$:

$$x(s) = \sum_{t \in H(s)} WC(t) + \max_{t \in H(s)} \{ NF(t) \} \quad s=1, \dots, 7$$

Lissage exponentiel double :

$$y(s) = (1-k) \cdot y(s-1) + k \cdot x(s) \quad s=1, \dots, 7$$

$$z(s) = (1-k) \cdot z(s-1) + k \cdot y(s) \quad s=1, \dots, 7$$

Estimations de $x(s)$:

$$\hat{x}(s) = y(s) + (y(s) - z(s)) = 2 \cdot y(s) - z(s)$$

Estimations faites en s de $x(s+1)$:

$$\hat{x}(s+1) = \hat{x}(s) + \hat{a} = y(s) + (y(s) - z(s)) \left(1 + \frac{k}{1-k}\right)$$

Remarques :

- Nous n'effectuons pas la somme des non fournis, car ils peuvent être comptabilisés plusieurs périodes de suite.
- Les prévisions périodiques sont calculées à partir des besoins globaux de l'horizon.
- On a constaté que certains CR chargent davantage à certaines périodes de la semaine. Aussi, pour chacune des périodes de l'horizon, on calcule sur les sept horizons précédents, une proportion moyenne. Cette proportion est appliquée à la valeur des prévisions périodiques. Si l'horizon comporte un jour férié, la valeur des besoins est annulée pour ce jour, puis reportée sur les périodes précédentes et/ou suivantes.
- Afin de ne pas sous-estimer les besoins, on ajoute à chaque prévision périodique, une valeur constante égale à 20 % de la valeur moyenne du chargement périodique d'un centre sur les 7 derniers horizons.

Calcul des disponibles

Les wagons disponibles, pour une gare et une période donnée, sont des wagons présents vides, c'est à dire pouvant être chargés (consommés sur place) ou taxés (expédiés vides vers d'autres gares). Un wagon non consommé ou taxé pour une période, reste disponible pour la période suivante. Le calcul des disponibles a deux sources d'information : le traitement répartition et les données ci-dessous.

- le nombre de wagons présents vides, par gare;
- le nombre de wagons attendus vides, par gare;
- le nombre de wagons en garage;
- le nombre de wagons sortis de l'atelier, par période.

Lors du premier traitement Répartition, la valeur des ressources effectives est initialisée à la somme des présents vides, des attendus vides et des présents en garage. Cette valeur est répercutée sur les T périodes de l'horizon. Les ressources potentielles sont calculées à partir des arrivées de wagons taxés par le répartiteur les jours précédents.

Aux traitements suivants, on regroupe les ressources restantes de l'horizon précédent et les nouvelles ressources apparues depuis le dernier recensement.

Soit $H = [1, \dots, t, \dots, T]$ l'horizon courant de répartition. Nous supposons connus les départs de wagons taxés, ou consommés sur place, ainsi que les échanges avec l'étranger. Une taxe correspond à une diminution des ressources effectives, pour le CR expéditeur à la période $t-1$, et à une augmentation des ressources effectives, pour le CR destinataire à la période de réception prévue. Les départs pour l'étranger correspondent à une diminution des ressources effectives, à la période $t-1$, et les retours de l'étranger à une augmentation des ressources effectives à la période de réception prévue. On ajoute (resp. retranche), aux ressources effectives de la période t les les sorties (resp. entrées) d'atelier.

Du fait de la rapidité des acheminements de wagons chargés, les ressources en fin d'horizon ne sont pas toutes connues. Afin de ne pas faire chuter les valeurs des ressources effectives en fin d'horizon, il convient donc d'établir une correction pour ces périodes. Soit C , la valeur moyenne des ressources périodiques sur les 3 dernières semaines. On ajoute C à la valeur des ressources effectives de T , $C/2$ à celles de $T-1$ et $C/4$ à celles de $T-2$, etc.

II.3 - Le modèle SNCF

II.3.1 - Modélisation du noyau

Gérer une répartition à moyen terme revient à prendre en compte des demandes effectives et des demandes prévisionnelles en vue d'assurer une régulation et une utilisation satisfaisantes des ressources effectives et des ressources potentielles. Le nombre, la variété et l'évolution des éléments de la répartition en font un système complexe et dynamique. La collecte des informations relatives à la répartition (besoins et ressources) s'avère primordiale. Une mise à jour des données (ressources et demandes) est effectuée avant les 2 "traitements" répartition journaliers.

Les demandes de la clientèle peuvent se faire pour le lendemain. Compte tenu des délais d'acheminement, les décisions d'affectation doivent être anticipées pour pouvoir être mises en oeuvre. Pour cela, une programmation des demandes, sur la base de données prévisionnelles, est nécessaire.

Par la suite, toutes les données (délais, coûts, ressources, demandes) feront référence à la même unité de temps, la période.

La partie dynamique de la formulation reflète les effets de décisions courantes sur les mouvements de wagons, à la fois sur la période actuelle et sur les périodes suivantes. Ces effets sont représentés par l'intermédiaire d'un réseau espace temps, dont les *noeuds* représentent les Centres de Répartition à chacune des périodes considérées tandis que les *arcs* modélisent les mouvements possible des wagons à la

fois dans l'espace et dans le temps. On considère que le *flot* partant de i à la période t et parcourant un arc (i,j) , arrive en j avec un certain délai d'acheminement. Ce type de problème a été étudié dans [For,1957] sous le nom de *Flots dynamiques*.

La recherche du meilleur acheminement se fait séparément du modèle. Il serait intéressant d'intégrer cette recherche dans le modèle opérationnel, mais dans le contexte actuel, la complexité de la structure ne le permet pas.

Soit le réseau $G = (N,E)$, un multigraphe connexe, avec $|N|=n$ et $|E|=m$.

On note,

$N = O \times D \times PI \times F$ l'ensemble des sommets du graphe,

$E = E_1 \times E_2 \times E_3 \times E_4$ l'ensemble des arcs du graphe.

CR l'ensemble des Centres de Répartition,

H l'ensemble des périodes de l'Horizon de Répartition,

ND les niveaux de priorité des demandes,

NR les niveaux de priorité des ressources.

Les sommets sont des couples (gare,période). On distingue les sommets ressources, les points intermédiaires, les sommets demandes, le sommet fictif b pour la balance ressource/demande et le sommet fictif r pour la racine initiale de l'arbre :

$O = \{ i = (g,t) / i \text{ est un sommet ressource, pour } g \text{ dans } CR, t \text{ dans } H \},$

$D = \{ j = (g,t) / j \text{ est un sommet demande, pour } g \text{ dans } CR, t \text{ dans } H \},$

$PI = \{ l = (g,t) / l \text{ est un point intermédiaire pour } g \text{ dans } CR, t \text{ dans } H \},$

$F = \{b,r\}$ l'ensemble des sommets fictifs du graphe.

A chaque sommet i de N , on associe un montant a_i :

$$a_i = \sum_{p=1}^{NR} a_{ip} > 0 \text{ si } i \text{ dans } O,$$

$$a_i = \sum_{p=1}^{ND} a_{ip} < 0 \text{ si } i \text{ dans } D,$$

$$a_i = 0 \text{ si } i \text{ dans } PI,$$

$$a_b = \sum_{i \in D} a_i - \sum_{i \in O} a_i, \text{ où } b \text{ est la balance,}$$

$$a_r = 0, \text{ où } r \text{ est la racine.}$$

Les Arcs de type 1 (ressource, point intermédiaire) : Pour respecter les règles de gestion portant sur les ressources (chapitre I), nous avons dédoublé chaque sommet ressource en deux éléments i et l , de O et PI . Ces deux sommets sont reliés par NR arcs multiples. Par suite, le coût unitaire d'un arc (ilp) sera fonction de la priorité p de la ressource élémentaire et la capacité de cet arc sera fonction du nombre de ressources élémentaires disponibles. Le flot qui passera par ces arcs respectera alors les contraintes décrites précédemment sur l'utilisation des ressources.

$$\begin{aligned} E_1 &= \{ (ilp) / l \in PI, i \in O \}, \\ c_{ilp} &= [(p-1) + (t-1) \cdot NR] \cdot M_1, (ilp) \in E_1 \text{ et } M_1 = \text{Max}\{c_{ij}/(ij) \in E_2\}, \\ k_{ilp} &= a_{lp} \quad \forall (ilp) \in E_1. \end{aligned}$$

Les Arcs de type 2 (point intermédiaire, point intermédiaire) : On joint le point intermédiaire i (période t_i) au point intermédiaire j (période t_j) si l'acheminement de durée d_{ij} est réalisable, c'est à dire si $(t_i + d_{ij}) \leq t_j$. Cette condition a pour conséquence que tous les acheminements entre les points de ressources et les points de demandes ne sont pas possible. Autrement dit, que le graphe biparti, muni des points de ressources et des points de demandes, n'est pas complet. Dans le pire des cas, pour un problème équilibré décrit précédemment, on peut envisager que la demande ne soit pas satisfaite, même si la solution optimale existe. Autrement dit, si on retire du graphe tous les arcs fictifs, on peut mettre en évidence des composantes connexes non équilibrées. La solution de ce problème se trouve dans la structure même du graphe muni des sommets fictifs et des arcs fictifs de type 4.

$$\begin{aligned} E_2 &= \{ (ij) / i \in PI, j \in PI, (t_i + d_{ij}) \leq t_j \}, \\ c_{ij} &= c_{ij} \text{ si } (ij) \in E_2, \\ k_{ij} &= \infty \quad \forall (ij) \in E_2 \end{aligned}$$

Les Arcs de type 3 (point intermédiaire, demande) : Pour respecter les règles de gestion portant sur les demandes (partie I), nous avons dédoublé chaque sommet demande en deux éléments i et l , de D et PI . Ces deux sommets sont reliés par ND arcs multiples. Par suite, le coût unitaire d'un arc (ilp) sera fonction de la priorité p et de la période t de la demande élémentaire et la capacité de cet arc sera fonction du nombre de demandes élémentaires. Le flot qui passera par ces arcs respectera alors les contraintes décrites précédemment sur la satisfaction des demandes.

$$\begin{aligned} E_3 &= \{ (ljp) / l \in PI, j \in D, p \in P \}, \\ c_{ljp} &= [(p-1) + (t-1) \cdot ND] \cdot M_3, (ljp) \in E_3 \text{ et } M_3 = \text{Max}\{c_{ilp} + c_{ij}/(ilp) \in E_1\}, \end{aligned}$$

$$c_{ljp} = T.NR.M_1,$$

$$k_{ljp} = a_{jp} \quad \forall (ljp) \in E_3.$$

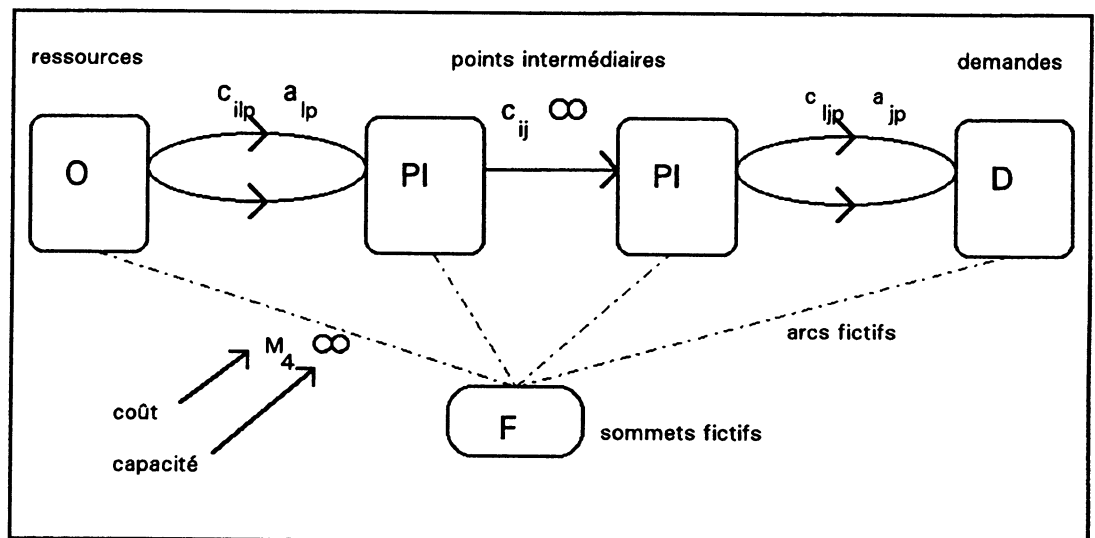
Les Arcs de type 4 (sommet fictif, autre sommet) : En pratique, la somme des ressources est souvent différente de la somme des demandes. Supposons que l'ensemble des ressources disponibles dépasse le total des demandes. Dans ce cas, nous créons une destination artificielle b munie du montant $a_b < 0$ et des arcs artificiels entre les points de ressources et b . Le coût de ces arcs est M et la capacité est infinie. La variable x_{ib} , dans la solution optimale, montre comment sont distribuées les ressources excédentaires. Supposons, au contraire, que l'ensemble des demandes dépasse le total des ressources disponibles. Dans ce cas, nous créons une origine artificielle, b munie du montant $a_b > 0$ et des arcs artificiels entre b et les points de demandes. Le coût de ces arcs est M et la capacité est infinie. La variable x_{bj} , optimale, désigne l'ensemble des demandes j de D non satisfaites. Pour obtenir une solution initiale réalisable, on crée un arbre fictif dont la racine est r et les feuilles sont tous les sommet du graphe. D'une manière générale, les demandes non satisfaites à cause d'acheminements irréalisables, sont couvertes par des ressources acheminées sur des arcs fictifs de type 4..

$$E_4 = \{(i,b) \cup (i,r) \cup (b,j) \cup (r,j) / i \in O, j \in PI \text{ ou } j \in D\}.$$

$$c_{ij} = M_4, (ij) \in E_4 \text{ et } M_4 = \text{Max}\{c_{ilp} + c_{ij} + c_{ljp} / (ilp) \in E_1, (ij) \in E_2, (ljp) \in E_3\}$$

$$c_{ij} = T^2.NR.ND.M_1$$

$$k_{ij} = \infty \quad \forall (i,j) \in E_4.$$



- graphe du problème -

Le problème de la Répartition (P1), correspondant au graphe ci-dessus, est alors le suivant :

$$\begin{aligned}
 \text{Minimiser } z &= \sum_{e \in E-E_4} c_e \cdot x_e + M \cdot \sum_{e \in E_4} c_e \cdot x_e \\
 \text{sc:} & \\
 \sum_{\{j/(ijp) \in E_1\}} x_{ijp} + \sum_{\{j/(ij) \in E_4\}} x_{ij} &= a_i & \forall i \in O \\
 - \sum_{\{j/(jip) \in E_1\}} x_{jip} - \sum_{\{j/(ji) \in E_4\}} x_{jip} + \sum_{\{j/(ij) \in E_2\}} x_{ij} &= a_i & \forall i \in PI \\
 - \sum_{\{j/(ji) \in E_2\}} x_{jip} - \sum_{\{j/(ji) \in E_4\}} x_{ji} + \sum_{\{j/(ijp) \in E_3\}} x_{ij} &= a_i & \forall i \in PI \\
 - \sum_{\{j/(jip) \in E_3\}} x_{jip} - \sum_{\{j/(ji) \in E_4\}} x_{ji} &= a_i & \forall i \in D \\
 \sum_{\{j/(ij) \in E_4\}} x_{ij} - \sum_{\{j/(ji) \in E_4\}} x_{ji} &= a_i & \forall i \in F \\
 0 \leq x_e \leq k_e & & \forall e \in E
 \end{aligned}$$

II.3.2 - Algorithme Répartition (n°1)

Lorsque nous avons débuté notre programme de recherche, nous avons observé que toutes les contraintes ne se situaient pas au même niveau dans la procédure de planification, et qu'elles ne faisaient pas appel aux mêmes données. Nous avons noté aussi qu'il y avait un ordre à respecter dans la séquence des décisions. Nous proposons donc une approche de résolution en plusieurs étapes, avec plusieurs modèles spécifiques à chaque type de catégorie.

ALGORITHME REPARTITION (n°1)

```
{  
    Initialiser l'horizon;  
    Traiter un groupe de catégories indépendantes (Ressources et  
    Demandes initiales);  
    Traiter un groupe de catégories indépendantes (Ressources et  
    Demandes de substitution);  
    Traiter un groupe de catégories indépendantes (Ressources et  
    Demandes de rééquilibrage);  
}
```

Traiter un groupe de catégories indépendantes (Ressources et Demandes)

```
{  
    Tant Que ( il existe une catégorie) Faire  
    {  
        Calculer la répartition;  
    }  
}
```

Initialiser l'horizon

```
{  
     $H_i = \{t, \dots, T\}$ ;          (Horizon courant)  
}
```

Ressources et Demandes initiales

```
{  
    Regrouper, les ressources restantes de l'horizon  $H_{i-1}$ , les nouvelles ressources  
    apparues depuis le dernier recensement et les ressources prévisionnelles.  
    Regrouper, les demandes non satisfaites à l'horizon  $H_{i-1}$ , les demandes apparues  
    depuis le début de l'horizon  $H_{i-1}$  et les demandes prévisionnelles.  
}
```

Ressources et Demandes de substitution

```
{  
    les ressources restantes de l'horizon  $H_i$ ;  
    les demandes non satisfaites à l'horizon  $H_i$ , substituables par une autre catégorie ;  
}
```

Ressources et Demandes de rééquilibrage

```
{  
    les ressources restantes de l'horizon  $H_i$ ;  
    les seuils de rééquilibrage;  
}
```

Calculer la répartition

```
{  
    Construire le graphe : déterminer les ensembles O, PI, D, F et E1, E2, E3, E4;  
    Utiliser l'algorithme de Bazara, pour ce graphe;  
}
```

II.3.3 - Résultats numériques

Nous avons effectué des mesures sur l'application chargée de l'acheminement du FRET. Les mesures ont été faites sur une semaine type du mois de Juillet 1990. Nous avons sélectionné tous les wagons répartis dans des 120 catégories de référence. Les ressources ont été déterminées grâce aux arrivées des wagons chargés et les demandes grâce aux départs des wagons chargés. Ces données nous ont permis d'estimer le nombre moyen de demandes et de ressources, par gare et par catégorie.

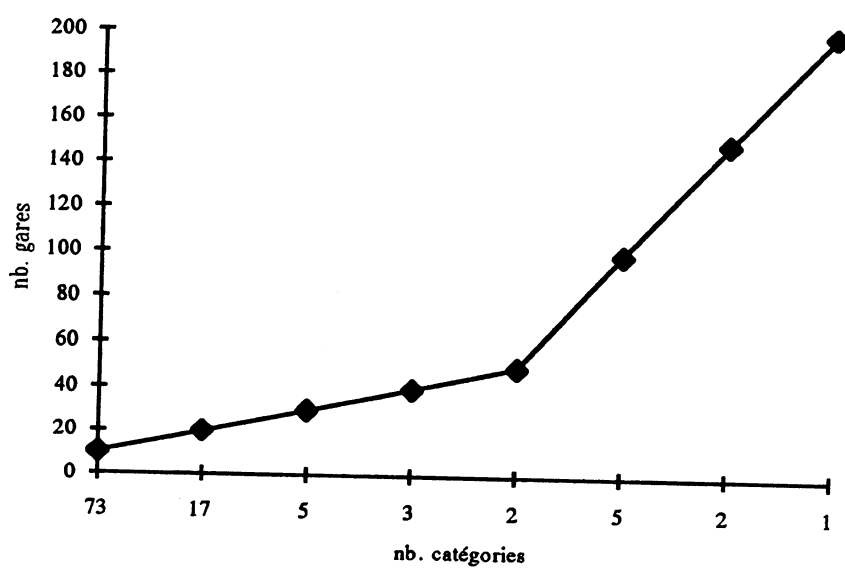
Pour tester le noyau (codé en langage C), nous avons généré différents jeux de données relatifs à une catégorie de wagons. Les machines de test sont des SUN SPARC Station 2, équipées de 32 Mo. de RAM.

Les temps de traitement inscrits dans les tableaux correspondent à la lecture des fichiers de données, au calcul de la solution et à l'écriture de la solution dans le fichier pour la BIG. Nous n'avons pas pris en compte les temps de transfert entre les deux machines, ni le traitement des données reçues de la BIG (procédures de tri, etc.).

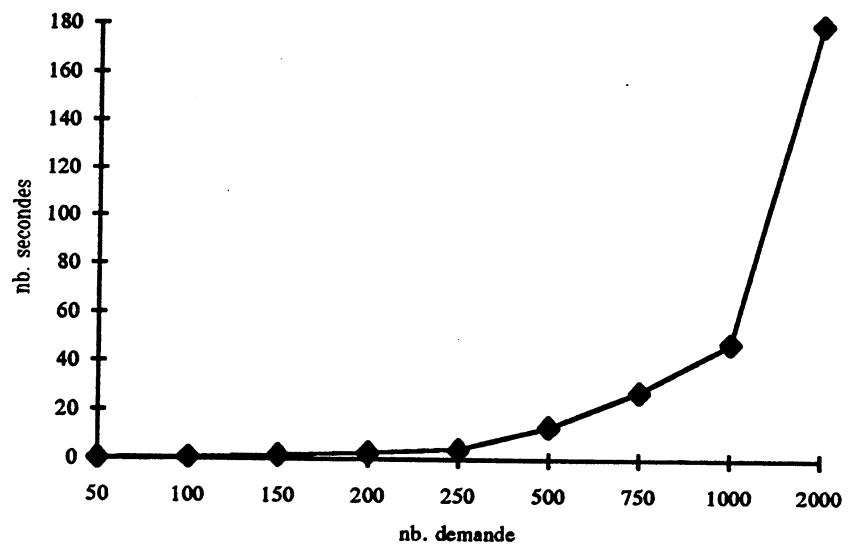
Les nombres d'arcs et de sommets générés pour satisfaire toutes les contraintes de l'application nous ont posé de gros problèmes d'espace mémoire. Après différents tests, nous avons choisi d'introduire le calcul de la dimension exacte du graphe dans l'algorithme et d'utiliser un système d'allocation dynamique de mémoire pour les tableaux d'arcs et de sommets. Cette méthode nous a permis de traiter 2000 demandes et 4800 ressources, en 3 mn. de calcul.

Volume des dem.	nb. categ.	nb. gare	nb. dem.	nb. res.	temps (s)	coûts (k.F.)	nb. pivots
35%	73	10	50	120	0.4	154.92	128
	17	20	100	240	0.9	673.09	282
	5	30	150	360	1.8	751.74	534
	3	40	200	480	2.8	872.91	704
	2	50	250	600	4.3	1008.5	1000
65%	5	100	500	1200	13.3	2290.3	2331
	2	150	750	1800	27.9	3649.3	4264
	1	200	1000	2400	47.8	5392.0	6926
			2000	4800	180.0	11718	14500
			2500	6000	mémoire	saturée	

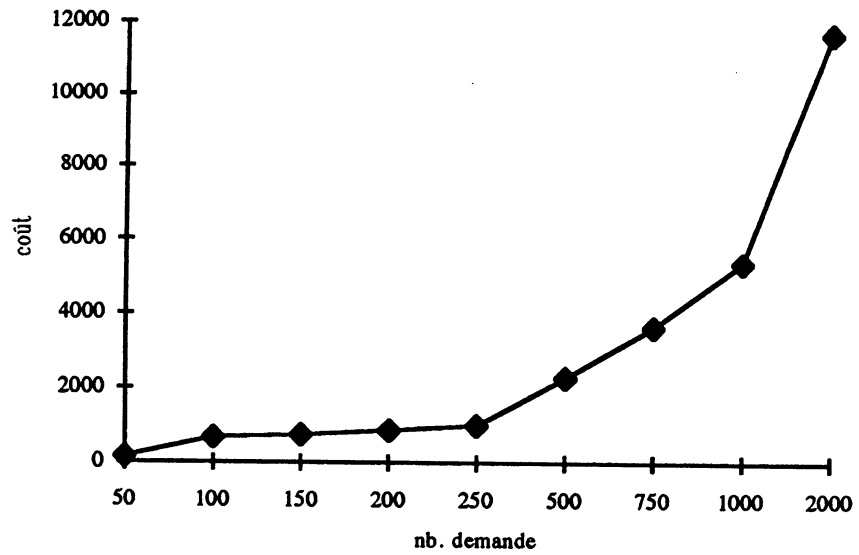
Nombre de catégories demandées par gare

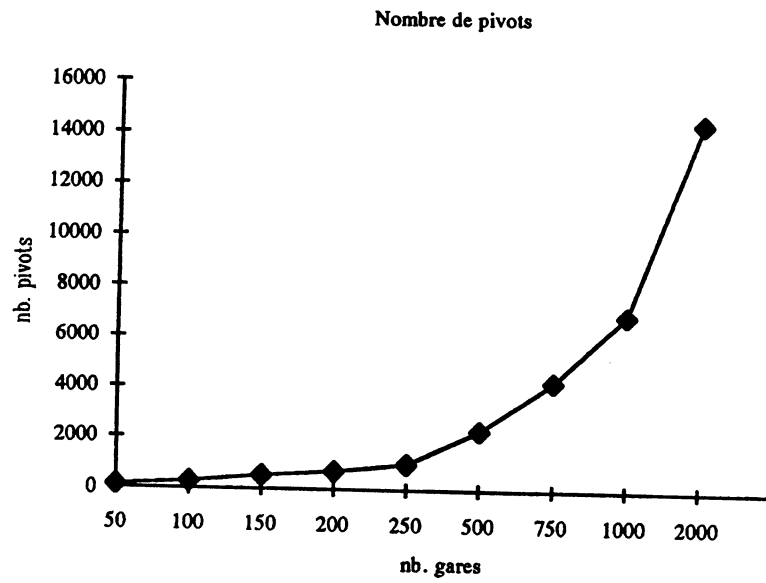


Temps d'exécution des programmes



Coûts des solutions





Grâce à ces résultats, nous avons estimé les temps moyens de traitement, pour les 120 catégories de référence :

nb. caté.	nb. dem.	temps unitaire (s)	temps total (s)
73	50	0.4	29.2
17	100	0.9	15.3
5	150	1.8	9
3	200	2.8	8.4
2	250	4.3	8.6
5	500	13.3	66.5
2	750	27.9	55.8
1	1000	47.8	47.8

Puis, nous avons estimé les temps moyens de traitement, pour les 210 catégories répartissables :

Traitement	temps (s)
90 sous-catégories	180
Substitutions entre sous-catégories	60
catégories étrangères	24
120 catégories de référence	240.6
substitution entre catégories de référence	80.2
rééquilibrages et excédents	240.6
Total	14 mn.

De ces résultats, nous pouvons conclure que l'ensemble du calcul de la Répartition demande au maximum un quart d'heure de temps.

En réalité, les 210 catégories de wagons seront traitées en parallèle, sur 7 stations de travail (SUN SPARC 10). Le temps de calcul proposé ci-dessus sera donc largement inférieur en opérationnel. Par contre le traitement de bout en bout (BIL -> BIG -> BIR -> BIG -> BIL) devrait être réalisé en une heure de temps, compte tenu des temps de transferts et des procédures sur les bases de données.

II.4 - Flexibilité de l'horizon

II.4.1 - Algorithme Répartition (n° 2)

Le principe de la méthode que nous proposons maintenant, est d'assurer une flexibilité de l'horizon. Lorsque le volume de données générées pour calculer la solution dépasse la capacité mémoire de la machine, nous satisfaisons les demandes de l'horizon H en plusieurs étapes :

ALGORITHME REPARTITION (n° 2)

```
{
    Initialiser l'horizon;
    Traiter un groupe de catégories indépendantes (Ressources et
    Demandes initiales);
    Traiter un groupe de catégories indépendantes (Ressources et
    Demandes de substitution);
    Traiter un groupe de catégories indépendantes (Ressources et
    Demandes de rééquilibrage);
}
Traiter un groupe de catégories indépendantes (Ressources et Demandes)
{
    Tant Que ( il existe une catégorie) Faire
        Tant Que (toutes les Périodes de l'horizon courant n'ont pas été traitées) Faire
        {
            Découpler l'horizon traité;
            Calculer la répartition;
            Faire glisser l'horizon traité;
        }
    }
Initialiser l'horizon
{
    Hi = {t,...,T};          (Horizon courant)
    Hj = {p1,...,pn} = Hi;  (Horizon traité)
}
Découpler l'horizon traité
{
    Si (Hj est trop "grand" pour la mémoire de la machine) Alors {pn = pn - 1};
}
Ressources et Demandes initiales
{
    Regrouper, les ressources restantes de l'horizon Hi-1, les nouvelles ressources
    apparues depuis le dernier recensement et les ressources prévisionnelles;
    Regrouper, les demandes non satisfaites à l'horizon Hi-1, les demandes apparues
    depuis le début de l'horizon Hi-1 et les demandes prévisionnelles;
}
Ressources et Demandes de substitution
{
    les ressources restantes de l'horizon Hi;
    les demandes non satisfaites à l'horizon Hi, substituables par une autre catégorie ;
}
Ressources et Demandes de rééquilibrage
{
    les ressources restantes de l'horizon Hi;
    les seuils de rééquilibrage;
}
Calculer la répartition
{
    Construire le graphe : déterminer les ensembles O, PI, D, F et E1, E2, E3, E4;
    Utiliser l'algorithme de Bazara, pour ce graphe;
}
Faire glisser l'horizon traité
{
    p1 = pn + 1; pn = T;
}
```

II.4.2 - Résultats numériques

Le principal intérêt d'un Horizon flexible réside dans le fait que nous sommes à même de déterminer une solution dans tous les cas de figure. En pratique, le simple fait d'obtenir une solution est nécessaire. Seulement la question que l'on se pose est de savoir de combien la solution est détériorée lorsqu'on travaille avec un Horizon flexible, par rapport à un Horizon complet. Nous allons maintenant étudier les coûts et les temps de calcul.

Proposition 1 :

Le coût total de la solution augmente avec le nombre de sous-Horizons.

Démonstration :

On considère le problème (P1) et l'horizon de Répartition H de T périodes. On appelle z , le coût de la solution optimale de (P1) sur H.

On considère maintenant que H est décomposable en k sous-horizons. On appelle H_i , le $i^{\text{ème}}$ sous-horizon de H et z_i le coût de la solution optimale de (P1) sur H_i . Soit d^k un découpage de H en k sous-horizons ($T \geq k$)

On note dz_k , la différence entre la somme des coûts des solutions partielles z_i ($i=1, \dots, k$) et le coût de la solution optimale z .

$$dz_k = \sum_{i=1}^k z_i - z$$

Notons que les demandes de l'horizon H1 sont satisfaites avant les demandes de l'horizon H2, conformément aux règles de gestion du chapitre I.

Considérons le graphe $G=[N,E]$, avec $N=\{1,2,3,4\}$ et $E=\{(1,3),(2,3),(1,4),(2,4)\}$, tel que :

- le sommet 1 est muni de α ressources, sur l'horizon H1;
- le sommet 2 est muni de α ressources, sur l'horizon H1;
- le sommet 3 est muni de α demandes, sur l'horizon H1;
- le sommet 4 est muni de α demandes, sur l'horizon H2;

- les arcs (1,3) et (2,3) sont munis d'un coût nul;
- l'arc (1,4) est muni d'un coût égal à 10;
- l'arc (2,4) est muni d'un coût très grand, égal à M.

La valeur de dz_2 , pour ce graphe, est égale à M. Nous pouvons alors affirmer qu'il existe un découpage de H en 2 sous-horizons, pour lequel la somme des solutions partielles est très éloignée de la solution optimale.

Pour $k=n-1$, supposons qu'il existe $d^2, \dots, d^{n-1}$ tels que $(0 < dz_2 < \dots < dz_{n-1})$.

pour $k=n$, montrons qu'il existe $d^2, \dots, d^n$ tels que $(0 < dz_2 < \dots < dz_{n-1} < dz_n)$.

D'après l'hypothèse, il reste à montrer qu'il existe d^{n-1}, d^n tels que $(dz_{n-1} < dz_n)$.

Soient d^n et d^{n-1} tels que $dz_n=0$ et $dz_{n-1} < 0$, alors il existe d^{n-1} et d^n tels que $(dz_{n-1} < dz_n)$.

Proposition 2 :

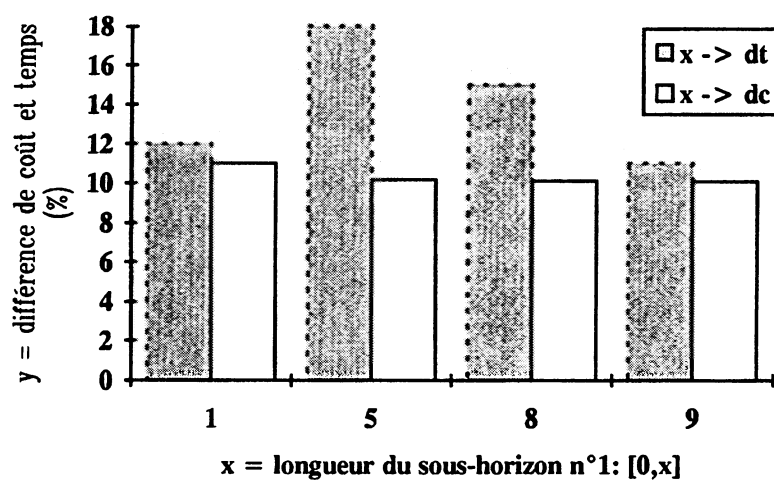
Le coût total de la solution diminue quand la longueur de H1 augmente.

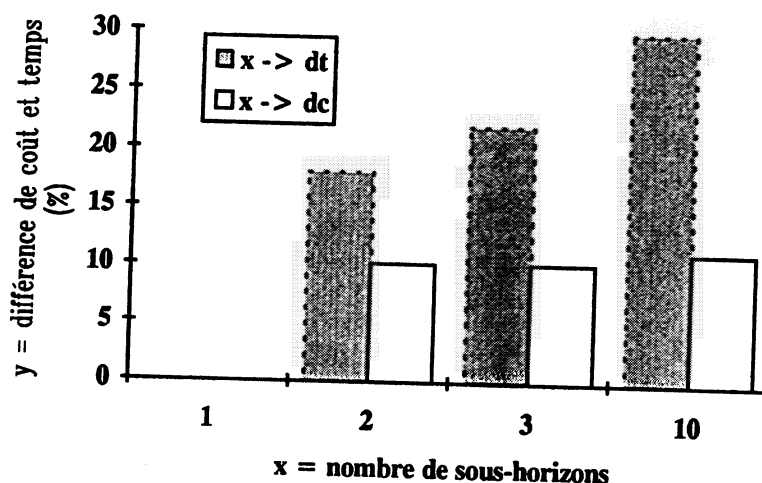
Démonstration :

Cette proposition découle directement des règles de gestions définies dans la partie I de ce rapport. En effet, grâce aux coûts fictifs et aux capacités sur les arcs de type 1 et 3, nous satisfaisons les demandes dans l'ordre chronologique des périodes et nous utilisons les ressources effectives avant les ressources prévisionnelles. Ainsi, si on découpe l'horizon après la première période, nous réduisons les possibilités d'affectation par rapport à un découpage après la $(T-1)^{\text{ème}}$ période.

Nous avons effectué différents tests comparatifs entre 2 méthodes : l'une assurant la solution optimale sur l'horizon complet et l'autre assurant une solution avec un horizon flexible. Pour cela, nous avons généré des jeux de données relatifs à un horizon H de 10 périodes, puis forcé le découpage de H en 2, 3 et 10 sous-horizons. Dans le cas d'un découpage en 2 sous-horizons (H1, H2), nous avons fait varier la longueur de H1 et reporté, dans le tableau qui suit, les différences de coûts et de temps obtenues, ainsi que le ratio entre ces deux valeurs.

nb. sous- horizons k	longueur des H _i i=1,...,k	temps sans décomp. (s)	temps avec décomp. (s)	%	coût sans décomp. (F)	coût avec décomp. (F)	%	ratio coût/temps
2	9,1	54	49	11	23064	23268	10.08	34
2	8,2	54	35	15	23064	23352	10.12	15
2	5,5	54	29	18	23064	23391	10.14	13
2	1,9	54	44	12	23064	25109	10.88	204.5
3	3,3,4	54	24	22	23064	23480	10.18	14
10	1,1,...,1	54	18	30	23064	25923	11.24	79





Ces résultats confirment la proposition 1 : plus on découpe l'horizon complet en sous-horizons, plus le coût est élevé (10 sous-horizons de 1 période). Par contre, plus on découpe en sous-horizons, plus les temps d'exécution sont faibles (10 sous-horizons de 1 période).

Ces résultats concordent avec la proposition 2 : lorsqu'on découpe l'horizon en 2 sous-horizons, la différence de coût décroît lorsque la longueur de H1 augmente ($H1=\{1,...,9\}$ et $H2=\{10\}$), mais la différence de temps est maximale lorsqu'on découpe l'horizon à la moitié ($H1=\{1,...,5\}$ et $H2=\{6,...,10\}$)

Si on combine le coût et le temps, il est possible de déterminer un découpage qui donne un rapport coût/temps optimal. Dans notre exemple, c'est un découpage en 2 sous-horizons tel que $H1=\{1,...,5\}$ et $H2=\{6,...,10\}$.

II.5 - Extension du modèle SNCF

Si on considère le problème de la répartition dans son ensemble, il apparaît clairement que la notion de catégorie de wagon est à la base du problème. Nous avons jusqu'à présent ignoré cette contrainte et considéré que le traitement d'une catégorie était indépendant des autres. Pour traiter toutes les catégories, on a le choix entre élaborer un traitement différencié par famille de catégories (fine, réseau, référence,...) ou mettre en place un modèle complet qui soit capable de traiter globalement le problème quelque soit le type de matériel demandé. Ce modèle complet nous permettrait de prendre en compte la notion de *substitution*, de *rééquilibrage* et d'*excédent*.

La possibilité de substituer une catégorie de wagon avec une autre est la caractéristique la plus difficile à formaliser. En effet, il faut distinguer, dans le modèle, les wagons qui sont physiquement de la catégorie traitée, de ceux qui sont substitués. Notons que le nombre des substitutions est borné (car elles sont demandées par le client) et que le nombre demandé est toujours entier.

II.5.1 - Modélisation du problème complet

Les traitements des catégories de référence, mis à part les substitutions de wagons, sont indépendants les uns des autres. L'objectif de cette nouvelle modélisation du problème, est de traiter une catégorie de wagons (comme dans le noyau du modèle), avec les substitutions, le rééquilibrage et les excédents. Notons qu'un modèle regroupant toutes les catégories n'est pas nécessaire.

Pour une gare et une période données, nous découpons la demande totale de la catégorie c en :

- 1 - demandes prioritaires sans substitution,
- 2 - demandes prioritaires avec substitutions,
- 3 - demandes fermes sans substitution,
- 4 - demandes fermes avec substitutions,
- 5 - demandes prévisionnelles sans substitution,
- 6 - demandes prévisionnelles avec substitutions,
- 7 - demandes prioritaires des autres catégories substituables par c ,
- 8 - demandes fermes des autres catégories substituables par c ,
- 9 - demandes prévisionnelles des autres catégories substituables par c .
- 10 - seuils minima.
- 11 - seuils maxima.

Nous associons alors un niveau de priorité à chaque type de demandes, qui seront satisfaites dans l'ordre ci-dessus. La modélisation est alors la même que précédemment, avec le problème (P1).

II.5.2 - Algorithme Répartition (n° 3)

```
ALGORITHME REPARTITION (n° 3)
{
    Initialiser l'horizon;
    Traiter un groupe de catégories indépendantes (Ressources et
    Demandes );
}

Traiter un groupe de catégories indépendantes (Ressources et Demandes)
{
    Tant Que ( il existe une catégorie) Faire
    {
        Tant Que (toutes les Périodes de l'horizon courant n'ont pas été traitées) Faire
        {
            Découpler l'horizon traité;
            Calculer la répartition;
            Faire glisser l'horizon traité;
        }
    }
}

Initialiser l'horizon
{
     $H_i = \{t, \dots, T\};$  (Horizon courant)
     $H_j = \{p_1, \dots, p_n\} = H_i;$  (Horizon traité)
}

Découpler l'horizon traité
{
    Si ( $H_j$  est trop "grand" pour la mémoire de la machine) Alors  $\{p_n = p_n - 1\};$ 
}

Ressources et Demandes
{
    Regrouper, les ressources restantes de l'horizon  $H_{i-1}$ , les nouvelles ressources
    apparues depuis le dernier recensement et les ressources prévisionnelles;
    Regrouper, les demandes non satisfaites à l'horizon  $H_{i-1}$ , les demandes apparues
    depuis le début de l'horizon  $H_{i-1}$ , les demandes substituables par une autre
    catégorie, les seuils de rééquilibrage et les demandes prévisionnelles;
}

Calculer la répartition
{
    Construire le graphe : déterminer les ensembles O, PI, D, F et E1, E2, E3, E4;
    Utiliser l'algorithme de Bazara, pour ce graphe;
}

Faire glisser l'horizon traité
{
     $p_1 = p_n + 1; p_n = T;$ 
}
}
```

II.5.3 - Les limites de l'approche globale

La formulation du problème complet nous permet de traiter les *substitutions* au niveau des demandes. Seulement le nombre de contraintes et de variables de ce problème est alors beaucoup trop élevé pour être traité globalement comme nous l'avons fait jusqu'à présent.

Nous proposons donc une autre approche, qui consiste à décomposer le problème général en sous-problèmes de dimensions et de complexité acceptables. Nous étudions, au chapitre suivant, plusieurs techniques de décomposition.

Approche par décomposition

Chapitre III

III.1 - Introduction

III.1.1 - Objectifs des méthodes de décomposition

L'objectif principal des méthodes de décomposition en Programmation Mathématique consiste à transformer des problèmes d'Optimisation de "grande dimension" en la résolution séquentielle ou parallèle de sous-problèmes de "plus petite taille". D'une manière générale, l'approche par décomposition est utilisée lorsqu'on recherche un ou plusieurs des cinq points suivants :

Réduire la dimension du problème, lorsque le nombre de variables et/ou de contraintes est trop élevé, afin de se ramener à la résolution de sous-problèmes de taille raisonnable (cf. Lasdon [Las,1970], Minoux [Min,1983]).

Hiérarchiser, dans le cadre d'un processus de décision à plusieurs niveaux, où chaque unité économique possède une unité de contrôle qui assure la coordination entre les sous-systèmes du niveau inférieur (cf. Titli [Titli,1975]).

Décentraliser un problème en le décomposant en sous-problèmes indépendants ayant leurs propres données, capables de trouver et reconnaître leurs propres solutions optimales (cf. Bensoussan, Crouhy et Proth [Ben,1974], Arrow et Hurwicz [Arr,1960]).

Partitionner, un problème aux données hétérogènes (par exemple avec des variables entières et discrètes) en sous-problèmes homogènes (cf. Benders [Ben,1962]).

Paralléliser, en accord avec l'évolution du matériel vers le calcul parallèle (multitâche, multiprocesseur) (cf. Bertsekas [Ber,1989]).

III.1.2 - Approche conceptuelle du problème

Une approche par décomposition comprend, en général, deux phases que nous présentons brièvement ci-dessous :

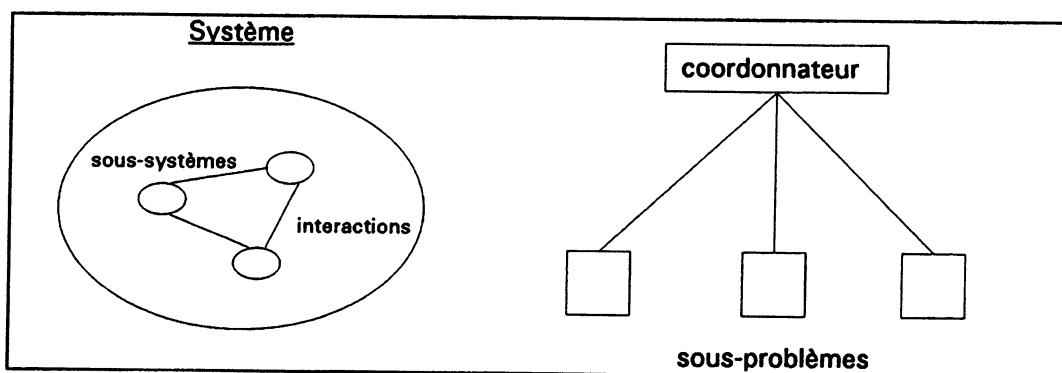
Phase 1 : Analyse structurale du problème

Pour transformer un problème d'Optimisation de grande dimension, cela sous-entend l'existence d'une structure sous-jacente difficile à exploiter dans une approche globale. Cette phase demande une analyse structurale du problème dans le but d'identifier une structure conduisant à un découplage maximal des sous-systèmes (régions, secteurs d'une économie, divisions d'une entreprise,...) supervisés par un coordonnateur (centre de planification, direction d'une entreprise,...) chargé d'établir et d'équilibrer les liens entre les sous-systèmes.

Dans le cadre de la Répartition, nous pouvons identifier différentes structures et sous-systèmes, relativement aux décompositions temporelles, spatiales et logiques (catégories de matériels).

Phase 2 : Organisation des calculs.

Cette phase consiste à manipuler le problème d'optimisation initial de manière à mettre en évidence un algorithme, généralement itératif, qui lui est adapté. Les sous-problèmes, traités en séquentiel ou en parallèle, sont supervisés par un niveau de coordination (programme maître), qui tient compte du couplage éventuel entre les sous-systèmes mesuré par un certain nombre de paramètres de coordination.



- Les deux phases de l'analyse -

III.1.3 - Transformation du problème initial

Séparabilité du problème

En réalité, le passage de la phase 1 à la phase 2 demande une manipulation intermédiaire, lorsque le problème initial n'est pas directement "séparable" en sous-problèmes. Nous allons tout d'abord définir d'une manière générale la séparabilité d'une fonction convexe, puis présenter la notion de faible séparabilité.

Soit (S0) le problème général d'optimisation, dans l'espace vectoriel X . On se place ici dans le cas où X est un espace de dimension fini, tel que $X = \mathbb{R}^n$.

$$(S0) \quad \begin{cases} \text{Minimiser } F(x) \\ \text{sc.} \\ x \in S \end{cases}$$

où F est une fonction définie de X dans $\mathbb{R}$ et S est un sous-ensemble de X .

Définition :

S'il existe une partition des variables en K sous-ensembles disjoints, telle que $X = X_1 \times X_2 \times \dots \times X_K$, on dit que F est une *fonction séparable* en x par rapport à cette partition, si et seulement si F est une combinaison linéaire de fonctions définies sur les espaces X_i ($i=1, \dots, K$).

Soit (S1) le problème d'optimisation :

$$(S1) \left\{ \begin{array}{l} \text{Minimiser } F(x) = \sum_{i=1}^K f_i(x_i) \\ \text{sc.} \\ x \in S = \{x = (x_1, \dots, x_K) / x_i \in S_i; i = 1, \dots, K\} \end{array} \right.$$

où les S_i sont des sous-ensembles des X_i ($i=1, \dots, K$) et f_i est une fonction définie sur X_i , à valeurs dans R .

Dans ce cas, la fonction F est une fonction séparable en x et pour résoudre le problème (S1) il suffit de résoudre les sous-problèmes $(S1_i)$ suivants :

$$(S1_i) \left\{ \begin{array}{l} \text{Minimiser } f_i(x_i) \\ \text{sc.} \\ x_i \in S_i \end{array} \right.$$

Définition :

Si la fonction F est séparable en x , par rapport à une partition $X = X_1 \times X_2 \times \dots \times X_K$, on dit alors que le problème (S1) est *séparable* pour cette partition.

Faible séparabilité du problème

En général, les problèmes réels ne sont pas séparables. Une situation favorable à laquelle on peut souvent se ramener est celle où une grande partie du problème est séparable et un nombre réduit de contraintes (ou de variables) réalisent le "couplage" entre les sous-systèmes, par la partition.

On dira qu'un problème est faiblement séparable s'il est de la forme suivante:

$$(S2) \left\{ \begin{array}{l} \text{Minimiser } F(x) = \sum_{i=1}^K f_i(x_i) \\ \text{s.c.} \\ \sum_{i=1}^K g_i(x_i) \leq 0 \\ x_i \in S_i \quad \forall i = 1, \dots, K \end{array} \right. \quad (1)$$

où les fonctions f_i sont définies sur X_i , à valeurs dans $\mathbb{R}$ et les fonctions g_i sont définies sur X_i , à valeurs dans $\mathbb{R}^m$.

Les contraintes (1) représentent le couplage entre les différents sous-systèmes associés à la partition $X = X_1 \times X_2 \times \dots \times X_K$.

Exemple :

Soit (S3) le programme linéaire de grande dimension, sous forme standard et faiblement séparable:

$$\left\{ \begin{array}{l} \text{Minimiser } z = c \cdot x = \sum_{i=1}^K c_i \cdot x_i \\ \text{s.c.} \\ \sum_{i=1}^K A_i \cdot x_i = b \quad (1) \\ D_i \cdot x_i = d_i \quad \forall i = 1, \dots, K \quad (2) \\ x_i \geq 0 \quad \forall i = 1, \dots, K \quad (3) \end{array} \right.$$

Ici $S_i = \{x_i \in X_i / D_i \cdot x_i = d_i, x_i \geq 0\}$ et la matrice Δ des contraintes du type (1) et (2) est dite bloc-angulaire :

$$\Delta = \begin{pmatrix} A_1 & A_2 & . & . & . & A_K \\ D_1 & & & & & \\ & D_2 & & & & \\ & & . & & & \\ & & & . & & \\ & & & & . & \\ & & & & & D_K \end{pmatrix}$$

Aux K blocs de la matrice Δ correspondent des partitions en blocs :

- . du vecteur c en sous-vecteurs $(c_1, c_2, \dots, c_K)$;
- . du vecteur x en sous-vecteurs $(x_1, x_2, \dots, x_K)$;
- . de la matrice A en sous-matrices $(A_1, A_2, \dots, A_K)$;
- . du vecteur d en sous-vecteurs $(d_1, d_2, \dots, d_K)$.

(S3) est un Programme Linéaire de grande dimension, difficile à exploiter dans une approche globale. Dans la pratique, il est excessivement rare de rencontrer une matrice de contraintes sans structure particulière. Ces matrices de grandes dimensions sont en général très creuses, avec K blocs en diagonale et des contraintes couplantes, ou K blocs en diagonale et des variables couplantes, ou enfin K blocs en diagonale avec contraintes et variables couplantes. Notons que les sous-matrices, associées aux K blocs, sont alors de faibles dimensions et plus facilement exploitables (cf. [Min,1983]).

III.1.4 - Classification de différentes méthodes

Le traitement des problèmes faiblement séparables demande une étude particulière. En effet il n'existe pas de recettes précises pour manipuler ces problèmes. Selon leurs structures, la transformation sera différente et c'est une des difficultés de la décomposition. Pour illustrer ce point, nous présentons deux manipulations classiques:

La *Relaxation Lagrangienne* des contraintes couplantes du problème, suivant une approche duale du problème (cf. Lasdon [Las,1970], Geoffrion [Geo,1974] et Magnanti [Mag,1976]).

L'*ajout de variables de couplage*, suivant une approche Primale du problème (cf., par exemple, Geoffrion [Geo,1968] et Silverman [Sil,1968]).

Illustrons ces manipulations par deux exemples :

Exemple 1 :

Reprenons le problème (S2), défini précédemment, avec une hypothèse supplémentaire de convexité. On suppose maintenant que les fonctions f_i et g_i (pour $i=1,\dots,K$) sont convexes, propres et semi-continues inférieurement et que les sous ensembles S_i de X_i sont convexes et fermés. La manipulation proposée, pour ce problème, est une relaxation des contraintes couplantes (1).

Soit (S4) le problème dual de (S2) où u est le vecteur des multiplicateurs de Kuhn et Tucker associé aux contraintes (1). Le "programme maître" (S4) est faiblement séparable et se définit comme suit:

$$\text{Maximiser } h(u) = \sum_{i=1}^K h_i(u^i)$$

et (S4i) les sous-problèmes associés :

$$h_i(u^i) = \text{Minimiser } (f_i(x_i) + \langle u^i, g_i(x_i) \rangle)$$

La fonction duale h est concave, mais généralement non différentiable.

Pour tout x réalisable et $u \geq 0$, on sait que $F(x) \geq h(u)$. Donc si (S2) a une solution optimale x^* , alors (S4) a également une solution optimale u^* et $F(x^*) = h(u^*)$.

Exemple 2 :

Reprenons le problème (S2) et l'hypothèse de convexité.

La manipulation proposée maintenant, consiste à ajouter des variables de couplage y_i (pour $i=1, \dots, K$). Nous obtenons le problème (S5). Ce problème est faiblement séparable et le couplage est cette fois représenté par un sous-espace de structure très simple.

Soit (S5) le problème maître, faiblement séparable et équivalent à (S2) :

$$\begin{cases} \text{Minimiser } V(y) = \sum_{i=1}^K v_i(y_i) \\ y = (y_1, \dots, y_K) \\ \text{s.c.} \\ \sum_{i=1}^K y_i = 0 \end{cases}$$

et (S5i) les sous-problèmes associés :

$$v_i(y_i) = \text{Minimiser} \{ f_i(x_i) / g_i(x_i) \leq y_i ; x_i \in S_i \}$$

où on suppose que pour tout y_i , il existe un optimum de valeur finie. De plus, la fonction V est convexe, mais généralement non différentiable.

Si (S2) a une solution optimale x^* , alors (S5) a également une solution optimale y^* et $F(x^*) = V(y^*)$.

L'organisation des calculs se fait ensuite autour d'un algorithme itératif entre le problème maître et les sous-problèmes. Dans les exemples ci-dessus, les itérations se font par rapport à u , pour le problème maître (S4) et par rapport à y , pour (S5), avec une projection sur le sous-espace relatif au couplage. Pour les sous-problèmes (S4i) et (S5i), les itérations se font par rapport à x , pour des valeurs de u et y fixées.

Nous présenterons tout d'abord les approches classiques pour résoudre les problèmes maîtres. Pour remédier aux difficultés inhérentes à ces méthodes (lenteur et décentralisation partielle), on a pu proposer les méthodes de décomposition mixte et de régularisation.

Méthodes classiques

Une grande partie des algorithmes classiques de décomposition trouvent leurs bases théoriques dans l'optimisation non différentiable. On trouve, par exemple, les trois méthodes suivantes :

Les *méthodes de sous-gradients*, étudiées par des chercheurs russes de l'école de Kiev dans les années 60 et maintenant classiques, consistent à mettre à jour les paramètres de couplage dans la direction d'un sous-gradient de la fonction objectif, en effectuant une série de "petits pas". La théorie (cf. Shor [Sho,1985] ou Minoux [Min,1983]) nous dit que ces pas doivent tendre vers zéro, mais pas trop vite. Bien que ces méthodes soient réputées lentes, elles sont couramment utilisées car elles demandent un faible coût d'implémentation, sont peu sensibles à la dimension du problème de coordination et génèrent des trajectoires en zigzag de part et d'autre des surfaces de non différentiabilité de la fonction. Cette méthode a été appliquée avec succès sur de nombreux problèmes de grandes dimensions. On peut citer en particulier les travaux de Held et Karp [Hel,1971], Held, Wolfe et Crowder [Hel,1974], Legendre et Minoux [Leg,1977]), Minoux et Serreault [Min,1981] et de Minoux [Min,1983].

Les *méthodes de plan de coupe*, appliquées à la maximisation de la fonction objectif, consistent à remplacer cette fonction par un ensemble de fonctions élémentaires. Elles sont définies comme des enveloppes inférieures, des supports affines de la fonction objectif, associées aux sous-gradients générés au cours des itérations précédentes. On trouve comme application l'algorithme de Dantzig-Wolfe [Dan,1961] ou l'algorithme de Benders [Ben,1962].

L'investissement vers des méthodes plus performantes comme les *méthodes de faisceaux* introduites par Lemaréchal en 1975 [Lem,1975] n'a été fait que tardivement dans le cas des techniques de décomposition (cf. Medhi [Med,1987], par exemple).

La non différentiabilité présente un des grands inconvénients de ces méthodes classiques, l'instabilité en est une conséquence. D'où la nécessité de trouver d'autres algorithmes permettant d'éviter la non différentiabilité, tout en préservant la séparabilité du problème. Les méthodes de décomposition mixte ou de régularisation proximale en sont des exemples.

Méthode de décomposition mixte

C'est l'analyse de la dégénérescence, introduite dans les sous-problèmes par la décomposition, qui est à l'origine de la méthode de décomposition mixte présentée plus loin.

Il s'agit d'une méthode mixte, dans le sens où elle combine une approche duale et primale du problème. Chaque sous-problème est décentralisé au voisinage de l'optimum. Une méthode de sous-gradient peut être utilisée dans une première phase pour approcher la solution optimale duale. Dans une dernière phase, on retrouve une procédure à un niveau, c'est à dire que les sous-problèmes s'échangent les paramètres de coordination sans l'aide d'un niveau supérieur de coordination. On trouve dans la littérature, les travaux de Cohen [Coh,1984], et Mahey [Mah,1986].

Méthode de régularisation

Il s'agit de la troisième ligne de recherche motivée par la construction de méthodes de décomposition décentralisées. Les travaux de Spingarn en 1983 [Spi,1983] et le

rapprochement avec les méthodes de Lagrangien Augmenté ont conduit aux techniques de régularisation proximale [Mah,1992]. Elles consistent à ajouter un noyau quadratique afin de transformer un problème convexe en une suite de problèmes fortement convexes et ainsi garantir la différentiabilité, la séparabilité et l'unicité de la solution.

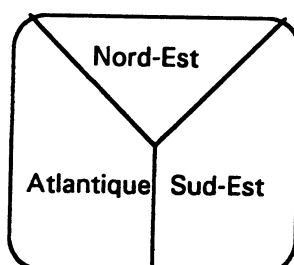
Pour résoudre le problème de transbordement de grande dimension (P0), nous proposons dans un premier temps, les deux manipulations classiques définies précédemment : la relaxation Lagrangienne, suivant une approche duale du problème et les variables de couplage, suivant une approche primale du problème. En accord avec ces manipulations, nous avons choisi d'expérimenter une méthode de sous-gradient. Dans un deuxième temps, nous proposons une méthode de décomposition mixte, à partir des résultats précédents.

III 1.5 - Structure du système SNCF

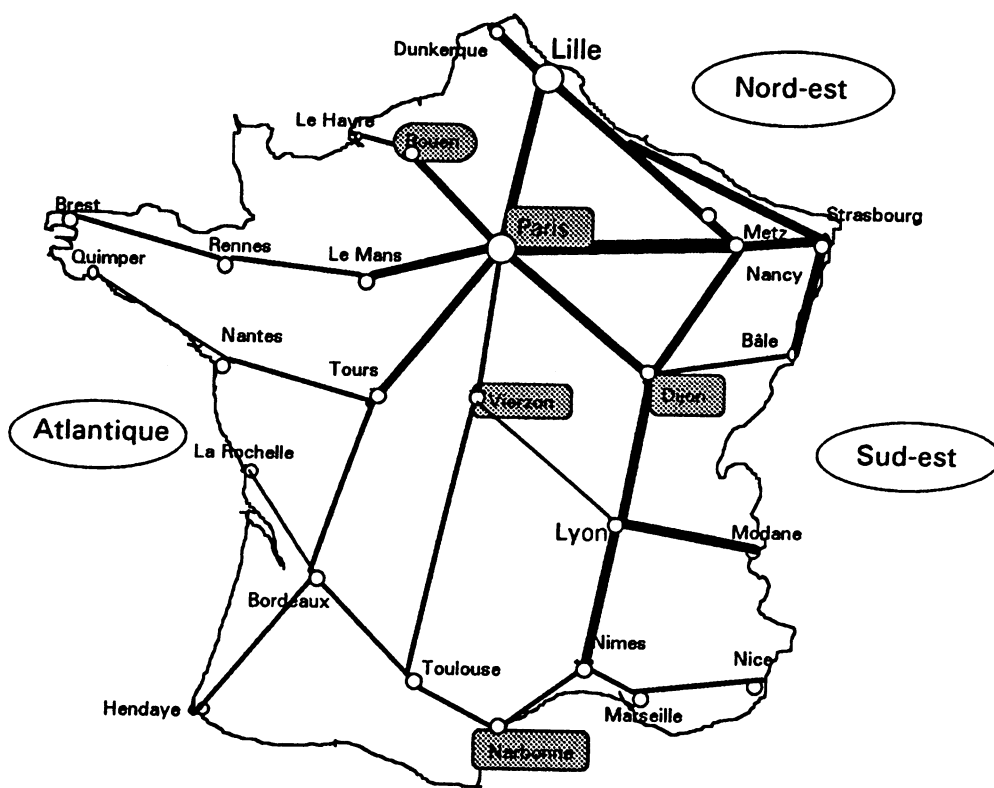
Dans la littérature, la plupart des problèmes de Transport de "grandes tailles", sont traités comme des problèmes de Multi-flots (cf. [Min,1983]), ou décomposés pour chaque arc (cf. [Ber,1992]). Dans ce chapitre, nous proposons une nouvelle approche, adaptée au problème de la SNCF.

Nous avons, dans ce rapport, présenté le problème de la Répartition des wagons vides à court terme. Nous avons vu que ce problème nous conduisait à la résolution d'un problème de Transbordement. Seulement, une approche globale de ce problème, bien que très efficace, génère dans le pire des cas un trop grand nombre de variables et de contraintes par rapport aux possibilités mémoire de l'ordinateur. Nous avons contourné ce problème en découpant l'horizon de planification, pour résoudre successivement des problèmes de Transbordement non indépendants et de dimension acceptable. Seulement, de cette manière, nous nous éloignons sensiblement de l'optimum. Nous proposons donc, dans ce chapitre, une décomposition spatiale du problème. Notre objectif est maintenant de décomposer le problème initial en sous problèmes de dimensions et de complexité acceptables, afin de calculer la solution optimale du problème.

Le réseau SNCF est structuré en trois grandes régions ou sous-systèmes :



Certains Centres de Répartition jouent le rôle de *Points Frontières* entre deux régions voisines (2 ou 3 par paire de zone). Pour simplifier les traitements, sans s'éloigner de la réalité, nous décomposons Paris en trois Points Frontières.



Légende :

- Trafic FRET par section de ligne (milliers de tonnes —)
- Définition des sous-systèmes 
- Interactions par les Points frontières 

III 1.6 - Notations

On note :

F, l'ensemble des points frontières;

R, l'ensemble des sous-systèmes ou sous-problèmes.

On appelle *sous-problème excédentaire*, un sous-problème pour lequel la somme des ressources est supérieure à la somme des demandes.

On appelle *sous-problème déficitaire*, un sous-problème pour lequel la somme des demandes est supérieure à la somme des ressources.

D'une manière générale, les notations relatives aux sous-systèmes excédentaires seront suivies de l'exposant + et les notations relatives aux sous-systèmes déficitaires seront suivies de l'exposant -. Par exemple, nous noterons :

R^+ , l'ensemble des sous-systèmes ou sous-problèmes excédentaires;

R^- , l'ensemble des sous-systèmes ou sous-problèmes déficitaires.

Par ailleurs, les notations relatives aux sous-systèmes seront indicées par r, à valeurs dans R. Par exemple, nous noterons :

$G_r = (N_r, E_r)$ le graphe de la zone r.

Enfin, les notations relatives aux points frontières seront indicées par f, à valeurs dans F. Par exemple, l'équilibre au point frontière f s'écrira :

$$\sum_{\{j/j \notin F\}} x_{fj} - \sum_{\{j/j \in F\}} x_{jf} = a_f$$

Par la suite, lorsqu'un qu'un sommet i sera élément du graphe de la zone r ($i \in N_r$), cela sous-entendra que i n'est pas un point frontière ($i \notin F$).

Nous reprenons le problème de transbordement (P0), avec les hypothèses suivantes :

$$\left\{ \begin{array}{l} \text{Minimiser } z = \sum_{(i,j) \in E} c_{ij} \cdot x_{ij} \\ \text{s.c.} \\ \sum_{\{j/(ij) \in E\}} x_{ij} - \sum_{\{j/(ji) \in E\}} x_{ji} = a_i \quad \forall i \in N \quad (1) \\ 0 \leq x_{ij} \leq k_{ij} \quad \forall (i,j) \in E \quad (2) \end{array} \right.$$

Hypothèses :

- La solution optimale de (P0) , notée $x^* = (x_1^*, \dots, x_m^*)$, existe , est unique et non dégénérée. Cette hypothèse implique que la solution optimale duale de (P0) est unique et non dégénérée. On notera u^* , le vecteur des multiplicateurs optimaux, associés aux contraintes (1).
- Tout arc du graphe relie deux gares de la même zone. Donc tous les flux inter-zones s'effectuent par les points frontières.
- Tout point frontière permet le passage d'une unique zone excédentaire vers une unique zone déficitaire.
- Les coûts d'acheminement à l'intérieur d'une zone sont tous inférieurs aux coûts d'acheminement, par un point frontière, avec une gare d'une autre zone.

III.2 - Décomposition duale

Nous proposons, dans un premier temps, une technique duale de décomposition. Dans cette approche duale, le niveau de coordination ajuste itérativement un vecteur de "prix" associé aux contraintes de couplage et l'introduit dans la fonction objectif de chaque décideur local. En terme de Programmation Mathématique, c'est une méthode duale consistant à traiter les contraintes couplantes de (P0) par relaxation Lagrangienne.

Le principe est d'obtenir indirectement une répartition optimale des ressources excédentaires entre les sous-systèmes, par l'intermédiaire de prix. On recherche itérativement une affectation minimisant la somme des coûts des solutions partielles obtenues. A chaque étape, les solutions des R sous-problèmes permettent, soit de conclure à l'optimalité de la répartition courante, soit de déterminer une nouvelle répartition conduisant à une diminution du coût.

III.2.1 - Relaxation des contraintes de couplage

La première étape de cette approche duale est de manipuler le problème (P0) pour obtenir le problème séparable (P1). Pour cela, nous transformons l'équation (1) en :

$$\sum_{\{j/(ij) \in E_r\}} x_{ij} - \sum_{\{j/(ji) \in E_r\}} x_{ji} = a_i \quad \forall i \in F, \forall r \in R \quad (3.1)$$

$$\sum_{\{j/j \in F\}} x_{fj} - \sum_{\{j/j \in F\}} x_{jf} = a_f \quad \forall f \in F, \quad (3.2)$$

La contrainte (3.2) est la contrainte de couplage du problème et nous définissons le vecteur u des multiplicateurs de Lagrange de ces contraintes :

$$u = (u_1, \dots, u_f, \dots, u_{|F|})$$

La difficulté du problème tient essentiellement à la prise en compte des contraintes (3.2). En effet, sans elles, le problème se décompose en R sous-problèmes indépendants, appartenant à la classe des problèmes de Transbordement.

Soit $L(x, u)$ le Lagrangien de (P0), au point x , associé à la relaxation des contraintes (3.2) :

$$L(x, u) = \sum_{\{(ij) \in E\}} c_{ij} \cdot x_{ij} + \sum_{f \in F} \left[a_f - \sum_{\{j/j \in F\}} x_{fj} + \sum_{\{j/(j \in F)\}} x_{jf} \right] \cdot u_f$$

$$L(x, u) = \sum_{r \in R^+} \left[\sum_{\{(ij) \in E_r\}} c_{ij} \cdot x_{ij} + \sum_{\{(if)/(i \in F, f \in F)\}} x_{if} \cdot u_f \right] + \sum_{r \in R^-} \left[\sum_{\{(ij) \in E_r\}} c_{ij} \cdot x_{ij} - \sum_{\{(fj)/(j \in F, f \in F)\}} x_{fj} \cdot u_f \right] + \sum_{f \in F} a_f \cdot u_f$$

On définit la fonction duale $h(u)$, au point u , en minimisant $L(x, u)$, pour tout x .

$$h(u) = \min_x \{ L(x, u) \}$$

$$h(u) = \sum_{r \in R^+} h_r^+(u) + \sum_{r \in R^-} h_r^-(u) + \sum_{f \in F} a_f \cdot u_f$$

Soit (P1) le problème, dit programme maître, dual de (P0) :

$$\text{Maximiser } h(u) \\ u \geq 0$$

et (P1r) les sous-problèmes associés :

$$\left\{ \begin{array}{l} h_r^+(u) = \min_x \sum_{(ij) \in E_r} c_{ij} \cdot x_{ij} + \sum_{\{(if)/f \in F\}} u_f \cdot x_{if} \\ \text{s.c.} \\ \sum_{\{j/(ij) \in E_r\}} x_{ij} - \sum_{\{j/(ji) \in E_r\}} x_{ji} = a_i \quad \forall i \in N_r \\ 0 \leq x_{ij} \leq k_{ij} \quad \forall (ij) \in E_r \end{array} \right.$$

et

$$\left\{ \begin{array}{l} h_r^-(u) = \min_x \sum_{(ij) \in E_r} c_{ij} \cdot x_{ij} - \sum_{\{(ff)/f \in F\}} u_f \cdot x_{ff} \\ \text{s.c.} \\ \sum_{\{j/(ij) \in E_r\}} x_{ij} - \sum_{\{j/(ji) \in E_r\}} x_{ji} = a_i \quad \forall i \in N_r \\ 0 \leq x_{ij} \leq k_{ij} \quad \forall (ij) \in E_r \end{array} \right.$$

La résolution du problème Dual se ramène à la recherche du maximum d'une fonction concave non partout différentiable.

Le programme dual est plus facile à résoudre que le primal. Sa solution (exacte ou approchée) fournit des bornes inférieures de $f(x^*)$. La résolution du dual donne généralement comme résultats intermédiaires des solutions admissibles du primal, souvent proches de l'optimum, quelquefois optimales. La connaissance d'une solution primale-admissible et d'une solution duale-admissible permet d'encadrer $f(x^*)$ et, par conséquent, d'interrompre les calculs si l'on se trouve dans les limites d'une tolérance fixée à l'avance.

Propriété :

Si (P1) a une solution optimale x^* , alors il existe un point-selle (x^*, u^*) et pour tout x et $u \geq 0$, on a :

$$h(u) \leq h(u^*) = f(x^*) \leq f(x)$$

Le coordonnateur doit donc rechercher le point-selle de la fonction h .

III.2.2 - Algorithme de sous-gradient

$g = (g_1, g_2, \dots, g_{|F|}) \in \mathbb{R}^{|F|}$ est un sous-gradient de la fonction concave h au point u^0 si et seulement si :

$$h(u) \leq h(u^0) + g^T \cdot (u - u^0)$$

On a en particulier, pour $f \in F$:

$$g_f = a_f + \sum_{\{j/j \in F\}} x_{jf} - \sum_{\{j/j \notin F\}} x_{jf}$$

où x_{jf} est solution du sous-problème (P2r), avec $r \in \mathbb{R}^+$ et x_{jf} est solution du sous-problème (P2r), avec $r \in \mathbb{R}^-$.

La fonction h à maximiser n'étant pas continûment différentiable, nous utilisons un algorithme de sous-gradient avec contraintes, car on sait facilement déterminer au moins un sous-gradient en chaque point de h .

A chaque itération t de l'algorithme, on choisit un ensemble de noeuds F , pour lequel on change le vecteur des prix u^t selon une certaine direction, celle du sous-gradient g^t . g^t appartient au sous-différentiel de h en u^t , $\partial h(u^t)$. Cette direction n'est pas nécessairement une direction de montée pour h , mais c'est une direction de descente pour la fonction distance à l'optimum.

L'algorithme s'arrête lorsque les résultats ont atteint un certain seuil de tolérance défini à l'avance. En pratique, la condition d'arrêt portent sur le niveau de précision des résultats, $|| u^{t+1} - u^t || < \tau$, et sur le nombre maximum d'itération, $t = t_{\max}$.

Soit $\text{Proj}_{/\Omega}$, l'opérateur de projection orthogonal sur l'ensemble $\Omega = \{u/u \geq 0\}$ convexe, tel que $\bar{u}^t = \text{Proj}_{/\Omega}(u^t)$ et :

$$\begin{aligned} \bar{u}_f^t &= u_f^t & \forall u_f^t &\geq 0 \\ \bar{u}_f^t &= 0 & \forall u_f^t &< 0 \end{aligned}$$

Algorithme 4 :

1 - A l'étape 0, initialiser :

$$t = 0, u^t, p^t$$

2 - A l'étape $t+1$, calculer :

$$u^{t+1} = \text{Proj}_\Omega(u^t + p^t g^t.)$$

$$\text{où } g^t = \partial h(u^t)$$

3 -Test d'arrêt :

Si ($\| u^{t+1} - u^t \| < \tau$ ou $t = t_{\max}$),

alors FIN.

Sinon $t = t+1$ et aller en 2.

Dans l'exemple qui nous intéresse, la calcul du sous-gradient correspond à un calcul matriciel relativement simple et extrêmement rapide à effectuer sur ordinateur. La principale difficulté est plutôt d'ordre théorique : comment choisir les déplacements p^t de façon à obtenir une convergence satisfaisante?

Conditions de convergence

La convergence de cette algorithme dépend essentiellement de la manière dont les pas de déplacement p^t sont choisis à chaque itération t . On peut extraire (cf. [Pol,1966]) de la suite $\{u^t\}$ une sous-suite convergeant vers une solution optimale du problème, sous les seules conditions :

$$\lim_{t \rightarrow \infty} p^t = 0$$

$$\sum_{t=0}^{\infty} p^t = \infty$$

Seulement, ces conditions théoriques entraînent une convergence trop lente de l'algorithme. En pratique, Shor [Sho,1968] propose une série convergente où p^0 dépend d'une estimation de la distance initiale à l'optimum et ρ^t est un coefficient strictement plus petit que 1 (cf. [Min,1983]):

$$p^t = p^0 \cdot \rho^t$$

III.3 - Décomposition primale

Pour mettre en place une procédure d'allocation spatiale de ressources, nous proposons une technique primale de décomposition par les ressources. L'avantage de ce type d'approche revient au coordonnateur pour qui les transferts d'informations sont relativement simples. L'inconvénient est le taux de convergence parfois très faible, générant un très grand nombre d'itérations.

Le principe est de satisfaire, pour chaque sous-problème (ou zone), les contraintes portant sur les ressources et les demandes. Pour cela, on utilise l'algorithme de transbordement, qui fournit les valeurs des variables duales de la solution optimale. Ces variables duales servent pour le test dual, portant sur les contraintes de couplage. Au niveau de coordination, on satisfait les contraintes de couplage à chaque itération, et on approche l'optimum itérativement.

III.3.1 - Variables de couplage

La première étape de cette approche primale est de manipuler le problème (P0) pour obtenir le problème séparable (P2). Pour cela, nous définissons les vecteurs des variables de couplage y^+ et y^- , représentant les flux inter-zones par l'intermédiaire des points frontières :

$$\begin{aligned} y^+ &= (y_1^+, \dots, y_f^+, \dots, y_{|F|}^+) \in \mathbb{R}^{|F|} \\ y^- &= (y_1^-, \dots, y_f^-, \dots, y_{|F|}^-) \in \mathbb{R}^{|F|} \end{aligned}$$

Nous transformons l'équation (1) en :

$$\sum_{\{j/(ij) \in E_r\}} x_{ij} - \sum_{\{j/(ji) \in E_r\}} x_{ji} = a_i \quad \forall i \in F, \forall r \in R \quad (3.3)$$

$$\sum_{\{i/i \in N_r^+\}} x_{if} = y_f^+$$

$$\sum_{\{i/i \in N_r^-\}} x_{fi} = y_f^- \quad \forall f \in F \quad (3.4)$$

$$y_f^+ - y_f^- = a_f$$

La contrainte (3.4) est la contrainte de couplage du problème et nous définissons le vecteur u des multiplicateurs de Lagrange de ces contraintes. On note :

$$u^+ = (u_1^+, \dots, u_f^+, \dots, u_{|F|}^+) \in \mathbb{R}^{|F|}$$

$$u^- = (u_1^-, \dots, u_f^-, \dots, u_{|F|}^-) \in \mathbb{R}^{|F|}$$

Pour une affectation y donnée, chaque sous-système r doit choisir son niveau d'activité $v_r(y)$ comme solution optimale du sous-problème.

Soit (P2) le problème maître, faiblement séparable et équivalent à (P0) :

$$\left\{ \begin{array}{l} \text{Minimiser}_y V(y) = \sum_{r \in R^+} v_r^+(y^+) + \sum_{r \in R^-} v_r^-(y^-) \\ \text{s.c.} \\ y_f^- - y_f^+ = a_f \quad \forall f \in F \end{array} \right.$$

et (S2r) les sous-problèmes associés :

$$\left\{ \begin{array}{l} v_r^+(y^+) = \text{Minimiser} \sum_{(ij) \in E_r^+} c_{ij} \cdot x_{ij} \\ \text{s.c.} \\ \sum_{\{i/i \in N_r^+\}} x_{if} = y_f^+ \quad \forall f \in F \\ \sum_{(ij) \in E_r^+} x_{ij} - \sum_{(ji) \in E_r^+} x_{ji} = a_i \quad \forall i \in N_r^+ \\ 0 \leq x_{ij} \leq k_{ij} \quad \forall (ij) \in E_r^+ \end{array} \right.$$

$$\left\{ \begin{array}{l} v_r^-(y^-) = \text{Minimiser} \sum_{(ij) \in E_r^-} c_{ij} \cdot x_{ij} \\ \text{s.c.} \\ \sum_{\{i/i \in N_r^-\}} x_{fi} = y_f^- \quad \forall f \in F \\ \sum_{(ij) \in E_r^-} x_{ij} - \sum_{(ji) \in E_r^-} x_{ji} = a_i \quad \forall i \in N_r^- \\ 0 \leq x_{ij} \leq k_{ij} \quad \forall (ij) \in E_r^- \end{array} \right.$$

La fonction V est convexe, mais généralement non différentiable.

III.3.2 - Algorithme de sous-gradient

$g = (g_1, g_2, \dots, g_{|F|}) \in \mathbb{R}^{2|F|}$ est un sous-gradient de la fonction convexe V au point y^0 si et seulement si :

$$V(y) \geq V(y^0) + g^T \cdot (y - y^0)$$

On a en particulier, pour $f \in F$:

$$g_f = \begin{pmatrix} u_f^+ \\ -u_f^- \end{pmatrix}$$

La fonction V à minimiser n'étant pas continûment différentiable, nous utilisons un algorithme de sous-gradient avec contraintes, car on sait facilement déterminer au moins un sous-gradient en chaque point de V .

A chaque itération t de l'algorithme, on choisit un ensemble de noeuds F , pour lequel on change le vecteur des flots inter-zone y^t selon une certaine direction, celle du sous-gradient g^t . En théorie, seule la connaissance complète du sous-différentiel en y ($y \neq y^*$) permettrait de définir une direction de déplacement descendante. g^t appartient au sous-différentiel de V en y^t , $\partial V(y^t)$. En pratique, nous définissons une direction de descente pour la fonction distance à l'optimum.

L'algorithme s'arrête lorsque les résultats ont atteint un certain seuil de tolérance défini à l'avance. En pratique, la condition d'arrêt portent sur le niveau de précision des résultats, $\|y^{t+1} - y^t\| < \tau$, et sur le nombre maximum d'itération, $t = t_{\max}$

Soit Proj_Ω , l'opérateur de projection orthogonale sur l'ensemble convexe Ω , défini par:

$$\Omega = \{y / y_f^+ - y_f^- = a_f; f \in F\}$$

Soit, après une translation de a_f :

$$\Omega_0 = \{y / y_f^+ - y_f^- = 0; f \in F\}$$

et

si $a_f > 0$

$$\begin{cases} \sum_{i \in N_r^+} x_{if} = y_f^+ \\ \sum_{i \in N_r^-} x_{fi} = y_f^- + a_f \\ y_r^- - y_r^+ = 0 \end{cases}$$

si $a_f < 0$

$$\begin{cases} \sum_{i \in N_r^+} x_{if} = y_f^+ - a_f \\ \sum_{i \in N_r^-} x_{fi} = y_f^- \\ y_r^- - y_r^+ = 0 \end{cases}$$

Le sous-gradient est alors le suivant :

$$\forall f \in F, \bar{g}_f^t = \text{Proj}_{/\Omega_0} (g_f^t) \text{ soit,}$$

$$\forall f \in F, \bar{g}_f^t = \text{Proj}_{/\Omega_0} \begin{pmatrix} u_f^+ \\ u_f^- \end{pmatrix}$$

$$\forall f \in F, \bar{g}_f^t = \frac{1}{2} \begin{pmatrix} u_f^+ - u_f^- \\ u_f^- - u_f^+ \end{pmatrix}$$

d'où l'algorithme primal de sous-gradient :

Algorithme 5 :

1 - A l'étape 0, initialiser :

$$t = 0, y^t, p^t$$

2 - A l'étape t+1, calculer :

$$y^{t+1} = \text{Proj}_{/\Omega} (y^t - p^t g^t.)$$

$$\text{où } g^t \in \partial V(y^t)$$

3 -Test d'arrêt :

$$\text{Si } (|| y^{t+1} - y^t || < \tau \text{ ou } t = t_{\max}),$$

alors FIN.

Sinon $t = t+1$ et aller en 2.

Conditions de convergence

Pour assurer une convergence assez rapide de cet algorithme nous proposons, comme précédemment pour l'algorithme 4, la méthode dite de la "série convergente" du type :

$$p^t = p^0 \cdot \rho^t$$

où p^0 dépend d'une estimation de la distance initiale à l'optimum et ρ^t est un coefficient strictement plus petit que 1 (cf. [Min,1983])

III.4 - Décomposition mixte

On va décrire ici une approche qui combine des allocations de prix et de ressources en visant la décentralisation des sous-problèmes. En effet, comme nous l'avons déjà observé dans les deux approches précédentes, la non unicité des solutions locales issue de la non différentiabilité des fonctions de coordination, ne permet pas aux décideurs locaux de calculer eux mêmes leur décision optimale globale. On dit alors que le schéma de décomposition n'est pas totalement décentralisé. Le niveau de coordination doit calculer une combinaison convexe des solutions locales, qui satisfait les contraintes de couplage. La nécessité de cette coordination, appelée programme maître dans la méthode de Dantzig-Wolfe, alourdit les calculs et pénalise la parallélisation potentielle à la décomposition.

L'analyse de la dégénérescence des sous-problèmes conduit à une allocation primale et duale qui permet d'éliminer le niveau de coordination. L'approche présentée ici à titre expérimental, s'inspire d'un algorithme de décomposition mixte décrit dans [Mah,1986].

III.4.1 - Dégénérescence des sous-problèmes

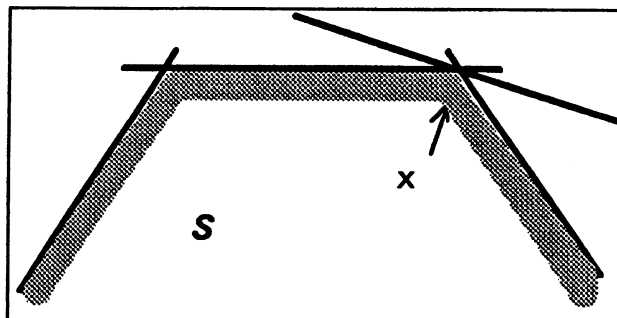
Rappelons brièvement comment caractériser la dégénérescence d'un problème de Programmation Linéaire.

Dégénérescence primale :

Une solution de base d'un programme linéaire comportant n variables et m contraintes (sous la forme standard) est dite primale dégénérée si le nombre de variables de base, strictement positives, est inférieur à m . La solution duale n'est alors pas unique.

Beaucoup de modèles sont naturellement dégénérés, mais nous nous intéressons ici à la dégénérescence introduite par la relaxation du problème en sous-problèmes.

Exemple :



- Exemple de dégénérescence dans $\mathbb{R}^2$ -
x est un point extrême correspondant à une solution
de base dégénérée.

Dégénérescence duale :

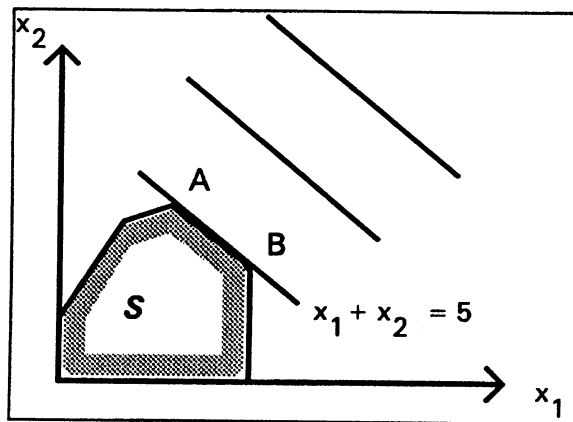
La solution optimale d'un programme linéaire est dite duale dégénérée, s'il existe des coûts réduits de variables hors base, à zéro. La solution primale n'est alors pas unique.

Exemple :

Soit (P) le problème :

$$\left\{ \begin{array}{l} \text{Minimiser } z = x_1 + x_2 \\ -3x_1 + 2x_2 \leq 2 \\ -x_1 + 2x_2 \leq 4 \\ x_1 + x_2 \leq 5 \\ x_1 \leq 3 \\ x_1 \geq 0 \\ x_2 \geq 0 \end{array} \right.$$

L'ensemble S, des solutions réalisables de (P), est représenté sur le schéma ci-dessous:



On se propose de déterminer la valeur de z^* optimale. Les courbes de niveau de la fonction z sont une famille de droites, d'équation générique :

$$z^* = x_1 + x_2 = \lambda = \text{cste.}$$

Dans notre exemple $z^*=5$. Or la droite d'équation $x_1 + x_2 = 5$ étant parallèle à la face du polygone des contraintes, défini par les deux points A et B, il en résulte que tous les points du segment [A,B] sont des solutions optimales de (P). La solution du problème (P) est dite dégénérée.

Reprenons maintenant le problème de transbordement (P0), avec des hypothèses d'unicité et de non dégénérescence des solutions optimales primales et duales. On note m son nombre de contraintes, m_f son nombre de contraintes couplantes et m_r son nombre de contraintes relatives aux sommets de la zone r différents des points frontières.

Par hypothèse, la solution optimale de (P0) contient exactement m variables de base strictement positives ($0 < x_{ij} < k_{ij}$), la solution optimale du problème (P1) contient exactement $m1$ variables de base strictement positives et la solution optimale du problème (P2) contient exactement $m2$ variables de base strictement positives.

Par l'approche Duale, on traite les sous-problèmes (P1r) suivants :

$$\left\{ \begin{array}{l} h_r^+(u) = \min_x \sum_{(ij) \in E_r} c_{ij} \cdot x_{ij} + \sum_{\{(if)/f \in F\}} u_f \cdot x_{if} \\ \text{s.c.} \\ \sum_{\{j/(ij) \in E_r\}} x_{ij} - \sum_{\{j/(ji) \in E_r\}} x_{ji} = a_i \quad \forall i \in N_r \\ 0 \leq x_{ij} \leq k_{ij} \quad \forall (ij) \in E_r \end{array} \right.$$

$$\left\{ \begin{array}{l} h_r^-(u) = \min_x \sum_{(ij) \in E_r} c_{ij} \cdot x_{ij} - \sum_{\{(if)/f \in F\}} u_f \cdot x_{if} \\ \text{s.c.} \\ \sum_{\{j/(ij) \in E_r\}} x_{ij} - \sum_{\{j/(ji) \in E_r\}} x_{ji} = a_i \quad \forall i \in N_r \\ 0 \leq x_{ij} \leq k_{ij} \quad \forall (ij) \in E_r \end{array} \right.$$

Par l'approche Primale on traite les sous-problèmes (P2r) suivants :

$$\left\{ \begin{array}{l} v_r^+(y^+) = \text{Minimiser} \sum_{(ij) \in E_r^+} c_{ij} \cdot x_{ij} \\ \text{s.c.} \\ \sum_{\{i/i \in N_r^+\}} x_{if} = y_f^+ \quad \forall f \in F \\ \sum_{(ij) \in E_r^+} x_{ij} - \sum_{(ji) \in E_r^+} x_{ji} = a_i \quad \forall i \in N_r^+ \\ 0 \leq x_{ij} \leq k_{ij} \quad \forall (ij) \in E_r^+ \end{array} \right.$$

$$\left\{ \begin{array}{l} v_r^-(y^-) = \text{Minimiser} \sum_{(ij) \in E_r^-} c_{ij} \cdot x_{ij} \\ \text{s.c.} \\ \sum_{\{i/i \in N_r^-\}} x_{if} = y_f^- \quad \forall f \in F \\ \sum_{(ij) \in E_r^-} x_{ij} - \sum_{(ji) \in E_r^-} x_{ji} = a_i \quad \forall i \in N_r^- \\ 0 \leq x_{ij} \leq k_{ij} \quad \forall (ij) \in E_r^- \end{array} \right.$$

Il en résulte que :

$$\begin{aligned}
m &= m_f + \sum_{r \in R} m_r \\
m_1 &= \sum_{r \in R} m_r = m - m_f \\
m_2 &= \sum_{r \in R} (m_r + m_f) = \sum_{r \in R} m_r + R \cdot m_f = m + (R - 1) \cdot m_f
\end{aligned}$$

Par conséquent, les solutions de base optimales des sous-problèmes associés aux décompositions primales et duales sont telles que :

la *méthode primale* affecte chaque contrainte de couplage dans chaque sous-problème. Ainsi $(R-1)m_f$ variables, en base et optimales, sont nulles. Par conséquent, au moins $(R-1)$ sous-problèmes sont dégénérés.

la *méthode duale* relaxe toutes les contraintes de couplage, en affectant un prix dans les fonctions objectifs des sous-problèmes. Le nombre total de contraintes a donc diminué de m_f . La solution optimale ne peut être formée par les solutions de base des sous-problèmes. Donc au moins un sous-problème possède des solutions optimales non uniques, c'est à dire que le dual est dégénéré.

Pour éliminer ces dégénérescences et ainsi garantir la décentralisation, on doit affecter chaque contrainte à exactement un sous-problème. Cette allocation peut être réalisée progressivement par une méthode de sous-gradient, ou bien déterminée directement par une méthode appropriée. Pour notre problème, une telle méthode d'allocation s'est avérée trop difficile à définir de manière rigoureuse. Sa conception repose sur l'analyse de la méthode duale et sur les conditions de dégénérescence. Nous nous sommes finalement limités à l'étude de différents cas de figure, pour lesquelles la méthode d'allocation est très efficace, et avons défini les deux propositions suivantes.

Proposition 3 :

Si une zone ne possède qu'un point frontière avec l'extérieur, alors elle n'est pas dégénérée.

Démonstration :

Nous avons :

$$|F| = q = q_1 = q_2 = 1 \quad \text{et} \quad |R| = 2$$

où q_r est le nombre de points frontières de la zone r ($r=1,2$). Nous savons, par hypothèse, que le problème global est non dégénéré et équilibré. Le nombre d'arcs en base est égal à :

$$n_1 + n_2 + 1 - 1 = n_1 + n_2$$

où n_r est le nombre de sommets du graphe de la zone r ($r=1,2$)

De même, le nombre d'arcs en base et non fictifs, pour les 2 sous-problèmes, s'élève à:

$$(n_1 + q_1 + 1 - 1) - q_1 + (n_2 + q_2 + 1 - 1) - q_2 = n_1 + n_2$$

Dans cette formulation, q_r ($r=1,2$) est le nombre de points frontières ainsi que le nombre d'arcs fictifs pointant sur la balance, dans le sous-problème r . Ainsi, quelque soit q_r , le nombre d'arcs non fictifs en base dans le sous-problème r , est égal à n_r . En déduit que les sous-problèmes ne sont pas dégénérés, car le nombre d'arcs en base est identique. La balance revient à satisfaire le couplage.

Remarque :

Si $q = 2$, alors l'arc supplémentaire de l'une des bases optimales locales, est un arc fictif incident à $f \in F$.

En effet, il existe une combinaison des solutions de bases locales qui satisfait le bilan au noeud f , alors que pour les autres sommets, le bilan est satisfait localement. Si l'arc supplémentaire appartient à la zone excédentaire, alors celle-ci est dégénérée, sinon la zone déficitaire est dégénérée.

Proposition 4 :

Si une zone possède ($q > 1$) points frontières avec ses ($|R|-1$) zones voisines, alors au moins une des $|R|$ zones est dégénérée et la dégénérescence totale est égale à ($q-1$).

Démonstration :

Nous avons :

$$|F| = q > 1 \quad |R| = p > 1$$

Nous savons, par hypothèse, que le problème global est non dégénéré et équilibré. Le nombre d'arcs en base est égal à :

$$(n_1 + n_2 + \dots + n_p) + q - 1$$

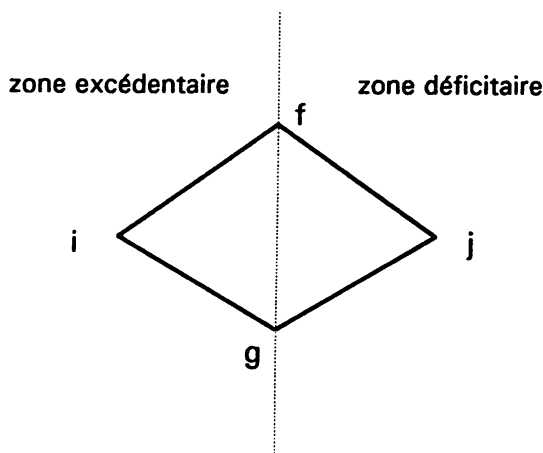
De même, le nombre d'arcs non fictifs en base, dans les $|R|$ sous-problèmes, est égal à:

$$(n_1 + q_1 + 1 - 1) - q_1 + \dots + (n_p + q_p + 1 - 1) - q_p = (n_1 + \dots + n_p)$$

Si $q > 1$, on en déduit qu'au moins un des sous-problèmes est dégénéré et le niveau de dégénérescence est égal à $(q-1)$.

Exemple :

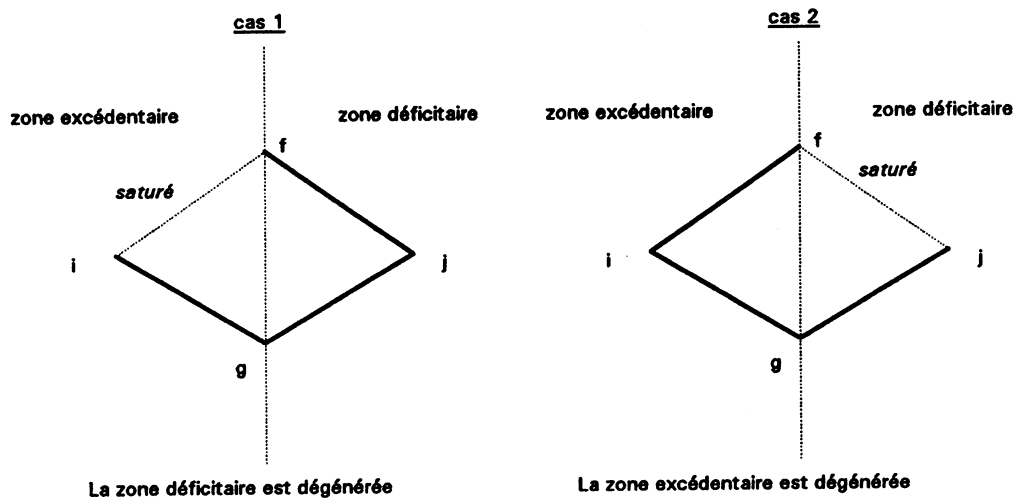
Considérons le graphe suivant, où f et g sont les points frontières entre les zones i et j .



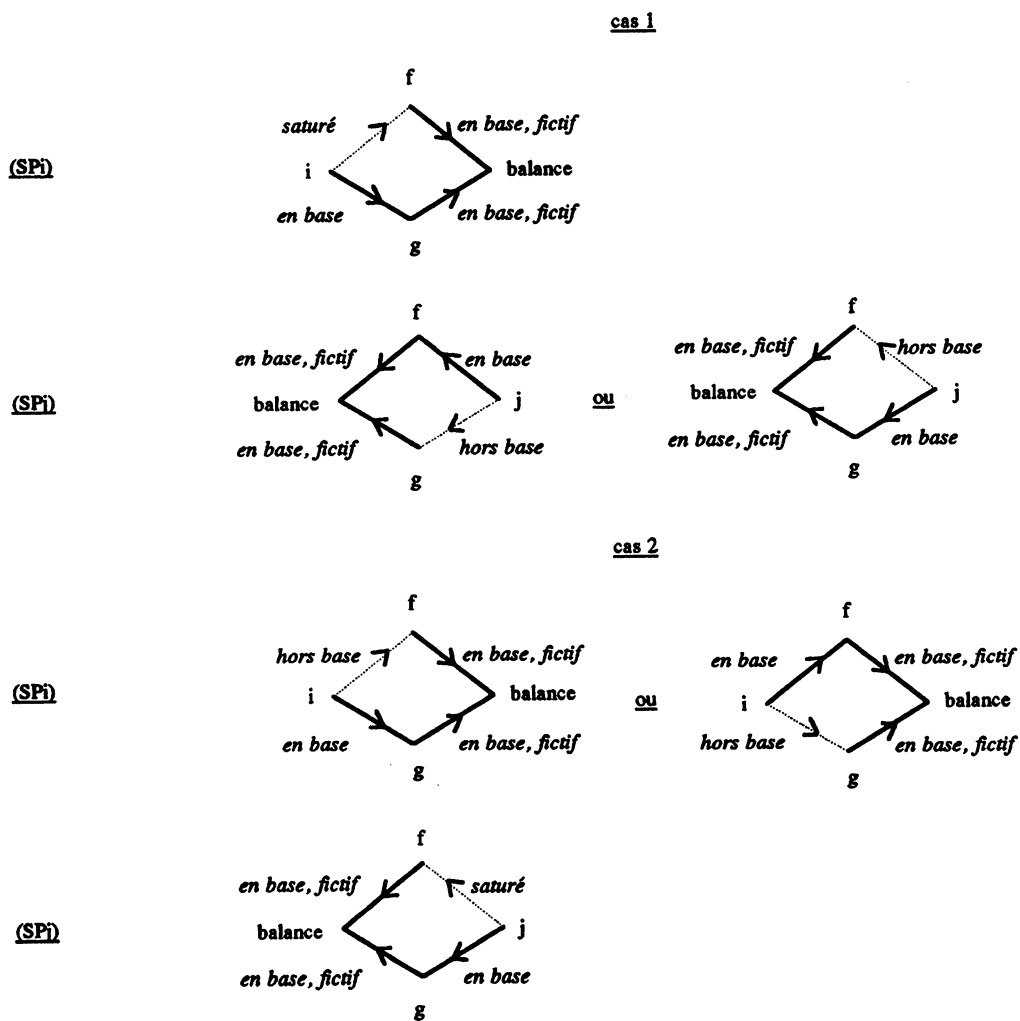
Nous supposons que le problème global n'est pas dégénéré. Comme ce graphe a deux sommets et deux points frontières, la solution optimale aura trois arcs en base et un arc saturé hors base. Dans ces conditions, la zone dégénérée est relativement simple à déterminer. Etudions, par exemple, les deux configurations suivantes et les sous-problèmes associés, (SP_i) et (SP_j) :

Cas 1 : si au noeud frontière f , les capacités saturent un arc de la zone excédentaire, alors celle-ci n'est pas dégénérée. Par contre, la zone déficitaire est dégénérée. En effet, comme le montre le graphe du sous-problème, la solution n'est pas unique car le flot peut passer indifféremment sur l'arc (fj) ou (gj) .

Cas 2 : dans le cas contraire, où les capacités saturent un arc de la zone déficitaire, c'est la zone excédentaire qui est dégénérée.



Les graphes des sous-problèmes (SPi) et (SPj) sont alors :



III.4.2 - Algorithme de décomposition mixte

Reprenons les contraintes de couplage aux noeuds frontières :

$$\sum_{\{j/j \notin F\}} x_{fj} - \sum_{\{j/j \in F\}} x_{jf} = a_f \quad \forall f \in F$$

Puis reprenons les résultats de l'approche primale, pour définir les contraintes suivantes:

si $a_f > 0$:

$$\begin{cases} \sum_{i \in N_r^+} x_{if} = y_f \\ \sum_{i \in N_r^-} x_{fi} = y_f + a_f \end{cases}$$

si $a_f < 0$:

$$\begin{cases} \sum_{i \in N_r^+} x_{if} = y_f - a_f \\ \sum_{i \in N_r^-} x_{fi} = y_f \end{cases}$$

Exemple :

Remarque préliminaire : pour simplifier les écritures, nous ne donnons ci dessous que les termes liés aux points frontières.

Prenons le cas où $a_f > 0$ et supposons que la zone r^+ est dégénérée. Dans ces conditions, il faut garder explicitement la contrainte $\sum_{i \in N_r^+} x_{if} = y_f$ dans $(P3r^+)$ et dualiser la contrainte $\sum_{i \in N_r^-} x_{fi} - y_f - a_f = 0$ dans $(P3r^-)$. Les sous-problèmes s'écrivent

alors comme ceci:

$$\begin{aligned}
(P3r^+) = \left\{ \begin{array}{l} k_f^+(\bar{u}_f, \bar{y}_f) = \underset{x}{\text{Minimiser}} \sum_{(ij) \in E_r^+} c_{ij} \cdot x_{ij} \\ \text{s.c.} \\ \sum_{\{i/i \in N_r^+\}} x_{if} = \bar{y}_f \quad \forall f \in F \\ \sum_{(ij) \in E_r^+} x_{ij} - \sum_{(ji) \in E_r^+} x_{ji} = a_i \quad \forall i \in N_r^+, i \notin F \\ 0 \leq x_{ij} \leq k_{ij} \quad \forall (ij) \in N_r^+ \end{array} \right. \\
(P3r^-) = \left\{ \begin{array}{l} k_f^-(\bar{u}_f, \bar{y}_f) = \underset{x}{\text{Minimiser}} \sum_{(ij) \in E_r^-} c_{ij} \cdot x_{ij} - \sum_{\{(fj)/f \in F\}} \bar{u}_f \cdot x_{fj} \\ \text{s.c.} \\ \sum_{(ij) \in E_r^-} x_{ij} - \sum_{(ji) \in E_r^-} x_{ji} = a_i \quad \forall i \in N_r^-, i \notin F \\ 0 \leq x_{ij} \leq k_{ij} \quad \forall (ij) \in N_r^- \end{array} \right.
\end{aligned}$$

Nous avons alors :

$$\begin{aligned}
k_f(\bar{u}_f, \bar{y}_f) &= k_f^+(\bar{u}_f, \bar{y}_f) + k_f^-(\bar{u}_f, \bar{y}_f) + u_f \cdot (a_f + y_f) \\
K(u, y) &= \sum_{f \in F} k_f(\bar{u}_f, \bar{y}_f)
\end{aligned}$$

Pour terminer, reprenons les résultats de l'approche duale pour définir le Lagrangien:

Si la zone $r \in R^+$ est dégénérée :

$$\begin{aligned}
L(x, u, y) &= \sum_{\{(ij) \in E\}} c_{ij} \cdot x_{ij} + \sum_{f \in F} \left[a_f - \sum_{\{j/j \in F\}} x_{fj} + y_f \right] \cdot u_f \\
L(x, u, y) &= \sum_{r \in R^+} \left[\sum_{\{(ij) \in E_r\}} c_{ij} \cdot x_{ij} \right] + \sum_{r \in R^+} \left[\sum_{\{(ij) \in E_r\}} c_{ij} \cdot x_{ij} - \sum_{\{(fj)/j \in F, f \in F\}} x_{fj} \cdot u_f \right] + \sum_{f \in F} (a_f + y_f) \cdot u_f
\end{aligned}$$

Si la zone $r \in R^-$ est dégénérée :

$$\begin{aligned}
L(x, u, y) &= \sum_{\{(ij) \in E\}} c_{ij} \cdot x_{ij} + \sum_{f \in F} \left[a_f - y_f + \sum_{\{j/j \in F\}} x_{jf} \right] \cdot u_f \\
L(x, u, y) &= \sum_{r \in R^+} \left[\sum_{\{(ij) \in E_r\}} c_{ij} \cdot x_{ij} + \sum_{\{(jf)/j \in F, f \in F\}} x_{jf} \cdot u_f \right] + \sum_{r \in R^-} \left[\sum_{\{(ij) \in E_r\}} c_{ij} \cdot x_{ij} \right] + \sum_{f \in F} (a_f - y_f) \cdot u_f
\end{aligned}$$

On minimise le Lagrangien au point (u,y), associé à la relaxation des contraintes de couplage et à l'ajout des variables de couplage.

$$k(u,y) = \min_x \{ L(x,u,y) \}$$

Le calcul de la fonction k se décompose en la résolution de R sous-problèmes (P3r).

Pour une affectation $(\bar{y}, \bar{u})$ donnée, chaque sous-système r doit choisir son niveau d'activité comme solution optimale du sous-problème (P3r).

Si (P0) a une solution optimale x^* , alors il existe un point-selle et pour tout u et y :

$$k(u,y^*) \leq k(u^*,y^*) = f(x^*) \leq k(u^*,y)$$

Le coordonnateur doit donc rechercher le point-selle de la fonction h. La fonction $k(u,y)$ est une fonction convexe en y et concave en u, seulement elle n'est, en général, pas différentiable (notamment à l'optimum). A l'intérieur du domaine de k, on peut définir le sous-différentiel de k en (u,y) :

$$k(u,y) = \partial_u k(u,y) \cdot \partial_y k(u,y)$$

Les sous-problèmes s'échangent les paramètres sans l'aide d'un niveau supérieur de coordination. Chaque sous-problème est décentralisé au voisinage de l'optimum. On montre que chaque sous-problème, auquel ont été affectés les paramètres de coordination optimaux, possède une unique paire de solutions primales-duales optimales, également optimales pour le problème de départ (chaque contrainte, non affectée dans un sous-problème, intervient dans sa fonction objectif par l'intermédiaire d'un prix).

Dans notre exemple, deux zones voisines communiquent par l'intermédiaire d'un ou plusieurs points frontières si l'une est déficitaire et l'autre est excédentaire.

Algorithme 6 :

1 - A l'étape 0, initialiser :

$$t = 0, y^t, u^t, p^t$$

2 - A l'étape $t+1$:

$$\forall \{r, s\} / r \in R, s \in R$$

Si $r \in R^+$, et r est dégénérée :

résoudre (P3r) avec y_f^t

résoudre (P3s) avec u_f^t

$$y_f^{t+1} = \sum_{\{i/i \in N_r^-\}} x_{fi}$$

$$u_f^{t+1} = u_f^+$$

Si $s \in R^-$, et s est dégénérée :

résoudre (P3s) avec y_f^t

résoudre (P3r) avec u_f^t

$$y_f^{t+1} = \sum_{\{i/i \in N_s^+\}} x_{if}$$

$$u_f^{t+1} = u_f^-$$

3 -Test d'arrêt :

Si $(|| y^{t+1} - y^t || < \tau \text{ ou } t = t_{\max})$,

alors FIN.

Sinon aller en 2.

Convergence de l'algorithme

Proposition :

Si $|R| = 3$, l'algorithme converge en un nombre fini d'itérations.

Démonstration :

On considère deux zones excédentaires (s et r) et une zone déficitaire (p).

Soient f le point frontière entre s et p , g le point frontière entre r et p et h le point frontière entre s et r (inactif).

Soit x la solution optimale de (P0) et u les variables duales optimales. Soit $\bar{c}_N$, le coût réduit des variables hors base, tel que $\bar{c}_N \geq 0$. Soit $\bar{c}_B$, le coût réduit des variables en base, tel que $\bar{c}_B = 0$

Soit a_s , la quantité excédentaire de s , telle que :

$$\sum_{\{i/i \in N_s^+\}} x_{if} = a_s$$

Soit a_r , la quantité excédentaire de r , telle que :

$$\sum_{\{i/i \in N_r^+\}} x_{ig} = a_r$$

Soit a_p , la quantité déficitaire de p , telle que :

$$a_p = a_s + a_r$$

Soit I l'ensemble des sommets demandes de la zone p , adjacents au sommet f .

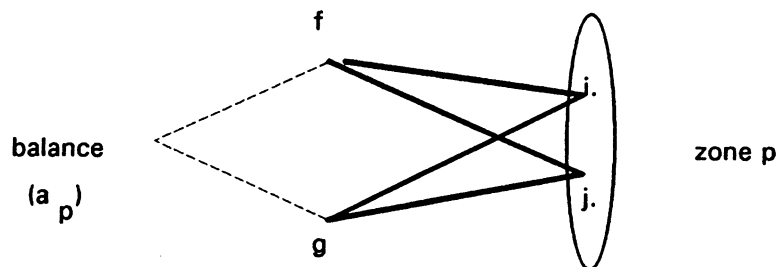
Soit J l'ensemble des sommets demandes de la zone p , adjacents au sommet g .

Nous faisons l'hypothèse, qu'à l'optimum :

$$\begin{aligned} \forall i \in I, \text{ les arcs } (fi) \text{ sont en base et les arcs } (gi) \text{ sont hors base. } \sum_{i \in I} a_i &= a_s \\ \forall j \in J, \text{ les arcs } (gj) \text{ sont en base et les arcs } (fj) \text{ sont hors base. } \sum_{j \in J} a_j &= a_r \end{aligned}$$

Etape $t=1$:

Traitement de la zone p par l'approche primale-duale :



Faisons l'hypothèse suivante :

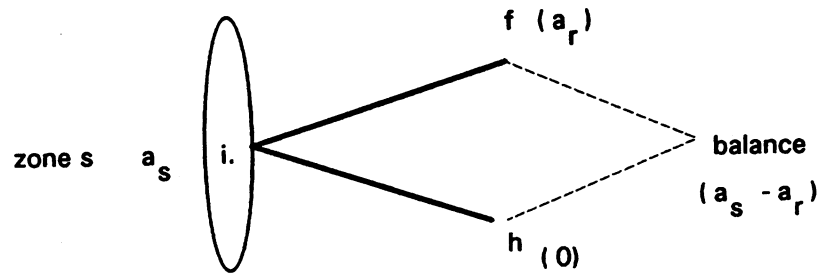
$$\sum_{i \in I} x_{gi} = a_s \quad \text{et} \quad \sum_{j \in J} x_{fj} = a_r$$

Donc :

$$y_f^{t+1} = a_r \quad \text{et} \quad y_g^{t+1} = a_s \quad \text{et} \quad y_h^{t+1} = 0 \quad \text{avec} \quad (t+1 = 2)$$

Etape $t=2$:

Traitement des zones s et r par l'approche primale-duale :



$$\sum_{\{i/i \in N_i^+\}} x_{if} = \text{Min}\{a_s, a_r\} = \alpha$$

$$\sum_{\{i/i \in N_i^+\}} x_{ig} = \text{Min}\{a_s, a_r\} = \alpha$$

Etape $t = 3$:

Traitement de la zone p. On a :

$$\forall i \in I, c_{fi} < c_{gi}, \bar{c}_{fi} = 0 \Rightarrow c_{fi} - u_f < c_{gi} - u_g$$

$$\forall j \in J, c_{gj} < c_{fj}, \bar{c}_{gj} = 0 \Rightarrow c_{gj} - u_g < c_{fj} - u_f$$

et

$$\forall i \in I, c_{fi} < c_{fj} \text{ et } c_{fi} - u_f < c_{fj} - u_f$$

$$\forall j \in J, c_{gj} < c_{gi} \text{ et } c_{gj} - u_g < c_{gi} - u_g$$

Donc la solution optimale est trouvée et l'algorithme s'arrête.

III.5 - Résultats numériques

Nous avons effectué différents tests sur les trois méthodes définies dans ce chapitre. Nous présentons dans les tableaux ci-dessous les résultats obtenus. On y trouve les informations suivantes :

La colonne (1) représente le taux de convergence de l'algorithme de sous-gradient.

La colonne (2) représente le pas initial de l'algorithme de sous-gradient.

La colonne (3) représente le numéro du jeu de données.

La colonne (4) représente le numéro de la méthode utilisée.

La colonne (5) représente le nombre d'itérations (algorithme primal, dual ou mixte)

La colonne (6) représente le temps d'exécution de l'algorithme.

La colonne (7) représente la balance ressource/demande (zone1/zone2/zone3)

La colonne (8) représente les flots inter-zone optima :

(zone12-zone21/zone13-zone31/zone23-zone32)

(zone12-zone21/zone13-zone31/zone43-zone34/

zone42-zone24/zone14-zone41/zone23-zone32)

Remarques :

- Le niveau de précision des données est $\tau = 10^{-7}$.

- Le nombre d'itérations maximum autorisé est $t_{\max} = N = 200$.

Résultats obtenus pour 3 zones et 3 points frontières :

taux de conv.	pas initial	jeux	méthode	itération	temps	balance	flot
0.8	10 100	1	1	112	32.5	12/-6/-6	6/10/0
			2	N	39.8		0-6/16-10/0
			2	3	9.7		6/10/0
0.8	10 100	2	1	113	30.9	-12/- 14/26	0/12/14
			2	39	15.9		0/12/14
			3	3	9.5		0/12/14
0.8	10 100	3	1	112	30.8	-20/- 14/34	0/20/14
			2	N	39.4		0/18-20/16-14
			3	2	9.4		0/20/14
0.8	10 100	4	1	112	30.8	20/-10/- 10	10/10/0
			2	10	10.7		10/10/0
			3	3	9.4		10/10/0
0.8	10 100	5	1	112	31.2	-4/-2/6	0/4/2
			2	9	10.6		0/4/2
			3	3	9.5		0/4/2
0.7	10	6	1	70	24.3	-4/10/6	4.8/2.2/3.7
			2	N	26.5		4-0/0/6-10
			3	3	10.2		4/0/6
0.7	10	7	1	71	26.3	-2/-36/38	0/36/2
			2	N	28.2		0/38-36/0-2
			3	3	10.4		0/36/2
0.7	10	8	1	70	24.6	12/-6/-6	6/6/0
			2	9	11.4		6/6/0
			3	3	10.5		6/6/0
0.7	10	9	1	70	24.9	8/-6/-2	5.9/2/0.1
			2	N	27.4		4-6/4-2/0
			3	3	10.3		6/2/0
0.7	10	10	1	71	24.6	-12/14/-2	12/0/2
			2	N	27.8		14-12/0/0-2
			3	4	10.3		12/0/2

Résultats obtenus pour 4 zones et 6 points frontières :

taux de conv.	pas initial	jeux	méthode	itération	temps	balance	flot
0.8	100	1	1	113	20.4	-2/-12	0.36/-1.38/0.83/6.58/-0.58/-4.61
			2	N	26.4	6/8	0/0/0/12-0/2-8/0-6
			2	N			
0.8	100	2	1	116	19.8	-6/-12/	0/0/2/11.98/6/0
			2	N	26.2	-2/20	0/0/2-0/12-0/6-20/0
			3	4	9.5		0/0/2/12/6/0
0.8	100	3	1	112	20.5	22/-16	7.6/0.26/-2/0/14.4/-7.6
			2	N	29.1	10/16	16.22/0/0-10/0/16-0/0
			3	N			
0.8	100	4	1	114	18.5	-2/-12	0/-2.9/-4/0.2/0/1--.7
			2	N	26.0	18/-4	0/2-0/4/0/0/12-14
			3	4	9.6		0/2/4/0/0/12
0.8	100	5	1	112	19.1	2/-24	2/-0.4/0/12/0/-9.6
			2	N	28.0	10/12	0-2/0/0/14-12/0/10
			3	3	10.0		2/0/0/12/0/10
0.8	100	6	1	101	19.9	-6/12	6/0/2/0/4/0
			2	N	13.2	-2/-4	6-0/0/2-12/0/4-0/0
			3	4	8.8		6/0/2/0/4/0
0.8	100	7	1	112	22.2	14/-20	14/0/-2/0/4/0
			2	N	13.2	2/4	18-14/0/0-2/0/2-4/0
			3	2	8.9		14/0/2/0/4/0
0.8	100	8	1	112	18.6	-14/-20	0/0/0/-14/-20/-2
			2	N	12.9	-2/36	0/0/0/14-0/20-36/2-0
			3	2	9.1		0/0/0/14/20/2
0.8	100	9	1	112	23.0	10/36	14.7/0/17.3/3.3/0/0.7
			2	N	13.6	-10/-36	18/0/14-18/0/0/4-0
			3	N			

Moyennes des résultats ci-dessus :

nb. zones	taux de conv.	pas initial	méthode	itération	temps	nb. optimum	nb. jeux
3	0.8	10	1	112.2	31.24	5	5
3	0.8	100	2	19.33	12.4	3	5
3			3	3	9.5	5	5
3	0.7	10	1	70	25	5	5
3	0.7	10	2	9	11.4	1	5
3			3	3	10.3	5	5
4	0.8	100	1	111.6	20.4	9	9
4	0.8	100	2	N	26	0	9
4			3	3	9.4	5	9

Analyse des résultats :

Ces résultats confirment nos propositions précédentes, concernant l'approche primale-duale:

Dégénérescence des sous-problèmes :

Si le nombre de zones excédentaires est égal à 1, alors

La zone excédentaire est dégénérée et traitée par l'approche primale.

Les zones déficitaires ne sont pas dégénérées et sont traitées par l'approche duale.

Si le nombre de zones déficitaires est égal à 1, alors

La zone déficitaire est dégénérée et traitée par l'approche primale.

Les zones excédentaires ne sont pas dégénérées et sont traitées par l'approche duale.

Si le nombre de zones excédentaires est égal au nombre de zones déficitaires alors

Si le nombre de zones est égal à 2

Les zones excédentaires sont dégénérées et traitées par l'approche primale.

Les zones déficitaires ne sont pas dégénérées et sont traitées par l'approche duale.

Sinon

On ne sait pas définir avec certitude les zones dégénérées.

Si le nombre de zones excédentaires est différent du nombre de zones déficitaires alors

On ne sait pas définir avec certitude les zones dégénérées.

Solutions exactes ou approchées :

La méthode 1 (approche primale) génère des solutions approchées, dans les conditions suivantes :

- L'algorithme de transbordement fonctionne avec des coûts entiers et des flots réels sur les arcs. La condition d'arrêt, portant sur les coûts réduits en nombres entiers, est exacte. La solution générée est optimale.

- L'algorithme de sous-gradient a une condition d'arrêt portant sur des flots en nombres réels. La solution générée (les flots inter-zone) est optimale à $\tau = 10^{-k}$ près ($k = 7$).

La méthode 2 (approche duale) génère des solutions approchées, dans les conditions suivantes :

- L'algorithme de transbordement fonctionne avec des coûts réels et des flots entiers sur les arcs. La condition d'arrêt, portant sur les coûts réduits en nombres réels, est inexacte. La solution générée est optimale à $\tau = 10^{-k}$ près ($k = 4$).

- L'algorithme de sous-gradient a une condition d'arrêt portant sur des variables duales en nombres réels. La solution générée (les variables duales des points frontières) est optimale à $\tau = 10^{-k}$ près ($k = 7$). Les flots inter-zones sont optima à $\tau = 10^{-k}$ près ($k = -2$). Cette différence de flot vient en général d'une zone telle que, par l'approche globale, on ait à l'optimum :

$$(\exists i \in N_r / a_i < 0), (\exists f \in F, g \in F) / (x_{fi} \neq 0, x_{gi} \neq 0)$$

Dans ce cas, le flot de l'approche duale, passe, soit par l'arc (fi), soit par l'arc (gi).

Conclusion

Conclusion

Cette thèse a été conduite dans le but de répondre à deux besoins spécifiques :

- La modélisation d'un problème réel de transport dynamique, visant la mise au point d'un logiciel fiable.
- L'élaboration d'un modèle original répondant à un besoin théorique relatif à l'étude d'une nouvelle méthode de décomposition.

La partie automatisée de la répartition comprend une interface utilisateur, un automate pour le déroulement du traitement et le dialogue avec la Base de Données et enfin un modèle mathématique pour le calcul de la solution.

Le noyau de ce modèle est défini comme un problème de Transbordement dynamique et stochastique. Nous avons choisi un algorithme de Transbordement dans la littérature, étudié la planification et la prévision pour la partie dynamique et stochastique de la formulation et intégré les principales contraintes du problème. Nous avons ensuite généré différents jeux de données, relatifs à une catégorie de matériel, pour tester ce noyau. Les nombres d'arcs et de sommets générés, pour satisfaire toutes les contraintes de l'application, nous ont posé des problèmes d'espace mémoire. Après différents tests, nous avons atteint 2000 demandes et 4800 ressources en moins de trois minutes.

Nous proposons aussi un horizon flexible, pour traiter des problèmes de plus grande dimension. En pratique, cette approche heuristique de décomposition est intéressante. En effet, elle nous permet de résoudre des problèmes de grande dimension, n'ayant pu être traités par une approche globale avec un algorithme de transbordement.

Nous présentons ensuite un modèle général incluant les catégories de matériel. Seulement, les contraintes et les variables du problème généré sont alors trop nombreuses pour être traitées globalement.

Nous avons disposé pour cette étude, d'une structure sous-jacente propre à la décomposition. C'est pourquoi, nous proposons au dernier chapitre une décomposition spatiale du système, pour résoudre les problèmes de grande dimension. Nous présentons dans un premier temps deux manipulations classiques : la relaxation Lagrangienne, suivant une approche duale du problème et l'ajout de variables de couplage, suivant une approche primale du problème. En accord avec ces manipulations, nous avons choisi d'expérimenter une méthode de sous-gradient. Dans un deuxième temps, nous proposons une méthode de décomposition mixte, à partir des résultats précédents.

Des critères de convergence facilement exploitables doivent être découverts pour éviter la phase de sous-gradient qui précède l'algorithme à un niveau et les heuristiques de choix des contraintes allouées dans les sous-problèmes.

Afin d'enrichir ce travail, il serait intéressant de mener une phase expérimentale de validation des résultats obtenus, concernant l'approche primale-duale, par comparaison avec l'approche globale.

Différentes extensions de ce travail sont envisageables. Ces prolongements concernent notamment l'étude d'un modèle utilisant les techniques Proximales ou l'élaboration d'un modèle basé sur la planification hiérarchisée.

L'idée de prise en compte explicite de l'incertitude sur les demandes peut être reprise dans un cadre plus large. Ceci peut se faire à partir d'autres types de modélisation (approche probabiliste par exemple).

A plus long terme, il semble également intéressant d'étendre l'étude menée ici dans le cadre d'une agrégation spatiale, à d'autres modes d'agrégation, notamment les catégories de wagons.

Annexes

Annexe 1 - Extrait du rapport EURECO

INTRODUCTION :

Le problème de la répartition est à la base un problème typique de Recherche Opérationnelle. Différentes études ont été menées à la SNCF pour définir la méthode la plus adaptée au problème de la répartition. Dans un premier temps, la société EURECO (cf. [Dej,1989]) a réalisé une étude diagnostique de la répartition et mené une étude comparative entre deux approches, l'IA (un système expert) et la RO. Dans un deuxième temps, la société BULL a proposé un langage déclaratif d'expression de contraintes, le langage CHARME (annexe 2).

EXTRAIT DU RAPPORT DE LA SOCIETE EURECO :

Rappel sur les systèmes d'information et d'aide à la décision :

Au cours des dix dernières années, un nouveau concept est apparu en IA : celui du système expert (SE). Dans ce paradigme on conçoit des systèmes informatiques qui raisonnent comme l'expert humain. Les SE constituent l'une des meilleures disciplines de l'IA. Les problèmes qu'elle aborde sont ceux qui sont soit mal définis, peu structurés et faisant appel à une connaissance difficile à formuler d'une manière explicite, soit des problèmes dont la combinatoire est tellement vaste que la solution optimale ne peut être atteinte par les systèmes informatiques classiques, dans des délais de temps raisonnables ou à cause des limitations de place mémoire. Elle s'emploie à trouver une solution "acceptable" dans des délais de temps raisonnables.

Quels sont les domaines d'application des SE ? Le diagnostic est l'une des applications les plus fréquentes des SE. La mise en oeuvre des connaissances est en

général assez aisée, en raison de la nature même de la démarche de diagnostic que l'expert humain met en oeuvre.

Le tableau ci-dessous condense les principales caractéristiques des SE et présente une comparaison avec celle des algorithmes d'optimisation classiques.

	systèmes experts	algorithmes d'optimisation
type de problème	mal défini	bien défini
type de données	souvent imprécises	bien connues
solution obtenue	approximative	exacte
méthodologie	heuristique	algorithmique
objectifs	simuler le raisonnement de l'expert	trouver la solution optimale

L'expérience pilote d'un SE pour la répartition [Sncf,1989]

Le système proposé doit fournir une aide à la répartition des wagons vides d'une catégorie donnée. La fonction principale de ce système est de fournir un ensemble d'ordres d'acheminement permettant de résoudre des déficits ou des sureffectifs constatés dans les différentes zones de répartition. Le système donne la possibilité de gérer les échanges de matériels par anticipation sur les cinq jours à venir. Les ordres de répartition sont donnés par zone et non par gare. Le système fonctionne suivant le schéma suivant : à l'aide des données historiques des sept dernières semaines, il prévoit les besoins et les ressources de chaque zone, ce qui lui permet d'avoir une estimation de la balance pour chaque zone (voir schéma suivant).

Le SE développé a trois points caractéristiques :

- Il s'agit d'un système hautement interactif permettant au répartiteur de pouvoir modifier à sa guise les solutions proposées.
- La répartition se fait suivant de nombreux critères, le coût de l'acheminement est loin d'être déterminant.

- La répartition est faite en considérant un horizon de cinq jours, ce qui permet d'anticiper les non-fournis. Cette idée, bien que très intéressante, est imprécise, puisque la prévision des ressources est basée sur des données collectées à 3 heures du matin (très peu fiables) et sans tenir compte des différences d'acheminement.

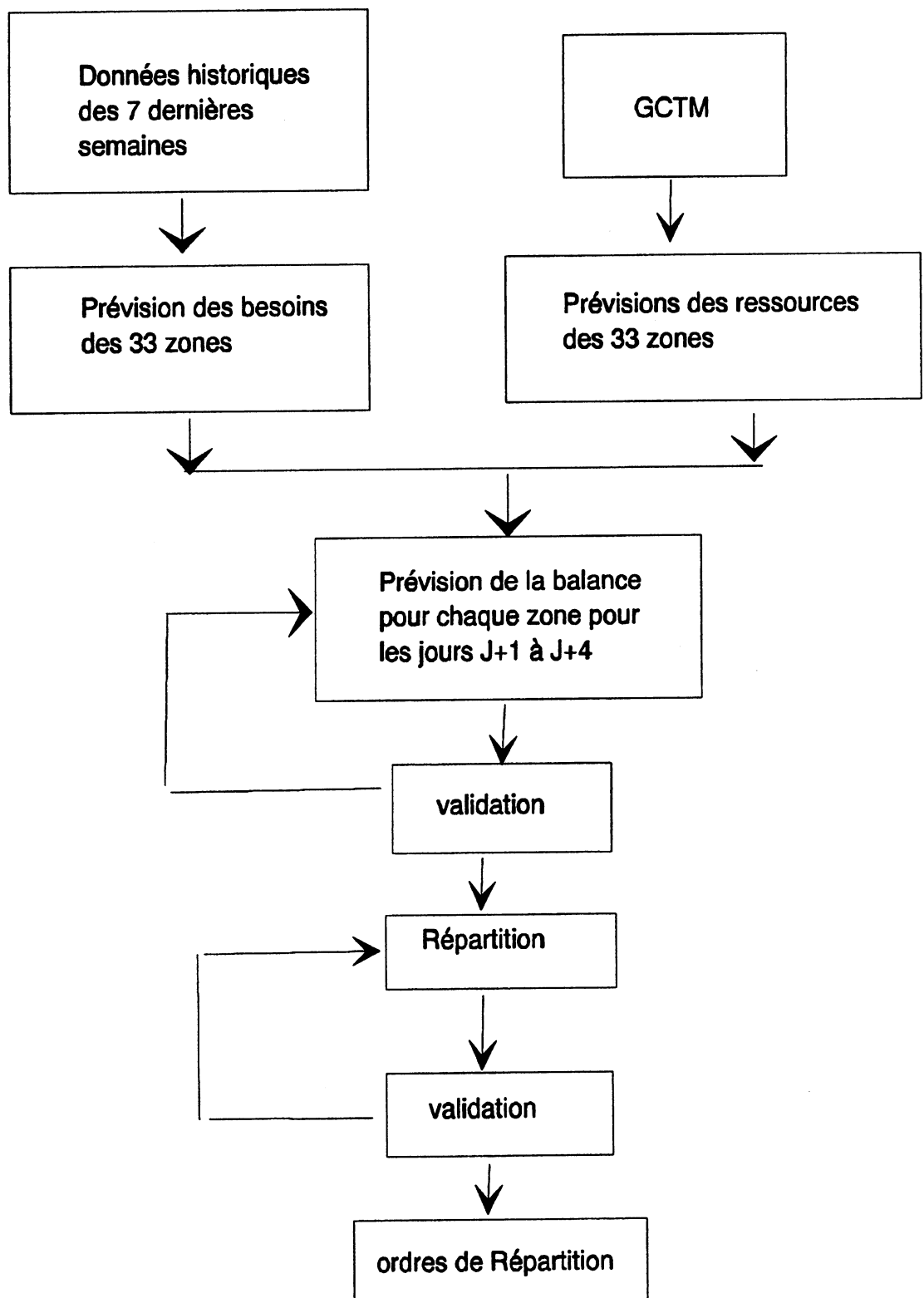
Une fois les données du problème correctement préparées, est-il pertinent d'effectuer la répartition proprement dite en utilisant les techniques des SE ? Nous pouvons poser la même question en d'autres termes, le jour où le SE sera parfaitement opérationnel, il effectuera la répartition en reproduisant la démarche des répartiteurs, est-ce la meilleure façon de résoudre le problème?

Il est parfaitement évident que la solution obtenue est loin d'être optimale.

Pourquoi chercher à simuler le comportement du répartiteur, alors que ce dernier, malgré tout son savoir-faire et sa bonne volonté, a beaucoup de mal à maîtriser la complexité intrinsèque de la tâche qui lui est assignée et cherche à trouver une solution réalisable, sans avoir la capacité de chercher une solution de coût minimal.

Nous pensons que l'utilisation des techniques de l'IA est parfaitement injustifiée pour le traitement de la répartition proprement dite. Le problème en soi est très complexe, mais parfaitement défini.

Une question se pose alors, peut-on résoudre ce problème grâce à un algorithme d'optimisation? Vu la complexité intrinsèque du problème, le nombre de variables de décision à déterminer est énorme et la résolution d'un tel problème par les techniques de l'optimisation aurait été impossible il y a une dizaine d'années. Mais aujourd'hui, avec le développement extraordinaire qu'a connu la programmation mathématique et l'optimisation combinatoire, une telle optimisation nous semble tout à fait possible. D'ailleurs un problème similaire a été très efficacement résolu par Dejax et Crainic [Cra, 1989] pour l'optimisation du transport de conteneurs vides et de nombreuses applications de l'optimisation à la gestion du ferroviaire vide sont citées en Europe et en Amérique du Nord.



Annexe 2 - Langage CHARME

MAQUETTE CHARME : première version

Une maquette a été réalisée par le Centre d'Expertise et de Développement en Intelligence Artificielle, Groupe BULL (BULL CEDIAG). Ce qui suit est extrait du dossier d'analyse "Maquette CHARME" et du cahier des charges "SNCF - BULL CEDIAG".

Objectifs de la maquette :

Prouver la faisabilité de l'application par l'utilisation de techniques d'IA ("programmation par contraintes") et trouver des stratégies permettant de minimiser les coûts.

Dans cette maquette, la répartition, effectuée sur 3 jours, doit mettre en évidence une méthode de programmation permettant de traiter le problème de façon très générale sur n jours.

La catégorie traitée est celle des E80, catégorie très importante, qui représente à elle seule près de 10% du parc total.

Les traitements sont différenciés selon la nature de la demande.

Le temps de traitement est inférieur à 1/4 d'heure.

Enjeux :

Les enjeux sont multiples :

- Assurer un coût minimum à chaque répartition grâce à l'utilisation de la meilleure stratégie.
- Effectuer 3 répartitions par jour sur l'ensemble des catégories de façon à prévoir beaucoup plus finement les mouvements des wagons, dans un temps très court.
- Intégrer ce système dans l'organisation déjà mise en place pour le NAW/BIG (notamment, les calculs d'ATHEO).

Algorithme général :

- Acquisition des données (demandes et ressources sur les 3 jours, délais, coûts).
- Préparation des matrices, représentant les mouvements des wagons par jour, en éliminant des colonnes les gares n'ayant pas exprimé de demandes pour le jour donné et des lignes les gares n'ayant pas de wagons disponibles pour le jour donné.
- Traitement incrémental par résolution de problèmes successifs : un problème par paire {jour de disponibilité, jour de la demande}, compte tenu des délais d'acheminement.

Temps de traitement : (3 périodes, 3 priorités)

nombre de gares	temps (mn,s) DPX 2000 (BULL)
50	0' 23"
100	1' 23"
175	5' 35"
200	limite taille mémoire

MAQUETTE CHARME : deuxième version

Objectifs de la maquette

- Obtenir une solution efficace, évolutive, maîtrisable et supportée.
- Intégrer un nouvel élément technologique dans la philosophie CHARME : un algorithme de transport totalement intégré dans CHARME.
- Rendre l'algorithme de transport totalement incrémental et sensible aux variations des données.

Algorithme général

- Exécution partielle de l'algorithme de transport pour réduire les domaines de définition des variables.
- Ajouter dynamiquement des équations sur des démons CHARME (par exemple, des capacités minima et maxima de stockage de wagons).

Temps d'exécution : 12 périodes, 6 priorités, rééquilibrage

nombre de gares	temps (mn,s) DPX 2000 (BULL)
50	0' 29"
100	7' 96"
150	40' 17"
200	126' 96"

Conclusion :

Les études, concernant la maquette CHARME, n'ont pas été prolongées, les temps d'exécution étant trop élevés.

Annexe 3 - Programmation Linéaire

Un problème de Programmation Linéaire consiste à Minimiser (Maximiser) une fonction linéaire sous contraintes linéaires, c'est à dire :

Minimiser $f(x)$

sous les contraintes (sc.)

$$g_i(x) = 0 \text{ (contraintes égalités)}$$

$$g_i(x) \leq 0 \text{ (contraintes inégalités)}$$

$$g_i(x) \geq 0 \text{ (contraintes inégalités)}$$

$$x = (x_1, x_2, \dots, x_n)^T \text{ dans } \mathbb{R}^n$$

où les f_i et g_i sont des fonctions linéaires affines des variables $x_1, x_2, \dots, x_n$. On peut toujours supposer que les variables x_j ($j=1, \dots, n$) sont astreintes à être non négatives.

On peut toujours mettre un Programme Linéaire sous forme standard (toutes les contraintes sont des égalités) en introduisant des variables supplémentaires appelées variables d'écart. Pour cette raison, on ne considérera dans ce qui suit, que des Programmes Linéaires sous forme standard du type :

(P) :

$$\text{Minimiser } z = c.x$$

sc.

$$A.x = b$$

$$x \geq 0$$

A est la matrice des contraintes de dimension $m \times n$ (n étant le nombre de variables et m le nombre de contraintes).

$c = (c_1, c_2, \dots, c_n)$ est le vecteur ligne des coûts.

$b = (b_1, b_2, \dots, b_m)^T$ est le vecteur colonne des seconds membres.

$z = c.x = \sum c_i.x_i$ est la fonction à optimiser (fonction objectif ou fonction économique).

$x = (x_1, x_2, \dots, x_n)^T$ est une *solution réalisable* de (P) ssi $A.x = b$ et $x \geq 0$.

Une *solution optimale* de (P) est une solution réalisable qui donne à la fonction $z=c.x$ une valeur maximale.

On appelle *base*, toute sous-matrice singulière ($m \times m$) extraite de A (il en existe au moins une si $\text{rang}(A) = m$).

Soit B une base telle, qu'en permutant les colonnes de A , on obtienne $A = [B, N]$ où N désigne les colonnes de A qui ne sont pas dans B . De même, on réalise une partition de x en $[x_B, x_N]^T$ et une partition de c en $[c_B, c_N]$.

x est une solution de (P) ssi $A.x = b$ et par suite $B.x_B + N.x_N = b$ (1).

On appelle *solution de base*, la solution particulière de (1) obtenue en faisant $x_N=0$.

x_B est alors déterminée de façon unique par la résolution du système de Cramer $B.x_B=b$, soit $x_B=B^{-1}b$

Une *solution de base* est dite *réalisable* si $x_B \geq 0$, autrement dit si $B^{-1}b \geq 0$.

Une base correspondant à une solution de base réalisable est appelée *base réalisable*.

Théorème :

Définissons le m vecteur-ligne $\pi = (\pi_1, \dots, \pi_m)$ par $\pi = c_B B^{-1}$

Une condition nécessaire et suffisante pour que B soit une *base réalisable optimale* est que :

$$\bar{c}_N = c_N - \pi \cdot N \geq 0$$

Le vecteur π est appelé le vecteur des *multiplieurs du Simplexe*. Les composantes $\bar{c}_j$ du vecteur $\bar{c}_N$ sont les *coûts réduits des variables hors-bases*. (démonstration : voir [Min,1979])

L'*algorithme Primal* du Simplexe s'interprète algébriquement comme la détermination d'une suite de bases adjacentes $B^0, B^1, \dots, B^q$ et de solutions de bases $x^0, x^1, \dots, x^q$ telles que

$$z(x^0) < z(x^1) < \dots < z(x^q) \dots$$

A chaque itération la solution courante est réalisable.

La notion de dualité est un concept fondamental en Programmation Linéaire et conduit à un résultat de grande portée théorique et pratique : le théorème de la dualité.

Le dual d'un Programme Linéaire sous forme standard s'écrit comme suit :

(D) :

Maximiser $w = u.b$

sc.

$u.A \geq c$

On associe à chaque contrainte i ($i=1,2,\dots,m$) de (P), une variable duale u_i positive, négative ou nulle.

Théorème de la dualité :

Etant donné un Programme Linéaire (P) et le Programme Dual (D) associé :

a/ Si (P) et (D) ont des solutions réalisables, alors chacun d'eux a une solution optimale et

$$z^* = \text{Max. } c.x = \text{Min } u.b = w^*$$

b/ Si l'un d'eux a un optimum non borné, l'autre n'a pas de solutions réalisables.
(Démonstration : voir [Min,1979]).

Théorème des écarts complémentaires :

Deux solutions (x,u) respectivement primale et duale réalisables sont optimales ssi

$$\forall j = 1, \dots, n \quad (u.A_j - c_j) \cdot x_j = 0$$

(Démonstration : voir [Min,1979]).

La variable duale u_i représente le coût d'une variation $\Delta(b_i) = 1$ du second membre de la $i^{\text{ème}}$ contrainte. On l'appelle le *coût marginal*.

L'algorithme dual consiste à appliquer l'algorithme Primal au Programme Dual. Les opérations de changement de base sont effectuées sur le tableau Primal.

L'algorithme Primal-Dual : Pour résoudre le Programme Linéaire (P), supposons que l'on connaisse une solution réalisable du dual, c'est à dire un vecteur $u=(u_1, u_2, \dots, u_m)$ vérifiant $u.A \leq c$. Comme toute solution réalisable de (P) vérifie $A.x=b$, alors x est solution optimale de (P) ssi x est solution optimale de :

(P') :

$$\text{Max. } z' = c'.x = (c - u.A) . x$$

sc.

$$A.x = b$$

$$x \geq 0$$

Annexe 4 - Données numériques

Evolution des effectifs du parc commercial, toutes catégories confondues et en répartition :

années	effectifs	chargements
janv. 88	104 594	134 174
janv. 89	99 820	131 098
janv. 90	95 548	123 838

Evolution (NB de wagons), rotation (Moyenne en jour), chargement (NB de wagons chargés dans le mois), % des effectifs et % des chargements pour les 10 plus grosses catégories (janv. 89) :

catégorie	effectifs	rotation	chargement	% effectifs	% chargement
G 40	8 504	30.65	7 695	8.9	6.21
K 50	7 933	23.80	9 437	8.3	7.62
E 80	7 253	21.02	9 548	7.59	7.71
R 9X	6 054	21.85	7 584	6.34	6.12
G 8	5 903	24.51	5 159	6.18	4.17
F 40	5 349	20.86	6 763	6.00	5.46
R 20	4 663	13.12	5 515	4.88	4.45
R 1X	3 731	21.55	4 878	3.90	3.94
R 80	2 796	16.44	4 795	2.93	3.87
G 10	2 466	19.61	2 973	2.58	2.04
Somme	54 652		64 34		
Moyenne		22.44			
%	57.2		51.96		

Les 10 plus grosses catégories représentent 57.2 % du total des effectifs et 51.96 % du total des chargements (janv. 89). Plus précisément, les G40 représentent 8.9 % du total des effectifs et 6.21 % du total des chargements.

Effectifs, rotation, chargement, % des effectifs et % des chargements pour les 10 plus petites catégories de wagons (janv. 90).

catégorie	effectifs	rotation	chargement	% effectifs	% chargement
R 47	5	22.53	7	< 0.1	< 0.06
H 38	8	8.66	361	< 0.1	2.9
H 81	16	23.21	9	0.16	0.07
S 63	20	13.09	116	0.21	0.9
S 61	24	6.94	82	0.25	0.6
S 30	25	47.79	22	0.26	0.2
F 71	25	46.09	0	0.26	
E 85	33	60.83	2	0.34	< 0.02
R 21	39	19.35	74	0.40	0.6
R 54	40	37.73	16	0.41	0.1
Somme	235		689		
Moyenne		32.34			
%	0.25		0.55		

On recense actuellement 120 catégories de wagons et 90 sous-catégories.

Bibliographie

- [Ahr,1980] Ahrens - Finke.
Primal Transportation and Transshipment Algorithms.
Zeitschrift für Operations Research. Vol 24, 1980, 1 - 32.
- [Ali, 1978] Ali - Helgason - Kennington - Lall.
Primal Simplex Network codes. State of the art.
Implementation technology. Networks. Vol 8, 1978, 315 - 339.
- [Ant, 1965] Anthony.
Planning and control systems : a framework for analysis.
Harvard University Press, Cambridge, Mass. 1965.
- [Arr,1960] Arrow - Hurwicz.
Decentralization and computation in resource allocation.
Essays in Economics and Econometrics, Pfouts ed., University of N. Carolina, 1960.
- [Bak, 1977] Baker.
Overview of computer based models applicable to freight car utilization.
Report prepared for the US department of transportation NTIS. Springfield. 1977.
- [Bal, 1962] Balas - Ivanescu.
On the transportation problem Part. I. Cahier du centre de RO. Vol 4 (2), 1962.
On the transportation problem Part II. Cahier du centre de RO. Vol 4 (3), 1962.
- [Baz, 1977] Bazarra - Jarvis.
Linear programming and network flows. New York, 1977.
- [Bea, 1984] Bean - Smith.
Conditions for the existence of planning horizons.
Math. of Op. Res. Vol 9 (3), 1984.
- [Bea, 1991] Beaujon - Turnquist.
A model for fleet sizing and vehicle allocation.
Transportation Sc. Vol 25 (1), 1991, 19 - 45.
- [Ben,1962] Benders.
Partitioning procedure for solving mixed variables programming problems.
Numerische Mathematik. Vol 4, 1962, 238 - 252.
- [Ben, 1973] Bennington.
An efficient minimum cost flow algorithm.
Man. Sci. Vol 19 (9),1973, 1042 - 1051.
- [Ben, 1983] Bensoussan - Crouhy - Proth.
Mathematical theory of production planning. North Holland. 1983.
- [Ber,1989] Bertsekas - Tsitsiklis.
Parallel and distributed computation, numerical methods.
Prentice - Hall, International Editions, 1989.
- [Ber,1992] Bertsekas.

- Linear Network optimization, 1992.
- [Bra, 1977] Bradley - Brown - Graves.
Design and implementation of large scale primal transshipment algorithms.
Man. Sci. Vol 24 (1), 1977, 1 - 32.
- [Big, 1985] Bigras.
La modélisation des flux de marchandises au Canada.
Univresité de Montréal - CRT. Publication 483, 1985.
- [Bus, 1961] Busacher - Gowen.
A procedure for determining a family of minimum cost flow patterns.
John Hopkins University. ORO technical report. Vol 15, 1961.
- [Cha, 1961] Charnes - Cooper.
Management models for industrial applications of linear programming.
John Wiley & Sons Inc, New York. Vol 2, 1961.
- [Coo, 1977] Cooper - Leblanc.
Stochastic tranportation problems and other network related convex problems.
Nav. Res. Log. Quat. Vol 24, 1977, 327 - 337.
- [Cra, 1982] Crainic.
Un modèle de planification tactique pour le transport ferroviaire des marchandises.
Thèse de doctorat. CRT. Montréal. Publication 247, 1982.
- [Cra, 1984] Crainic - Ferland - Rousseau.
A tactical planning model for rail freight transportation.
Trans. Sci. Vol 18, 1984, 165 184 .
- [Cra, 1986](a) Crainic - Rousseau.
Multimode, multicommodity freight transportation : a general modelling and algorithmic framework for the service network design.
Trans. Res. Vol 20 B, 1982, 225 - 242.
- [Cra, 1986](b) Crainic - Dejax - Delorme.
A warehouse location model for empty container distribution.
TIMS/ORSA national meeting, Miami, 1986.
- [Cra, 1987](a) Crainic.
Intercity multimode freight transportation : the state-of-the-art.
The logistics and transportation review. Vol 23, 1982, 189 - 206.
- [Cra, 1987](b) Crainic - Roy.
OR tools for the tactical planning of freight transportation.
CRT. Montréal. Publication 398, 1987.
- [Cra, 1987](c) Crainic - Florian - Guélat.
OR tools for the statagic planning of national freight systems : modelling.
CRT. Montréal. 1987.

- [Cra, 1987](d) Crainic - Florian - Guélat.
OR tools for the strategic planning of national freight systems : algorithms.
CRT. Montréal. 1987.
- [Cra, 1987](e) Crainic - Florian - Leal.
A model for the strategic planning of national freight transportation by rail.
CRT. Montréal. 1987.
- [Cra, 1989](a) Crainic - Dejax - Gendreau.
Transport terrestre multi-modal de conteneurs : problématique et modèles.
CRT. Montréal. Publication 656, 1989.
- [Cra, 1989](b) Crainic - Dejax - Delorme.
Models for Mutimode Multicommodity Location Problems with interdepot
balancing requirements.
Annals of Operations Research. Vol 18, 1989, 279 - 302.
- [Cra, 1990] Crainic - Dejax - Gendreau.
Modelling the container fleet management prpbem using a stochastic dynamic approch.
Operational Research'90, Bradley (ed.). Pergamon Press, N.Y., 473 - 486.
- [Cra, 1993] Crainic - Dejax - Gendreau.
Dynamic and stochastic models for the allocation of empty containers.
Op. Res. Vol 41 (1), 1993, 102 - 126.
- [Coh,1984] Cohen.
Decomposition et coordination en optimisation déterministe, différentiable et
non différentiable.
Thèse d'état. Paris. 1984.
- [Dan, 1951] Dantzig.
Application of the Simplex method to a transportation problem.
In "Activity Analysis of Production and Allocation".
Edit. T.C. Koopmans. Chap. XXIII. 1951.
- [Dan, 1961] Dantzig - Wolfe.
The decomposition algorithm for linear programming.
Econometrica. Vol 29 (4), 1961, 767 - 778.
- [Dan, 1963] Dantzig.
Linear Programming and extensions. Princeton University Press. 1963.
- [Dec, 1978] Decision system analysis, Inc.
Implementation studies via simulation of the computer based model for freight
car distribution.
Report prepared for U.S. Department of Transportation, FRA, DSAI, Md. 1978.
- [Dej, 1986](a) Dejax.
Methodology for facility location and distribution systems planning.

- Technical report. Ecole Centrale de Paris, France. LEIS, Vol 86 - 04, 1986.
- [Dej, 1986](b) Dejax - Bine - Gozlan.
Le problème du transport à vide.
Technical report. Ecole Centrale de Paris, France. LEIS, Vol 86 - 03, 1986.
- [Dej, 1986](c) Dejax - Durand - Servant.
A planning model for loaded and empty container ground transportation.
TIMS/ORSA National meeting, Los Angeles. 1986.
- [Dej, 1987] Dejax - Crainic.
A review of empty flows and fleet management models in freight transportation.
Trans. Sci. Vol 21, 1987, 227 - 247.
- [Dej, 1988] Dejax - Crainic - Delorme.
Strategic/tactical planning of container land transportation systems.
TIMS XXVIII/EURO IX Joint Meeting, Paris, July 6 - 8.
- [Dej, 1989] Dejax - Haouari - De Tocqueville.
Etude diagnostic de la Répartition actuelle des wagons réseau SNCF.
Société EURECO, Paris. 1989.
- [Dej, 1992] Dejax - Benamar - Crainic - Gendreau.
Short term container fleet management : issues, models and tools.
In proceedings of the 6th world conference on transport research.
Lyon, France, June 29 - July 3.
- [Ede, 1977] Edelstein - Melnyk.
The pool control system.
Interface 8, 1977, 21 - 36.
- [Edm, 1972] Edmonds - Karp.
Theoretical improvements in algorithmic efficiency for network flow problem.
Journal of the ACM. Vol 19 (2), 1972, 248 - 264.
- [Fee, 1957] Feeney.
Controlling the distribution of empty freight cars.
Proceedings of the tenth national meeting. OR society of America, Baltimore. 1957.
- [Flo, 1971] Flood.
A transportation algorithm and code.
Nav. Res. Log. Quat. Vol 8, 1971, 257 - 276.
- [Flo, 1984] Florian.
Transportation planning models.
North-Holland, Amsterdam, 1984.
- [Flo, 1986](a) Florian - Crainic - Guélat - Babin.
An interactive graphic system for strategic planning of freight transportation
at the national and regional level.

- CRT. Montreal. publication 495, 1986
- [Flo, 1986](b) Florian - Crainic - Sénécal - Nguyen.
A multi-product multimode freight assignment model for national transportation planning.
TIMS/ORSA national meeting, Los Angeles. 1986
- [For, 1957] Ford - Fulkerson.
A primal-Dual algorithm for the capacited Hitchcock problem.
Nav. Res. Log. Quart. Vol 4 (1), 1957,.
- [For, 1962] Ford - Fulkerson.
Flows in Network. Princeton University Press. 1962.
- [Ful, 1961] Fulkerson.
An out-of-kilter method for minimal cost flow problem.
Journal SIAM. Vol 9 (1), 1961, 18 - 27.
- [Geo, 1968] Geoffrion.
Primal ressource-directive approches for optimizing nonlinear decomposable system.
Memorendum RM - 5829 - PR, the RAND Corp., Santa Monica.
Operations Research. Vol 18, 1968, 375 - 403.
- [Geo, 1974] Geoffrion.
Lagrangian relaxation and its uses in integer programming.
Math. Prog. Study. Vol 2, 1974, 82 - 114.
- [Gia, 1981] Giard.
Gestion de la production - Calcul économique.
Ed. Economica, 1981.
- [Gli, 1985] Glickman - Sherali.
Large scale network distribution of pooled empty freight cars over time, with limited substitution and equitable benefits.
Trans. Res. Vol 19 B, 1985, 85 - 94.
- [Glo, 1974] Glover - Karney - Klingman - Napier.
A computational study on start procedures, basis change criteria and solution algorithms for Transportation Problems.
Man. Sci. Vol 20 (5), 1974, 793 - 813.
- [Glov 1974] Glover - Karney - Klingman.
Implementantion and coputational comparisons of Primal, Dual and Primal-Dual computer codes for minimum cost Network flow problem.
Network. Vol 4. 1974, 191 - 212.
- [Gri, 1977] Grinold.
Finite horizon approximations of finite horizon linear programs.
Math. Prog. Vol 12, 1977, 1 - 17.

- [Gri, 1983] Grinold.
Convex infinite horizon programs.
Math. Prog. Vol 25, 1983, 64 - 82.
- [Hel,1974] Held - Wolfe - Crowder.
Validation of subgradient optimization.
Math. Prog. Vol 6, 1974, 62 - 88.
- [Her, 1973] Herren.
The distribution of empty wagons by means of computer : an analytical model of the Swiss Federal Railways.
Rail International. Vol 4, 1973, 1005 - 1010.
- [Her, 1977] Herren.
Computer controlled empty wagon distribution on the SBB.
Rail International. Vol 8, 1977, pp 25 - 32.
- [Hop, 1973] Hopcroft - Karp.
An $O(n^5/2)$ algorithm for maximum matching in bipartite graphs.
SIAM J. Comp. Vol 2, 1973, 225 - 231.
- [Jor, 1983] Jordan - Turnquist.
A stochastic dynamic network model for railroad car distribution.
Trans. Sc. Vol 17, 1983, 123 - 145.
- [Jor, 1986] Jordan - Turnquist - Bustamante.
Mode choice and vehicle utilisation in large distribution networks.
TIMS/ORSA national meeting, Miami. 1986.
- [Kle, 1967] Klein.
A primal method for minimal cost flows with applications to the assignment and transportation problems.
Man. Sci. Vol 14, 1967, 205 - 221.
- [Knu, 1968] Knuth.
The art of computer programming. Vol 1. Another enumeration of trees.
Can. J. Math. Vol 20, 1968, 1077 - 1086.
- [Kuh, 1955] Kuhn.
The Hungarian method for the assignment problem.
Nav. Res. Log. Quat. Vol 2, 1955, 83 - 97.
- [Lan, 1974] Langley - Kennington - Shetty.
Efficient Computational Devices for the Capacited Transportation Problem.
Nav. Res. Log. Quat. Vol 21, 1974, 637 - 647.
- [Las,1968] Lasdon.
Duality and decomposition in mathematical programming.
IEEE. Trans. Syst. Sci. and Cybernetic. Vol 4, 1968, 86 - 100.

- [Las, 1970] Lasdon.
Optimization for large systems.
Mac Millan series in Operation Research, New York. 1970.
- [Las, 1986] Lasserre.
Detecting planning horizons in deterministic infinite horizon optimal control.
IEEE Transactions on Automatic Control. Vol 31 (1), 1986, 70 - 74.
- [Las, 1989] Lasserre.
A tentative model for consistent hierarchical production planning and scheduling.
Rapport de recherche LAAS 89030, 1989.
- [Leg, 1977] Legendre - Minoux.
Une application de la notion de dualité en PNE : sélection et affectation optimale d'une flotte d'avions.
RAIRO. Vol 11 (2), 1977, 201 - 222.
- [Lem, 1975] Lemaréchal.
An extension of Davidon's methods to non differentiable problems.
Math. Prog. Study. Vol 3, 1975, 95 - 109.
- [Mag, 1976] Magnanti - Shapiro - Wagner.
Generalized linear programming solves the dual.
Man. Sci. Vol 22, 1976, 1195 - 1203.
- [Mah, 1986] Mahey.
Méthodes de décomposition et décentralisation en Programmation Linéaire.
RAIRO. Vol 20 (4), 1986, 287 - 306.
- [Mah, 1992] Mahey - Nguyen - Oualibouch - Pham.
Proximal techniques for convex programming.
Rapport de recherche. Artemis/IMAG, Grenoble. 1992.
- [Mar, 1987] Martin - Intrand.
ETNA : Le rôle de l'informatique dans la gestion des priorités entre wagons dans un plan de transport unique.
Revue Générale des Chemins de Fer. 1987.
- [Med, 1987] Medhi.
Decomposition of structured large scale optimization problems and parallel optimization.
Ph. D. Thesis, University of Wisconsin, Madison. 1987.
- [Mer, 1987] Merce.
Cohérence des décisions en planification hiérarchisée.
Thèse, LAAS, Toulouse. 1987.
- [Min, 1979] Gondran - Minoux.
Graphes et algorithmes. Collection Eyrolles. 1979.

- [Min,1981] Minoux - Serreault.
Subgradient optimization and large scale programming : an application to optimum multicommodity, network synthesis with security constraints.
RAIRO. Vol 15 (2), 1981, 185 - 203.
- [Min, 1983] Minoux.
Programmation Mathématique (2 tomes). Col. Dunod. 1983.
- [Mis, 1972] Misra.
Linear programming of empty wagon disposition.
Rail International. Vol 3, 1972, 151 - 158.
- [Mod, 1978] Moder - Elmaghraby.
Handbook of Operations Research : Models and Applications. 1978.
- [Pic,1987] Picard.
Fret : un modèle de flux de marchandises au Canada.
Université de Montréal. Publication 539, 1987.
- [Pow, 1984] Powell - Sheffi - Thiriez.
The dynamic vehicle allocation problem with uncertain demands.
Ninth International symposium on transportation and traffic theory. VNU. Science press, Utrecht. 1984, 357 - 374.
- [Pow, 1986](a) Powell.
A local improvement heuristic for the design of less-than-truckload carrier network.
Trans. Sci. Vol 20, 1986, 246 - 257.
- [Pow, 1986](b) Powell.
A stochastic model of the dynamic vehicle allocation problem.
Trans. Sci. Vol 20, 1986, 117 - 129.
- [Pow, 1986](c) Powell - Sheffi.
Interactive optimization for motor carrier load planning.
J. Bus. Logist. Vol 7, 1986, 64 - 90.
- [Pow, 1987](a) Powell.
A comparative review of alternative algorithms for the dynamic vehicle allocation problem.
Princeton University. SOR. 1987.
- [Pow, 1987](b) Powell.
An operational planning model for the dynamic vehicle allocation problem with uncertain demands.
Trans. Res. Vol 21 B, 1987, 217 - 232.
- [Pow, 1988] Powell - Sheffi - Nickerson - Butterbaugh - Atherton.
Maximizing profits for North American Van Lines' Truckload Division : A new framework for pricing and operations.

- Interface. Vol 18, 1988, 21 - 41.
- [Roc,1970] Rockafellar.
Convex Analysis. Princeton University Press. 1970.
- [Roy, 1985] Roy.
Modélisation stratégique du transport de marchandises au Brésil : le mode ferroviaire et l'analyse de matrices de demandes.
CRT, Montréal. Publication 578, 1985.
- [Roy, 1986] Roy.
Modélisation stratégique du transport de marchandises au Brésil
M.Sc. thesis. CRT, Montréal. Publication 485, 1986.
- [San, 1990] Sanso - Desrochers - Desrosiers - Dumas - Soumis.
Modeling and solving routing and scheduling problems : GENCOL user guide.
GERAD. 1990.
- [Sch, 1991] Schlaepfer.
Das mathematische modell. Rechnergesteuerte Leerwagenverteilung (LWV).
Deutsche Bundesbahn (DB). 1991.
- [Sho,1968] Shor.
On the rate of convergence of the generalized gradient method.
Kibernetika. Vol 4 (3), 1968.
- [Sho, 1985] Shor.
Minimization methods for nondifferentiable functions. Springer V. 1985.
- [Sim, 1966] Simonnard.
Linear Programming. Col. Dunod. 1966.
- [SnCF, 1989] SNCF : Document interne.
Système Extert d'aide à la Répartition des wagons vides, 1989.
Direction des Etudes, de la Planification et de la Recherche.
Département : Recherche Cybernétique et Technique, Paris.
- [SnCF,1990] SNCF - Document interne.
Acheminement des wagons isolés, Paris. 1990.
- [Spi,1983] Spingarn.
Partial inverse of a monotone operator.
Appl. Math. and Optim. Vol 10, 1983, 247 - 265.
- [Sri, 1973] Srinivasen - Thompton.
Benefit cost analysis of coding techniques for the primal transportation algorithm.
JACM. Vol 20, 1973, 194 -213.
- [Tur,1975] Turnquist.
Research opportunities in transportation system characteristics and operations,
Trans. Res. Vol 19 A, 1985, 357 - 366.

- [Tur,1982] Turnquist - Jordan.
 A computer based method for railroad car distribution.
 Final report, DTRS5680-C-00013, U.S. Department of transportation,
 office of university research, 1982.
- [Tur,1985] Turnquist - Jordan.
 Fleet sizing under production cycles and uncertain travel times.
 General Motors Research Laboratories, Warren, Mich.
 Research Publication GMR - 5012.
- [Tur,1986](a) Turnquist.
 MOV - EM : A network optimization model for empty freight car distribution.
 School of civil and environmental engineering, Cornell University, N.Y. 1986.
- [Tur,1986](b) Turnquist - Bustamante.
 Mode choice and vehicle utilization in large production and distribution
 networks.
 School of civil and environmental engineering, Cornell University, N Y 1986.
- [Tur,1987] Turnquist - Jordan.
 Coordinating product distribution and equipment utilization decisions in large
 networks.
 General Motors Research Laboratories, Warren, Mich.
 Research Publication GMR - 5705.
- [Tit,1975] Titli.
 Commande hiérarchisée et optimisation des systèmes complexes.
 Dunod automatique. Paris. 1975.
- [Whi, 1969] White - Bomberault.
 A network algorithm for empty freight car allocation.
 IBM. Vol 8 (2), 1969, 147 - 169.
- [Whi, 1972] White.
 Dynamic transshipment network : an algorithm and its application for the
 distribution of empty containers.
 Networks. Vol 2, 1972, 211 - 236.