

INÉQUATIONS VARIATIONNELLES: MOTIVATION, ALGORITHMES DE RÉOLUTION ET QUELQUES APPLICATIONS

Patrice MARCOTTE

Cours donné à Zinal, Suisse, 4–8 mars mil-neuf-cent-nonante-sept
Révision février 2009

0. Introduction et notions de base
1. Motivation
 - 1.1 Premier exemple: équilibre de Nash (2 joueurs)
 - 1.2 Deuxième exemple: équilibre de trafic urbain
2. Formulations
 - 2.1 Caractérisation géométrique
 - 2.2 Formulation de point fixe
 - 2.3 Formulation primale-duale
 - 2.4 Formulation sous forme d'équation
 - 2.5 Formulations d'optimisation
3. Existence et unicité
4. Algorithmes
 - 4.1 Algorithmes de point fixe
 - 4.2 Algorithmes d'approximation
 - 4.3 L'extragradiant: algorithmes de Korpelevich et de Konnov
 - 4.4 Les algorithmes d'optimisation
 - 4.5 Les algorithmes de plans coupants
5. Applications
 - 5.1 Équilibre oligopolistique
 - 5.2 Équilibres sur des réseaux de transport
 - 5.3 Équilibre bicritère
 - 5.4 Équilibres stratégiques
 - 5.5 Équilibre dynamique
6. Programmation à deux niveaux

0. INTRODUCTION ET NOTIONS DE BASE

Ce document accompagne la présentation donnée à Zinal en mars 1997, dans le cadre des séminaires du troisième cycle romand. Il se veut une courte introduction aux modèles d'équilibre ainsi qu'un guide vers des ouvrages ou articles traitant de façon plus approfondie des aspects théoriques, algorithmiques ou pratiques liés à l'étude des inégalités variationnelles. Une section du document sera consacrée à chacun de ces aspects. Je mentionne quatre bonnes sources d'informations concernant les inéquations variationnelles, la première traitant plus particulièrement des inégalités variationnelles en dimension infinie : Glowinsky, Lions et Trémolières [**GloLT76**], Auslender [**Ausle76**], Pang et Chan [**PangC82**] Harker et Pang [**HarPa90**], Bertsekas et Tsitsiklis [**BerTs89**], Patriksson [**Patri94**].

Quelques notions de programmation mathématique (programmation linéaire, conditions de Kuhn-Tucker, conditions d'optimalité du premier ordre, taux de convergence), de topologie élémentaire (convexité, ensembles ouverts, fermés, bornés, compacts), d'analyse (continuité, théorème de Weierstrass, convergence des suites, développement en série de Taylor, gradient, matrice jacobienne, dérivée directionnelle) et d'algèbre linéaire (notation matricielle, valeurs propres) devraient suffire à assurer la compréhension du texte. Le seul résultat non 'élémentaire' qui sera utilisé est le théorème de point fixe suivant, dont une preuve constructive est présentée, entre autres, dans le joli petit livre de Border [**Borde85**] consacré aux théorèmes de points fixes et à leurs applications.

Théorème (Brouwer) Soit C un sous-ensemble convexe et compact de R^n et F une fonction continue de C dans C . Alors F possède un point fixe x^* appartenant à C , c'est-à-dire : $F(x^*) = x^*$.

J'ai évité de reproduire des preuves complexes. Cependant j'ai esquissé certaines démonstrations afin d'illustrer des techniques d'usage fréquent. Les principales notations utilisées sont résumées ci-dessous. De façon générale, les vecteurs qui se situent à gauche d'un produit scalaire sont des vecteurs ligne et les vecteurs qui se situent à droite des vecteurs colonne :

$x \in R^n :$	vecteur colonne
$\langle x, y \rangle = x^t y :$	produit scalaire usuel des vecteurs x et y
$f(x) : R^n \rightarrow R :$	fonction réelle de n variables
$F(x) : R^n \rightarrow R^n :$	fonction vectorielle (vecteur ligne)
$f'(x) :$	gradient de f (vecteur ligne des dérivées partielles $f'_i(x)$)
$f'(x; d) :$	dérivée directionnelle de f dans la direction d
$f'_i(x) :$	dérivée partielle par rapport à la i ème variable
$f''(x) :$	matrice hessienne de f (matrice des dérivées d'ordre 2)
$F'(x) :$	matrice jacobienne de F (matrice des dérivées partielles)

Je termine cette introduction par l'énoncé de deux conditions nécessaires d'optimalité dont l'incontournable condition d'optimalité due à Kuhn et Tucker (et Karush, pour les puristes) pour le programme mathématique :

$$\begin{aligned} \min_{x \in X} \quad & f(x) \\ X = \{x \in R^n : & g_i(x) \leq 0, i \in I\} \\ I = \{1, 2, \dots, & m\}, \end{aligned} \tag{1}$$

où les fonctions f et g_i sont continûment différentiables. Je pose :

$$G(x) = (g_1(x), \dots, g_m(x))^t.$$

Une direction d est ADMISSIBLE s'il existe un nombre positif α tel que $x + td \in C$ pour $t \leq \alpha$.

Théorème Soit x un minimum local pour (1). On a que

$$f'(x; d) \geq 0 \tag{2}$$

pour toute direction admissible d , c'est-à-dire qu'il n'existe pas de direction de descente admissible.

Tout point x de X satisfaisant la condition nécessaire (2) est appelé POINT STATIONNAIRE du programme mathématique (1).

Théorème (Kuhn et Tucker) Un point RÉGULIER x de X est stationnaire si et seulement si il existe un vecteur de MULTIPLICATEURS $\lambda \in R^m$ tel que :

$$\begin{aligned} G(x) & \leq 0 \\ f'(x) + \lambda G'(x) & = 0 \\ \lambda & \geq 0 \\ \langle \lambda, g(x) \rangle & = 0. \end{aligned} \tag{3}$$

Les démonstrations de ces deux résultats, ainsi que diverses notions de régularité, se trouvent dans le livre de Bazaara, Sherali et Shetty [**Bazar94**]. Un cas particulier important se présente lorsque les fonctions g_i sont affines :

$$X = \{x \in R^n : Ax \leq b\}$$

pour une certaine matrice A de dimension $m \times n$. Tout point de X est alors régulier et le système PRIMAL-DUAL (3) prend alors la forme :

$$\begin{aligned} Ax & \leq b \\ f'(x) + \lambda A & = 0 \\ \lambda & \geq 0 \\ \lambda(Ax - b) & = 0. \end{aligned} \tag{4}$$

Si les fonctions f et l'ensemble X sont convexes, un point régulier x de X est un minimum global pour le problème (3) si et seulement si il satisfait les CONDITIONS DE KUHN ET TUCKER (Karush-Kuhn-Tucker pour les puristes).

1. MOTIVATION

Soit $x \in R^n$. On dit que x^* est une solution de l'inégalité variationnelle $IV(F, X)$ associée à la fonction $F : R^n \rightarrow R^n$ et à l'ensemble convexe $X \subseteq R^n$ s'il satisfait l'inéquation :

$$IV(F, X) : \quad \langle F(x^*), x^* - x \rangle \leq 0 \quad \forall x \in X, \quad (5)$$

et je dénoterai par X^* le sous-ensemble de X satisfaisant (5).

Les inégalités variationnelles permettent la représentation mathématique d'équilibres physiques, chimiques, économiques, voire militaires. Un équilibre du système correspondra à une solution de l'inégalité variationnelle (5). Les situations que nous considérerons sont de nature économique ou relèvent de la théorie des jeux élémentaire.

1.1. Premier exemple : équilibre de Nash (2 joueurs)

Soient deux joueurs cherchant à optimiser, sur des ensembles disjoints, une fonction impliquant leurs stratégies respectives. Soient x_i le vecteur stratégique du joueur i , X_i son ensemble de stratégies admissibles et $f_i(x_1, x_2)$ sa fonction de coût convexe. La réaction $x_1(x_2)$ du joueur 1 à la stratégie x_2 adoptée par le joueur 2 satisfait :

$$\min_{x_1 \in X_1} f_1(x_1, x_2) = f_1(x_1(x_2), x_2)$$

et, symétriquement pour le joueur 2, j'obtiens :

$$\min_{x_2 \in X_2} f_2(x_1, x_2) = f_2(x_1, x_2(x_1)).$$

Si je pose :

$$F_1(x_1, x_2) = f_1'(x_1(x_2), x_2)$$

(le gradient est calculé par rapport au vecteur x_1) et

$$F_2(x_1, x_2) = f_2'(x_1, x_2(x_1))$$

(ce deuxième gradient est calculé par rapport au vecteur x_2), les 'réactions' $x_1(x_2)$ et $x_2(x_1)$ sont caractérisées par les conditions du premier ordre suivantes :

$$\begin{aligned} \langle F_1(x_1(x_2), x_2), x_1(x_2) - x_1 \rangle &\leq 0 & \forall x_1 \in X_1 \\ \langle F_2(x_1, x_2(x_1)), x_2(x_1) - x_2 \rangle &\leq 0 & \forall x_2 \in X_2. \end{aligned}$$

Un équilibre de Nash est obtenu lorsque :

$$\begin{aligned} x_1(x_2^*) &= x_1^* \\ x_2(x_1^*) &= x_2^*. \end{aligned}$$

Le système de conditions nécessaires ci-dessus se reformule comme (5) si je pose :

$$x^* = \begin{pmatrix} x_1^* \\ x_2^* \end{pmatrix} \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad F(x) = \begin{pmatrix} F_1(x_1, x_2) \\ F_2(x_1, x_2) \end{pmatrix} \quad X = X_1 \times X_2.$$

Consulter Owen [Owen] pour plus de détails.

1.2. Deuxième exemple : équilibre de trafic urbain

Soit un réseau composé de n chemins (pas forcément disjoints) reliant tous une origine A à une destination B , et d le débit de véhicules, exprimé en véhicules par minute, souhaitant quitter A pour relier B . Le temps de parcours du chemin j est donné par la fonction de délai $F_j(x)$, où $x = (x_1, \dots, x_n)^t$. Un ÉQUILIBRE DE WARDROP x^* est atteint lorsque les temps de parcours des chemins utilisés sont tous égaux, et inférieurs aux coûts des chemins non utilisés. Mathématiquement, ces conditions prennent la forme :

$$\begin{aligned} F_j(x^*) &= u^* & \forall j : x_j^* > 0 \\ F_j(x^*) &\geq u^* & \forall j \\ \sum_{j=1}^n x_j^* &= d \\ x^* &\geq 0. \end{aligned}$$

En posant $X = \{x \in R_+^n : \sum_{j=1}^n x_j = d\}$, on peut facilement démontrer que les solutions du système ci-dessus correspondent aux solutions de $IV(F, X)$.

Ce MODÈLE D’AFFECTATION DU TRAFIC peut mener à certaines situations qui peuvent sembler paradoxales. Considérons en effet le réseau illustré à la figure 1 où les fonctions de délai sur les arcs (dont l’unique paramètre est le flot d’arc v) sont indiquées. Le délai sur le chemin j est obtenu en additionnant les délais sur les arcs formant le chemin j . Le réseau est formé des trois chemins ACD, ABD et ABCD, dont les fonctions de délai sont données à la table 1. La demande est fixée à 6 unités de flot entre l’origine A et la destination D. A l’équilibre, 2 unités sont affectées à chacun des trois chemins. Ceci résulte en un délai de 92 pour chaque usager, et le coût total pour le système est égal à $6 \times 92 = 552$. On constate qu’il aurait été plus efficace, du point de vue du système et des usagers, d’affecter 3 de ces derniers au chemin ABD, et 3 au chemin ACD, au coût individuel de 83, et au coût total de 498. Malheureusement, cette solution est ‘instable’ car certains usagers seront inmanquablement attirés par le chemin ABCD de coût $70 < 83$. Sans coopération entre les usagers (ce qui serait profitable pour tous) la solution optimale est irréalisable. Un planificateur astucieux, quant à lui, interdirait l’accès à l’arc diagonal BC, en établissant par exemple un péage prohibitif.

Consulter Patriksson [Patri94] pour une revue exhaustive des modèles d’affectation du trafic.

2. FORMULATIONS

Dans cette section, nous allons donner plusieurs formulations équivalentes de l’inégalité variationnelle $IV(F, X)$.

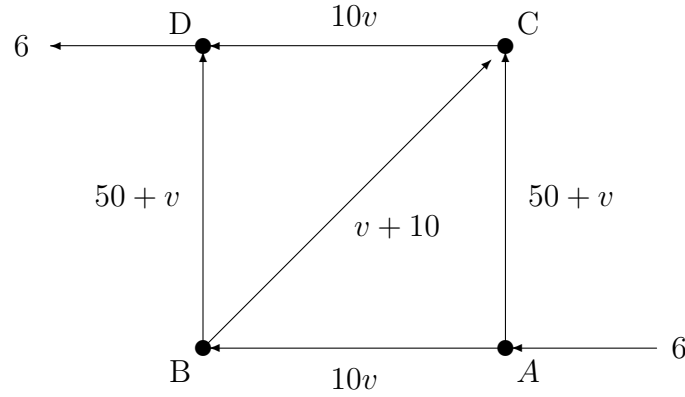


Figure 1: LE RÉSEAU DU PARADOXE DE BRAESS

chemin	flot	fonctions de délai
ACD	x_1	$F_1(x) = [50 + x_1] + [10(x_1 + x_3)] = 11x_1 + 10x_3 + 50$
ABD	x_2	$F_2(x) = [50 + x_2] + [10(x_2 + x_3)] = 11x_2 + 10x_3 + 50$
ABCD	x_3	$F_3(x) = [10(x_2 + x_3)] + [x_3 + 10] + [10(x_1 + x_3)]$ $= 10x_1 + 10x_2 + 21x_3 + 10$

Table 1: LES FONCTIONS DE DÉLAIS SUR LES CHEMINS

2.1 Caractérisation géométrique

Soit $N_X(x)$ le CÔNE NORMAL à l'ensemble X au point $x \in X$, défini par :

$$N_X(x) = \{d \in \mathbb{R}^n : \langle d, y - x \rangle \leq 0 \quad \forall y \in X\}. \quad (6)$$

Le cône normal définit l'ensemble des directions formant un angle 'obtus' avec l'ensemble $X - x = \{y - x : y \in X\}$. On a alors que (5) est équivalent à (voir figure 2) :

$$-F(x^*) \in N_X(x^*). \quad (7)$$

L'expression précédente, qui généralise la notion habituelle de système d'équations non linéaires, a été nommée ÉQUATION GÉNÉRALISÉE par Robinson [Robin79]. Si la fonction F est le gradient d'une fonction f , alors j'obtiens la condition nécessaire d'optimalité :

$$-f'(x^*) \in N_X(x^*), \quad (8)$$

qui exprime qu'au point x^* , il n'existe pas de direction de descente admissible d , c'est-à-dire : $f'(x^*; d)$ n'est jamais négatif.

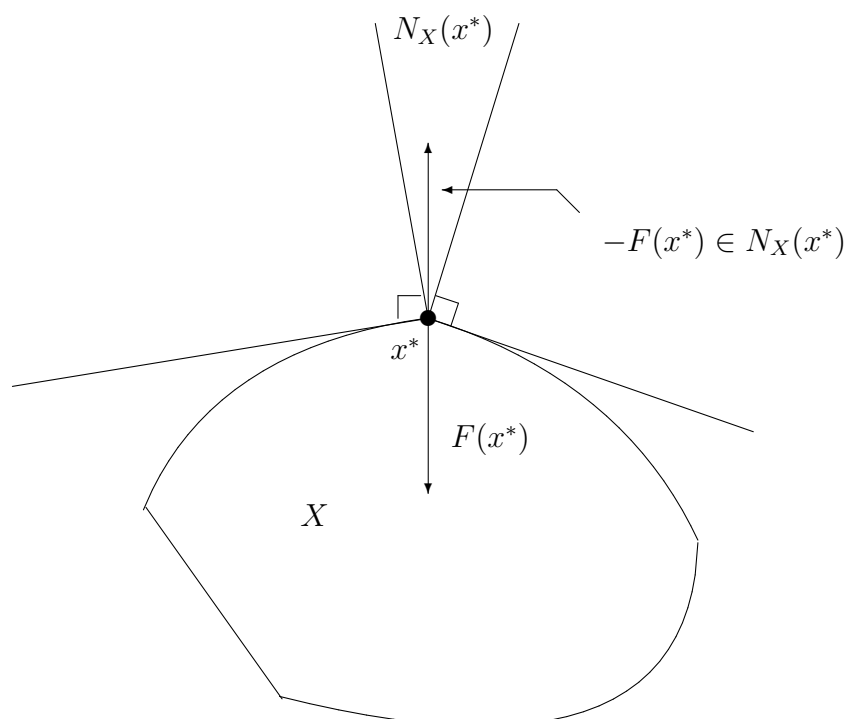


Figure 2: REPRÉSENTATION GÉOMÉTRIQUE

2.2 Formulation de point fixe

Définissons l'ensemble :

$$\Gamma(x) = \arg \max_{y \in X} \langle F(x), x - y \rangle. \quad (9)$$

Il est facile de voir que x est solution de $IV(F, X)$ si et seulement si :

$$x \in \Gamma(x). \quad (10)$$

Considérons maintenant le problème d'optimisation convexe :

$$\min_{p \in X} \frac{1}{2} \|p - (x - \alpha F(x))\|^2, \quad (11)$$

dont la solution unique $p_\alpha(x) = \text{proj}_X(x - \alpha F(x))$ est la projection suivant la norme euclidienne du point $x - \alpha F(x)$ sur l'ensemble X (voir figure 3). Les conditions d'optimalité du problème (11) sont :

$$\langle p_\alpha(x) - (x - \alpha F(x)), p_\alpha(x) - y \rangle \leq 0 \quad \forall y \in X. \quad (12)$$

Un point x est solution de $IV(F, X)$ si et seulement si $x = p_\alpha(x)$. La première formulation de point fixe considérée correspond au cas limite $\alpha = +\infty$, en quelque sorte. On peut généraliser cette formulation en adoptant le produit scalaire et la

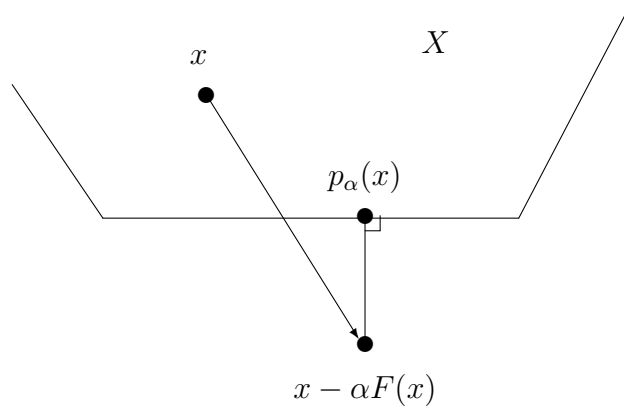


Figure 3: PROJECTION SUR L'ENSEMBLE DÉFINI PAR LES CONTRAINTES

norme induits par une matrice symétrique et définie positive K : $\langle x, y \rangle_K = x^t K y$, $\|x\|_K^2 = x^t K x$. On obtient alors le problème d'optimisation :

$$\min_{p \in X} \frac{1}{2} \|p - (x - \alpha K^{-1} F(x))\|_K^2, \quad (13)$$

qui est identique à :

$$\min_{p \in X} \frac{1}{2} \langle p - x + \alpha K^{-1} F(x), K(p - x + \alpha K^{-1} F(x)) \rangle \quad (14)$$

et dont la solution $p_\alpha^K(x)$ satisfait l'inéquation variationnelle :

$$\langle K(p_\alpha^K(x) - x) + \alpha F(x), p_\alpha^K(x) - y \rangle \leq 0 \quad \forall y \in X. \quad (15)$$

et résoud le problème d'optimisation :

$$\max_{p \in X} \langle F(x), x - p \rangle - \frac{1}{2\alpha} \|x - p\|_K^2, \quad (16)$$

dont la valeur optimale est notée $g_\alpha^K(x)$, et simplement $g_\alpha(x)$ si $K = I$. Comme auparavant, x est solution de l'inégalité variationnelle si et seulement si $x = p_\alpha^K(x)$.

Définissons maintenant, de façon plus générale, une approximation $\tilde{F}(z, x)$ de la fonction $F(z, x)$ au point x telle que :

$$\tilde{F}(x, x) = F(x). \quad (17)$$

pour x fixé et soit $\tilde{\Gamma}(x)$ l'ensemble des solutions de $\text{IV}(\tilde{F}, X)$. Il est clair que x est solution de $\text{IV}(F, X)$ si et seulement si $x \in \tilde{\Gamma}(x)$. En posant $\tilde{F}(z, x) = F(x)$ et $\tilde{F}(z, x) = F(x) + \frac{1}{\alpha} K(z - x)$ je retrouve les formulations de point fixe précédentes.

Certaines formes particulières de \tilde{F} ont retenu l'attention des chercheurs :

- $\tilde{F}_i(z, x) = F_i(x_1, \dots, x_{i-1}, z_i, x_{i+1}, \dots, x_n)$
(approximation de JACOBI) (18)

- $\tilde{F}(z, x) = F(x) + D(x)(z - x)$ avec
 $[D(x)]_{ii} = \frac{\partial F_i(x)}{\partial x_i} = (F'(x))_{ii}$
 $[D(x)]_{ij} = 0$ si $i \neq j$
 (approximation de JACOBI LINÉARISÉE) (19)

- $\tilde{F}_i(z, x) = F_i(z_1, \dots, z_{i-1}, z_i, x_{i+1}, \dots, x_n)$
(approximation de GAUSS-SEIDEL) (20)

- $\tilde{F}(z, x) = F(x) + L(x)(z - x)$ avec
 $[L(x)]_{ij} = \frac{\partial F_i(x)}{\partial x_j} = (F'(x))_{ij}$ si $i \geq j$
 $[L(x)]_{ij} = 0$ si $i < j$
 (approximation de GAUSS-SEIDEL LINÉARISÉE) (21)

- $\tilde{F}(z, x) = F(x) + F'(x)(z - x)$
(approximation de NEWTON) (22)

- $\tilde{F}(z, x) = F(x) + Q(x)(z - x)$ où $Q(x)$ est une matrice
(approximation de QUASI-NEWTON) (23)

- $\tilde{F}(z, x) = F(x) + M(z - x)$ où M est une matrice
(PROJECTION ASYMÉTRIQUE, voir Tseng [Tseng91]) (24)

- $\tilde{F}(z, x) = F(x) + \frac{1}{\alpha}(h'(z) - h'(x))$ où h est fortement convexe
(PROBLÈME AUXILIAIRE, voir Cohen [Cohen88]) (25)

L'approximation de Jacobi néglige les effets 'hors-diagonale' alors que l'approximation de Gauss-Seidel ne tient compte que des effets sous-diagonaux. L'approximation de Newton est une approximation de Taylor au premier ordre de la fonction F . La projection asymétrique approxime F par une matrice constante ; si cette matrice est symétrique et définie positive, je retrouve la projection classique suivant la norme induite par la matrice M . Je discuterai des avantages et inconvénients respectifs de ces approximations à la section traitant des algorithmes.

2.3 Formulation primale-duale

Un vecteur x de X est solution de $IV(F, X)$ si et seulement si :

$$x \in \arg \max_{y \in X} \langle F(x), x - y \rangle \quad (26)$$

$$= \arg \min_{y \in X} \langle F(x), y \rangle \quad (27)$$

$$= \Gamma(x). \quad (28)$$

Supposons que l'ensemble X soit compact, que je puisse écrire : $X = \{x : G(x) \leq 0\}$, et que la représentation de X soit régulière (au sens de la qualification des contraintes). Les conditions nécessaires et suffisantes d'optimalité du programme mathématique convexe $\min_{y \in X} \langle F(x), y \rangle$ sont alors :

$$\begin{aligned} G(y) &\leq 0 \\ F(x) + \lambda G'(y) &= 0 \\ \langle \lambda, G(y) \rangle &= 0 \\ \lambda &\geq 0. \end{aligned} \quad (29)$$

En substituant x à y dans le système précédent, j'en déduis la formulation primale-duale de l'inégalité variationnelle :

$$\begin{aligned} G(x) &\leq 0 \\ F(x) + \lambda G'(x) &= 0 \\ \langle \lambda, G(x) \rangle &= 0 \\ \lambda &\geq 0. \end{aligned} \quad (30)$$

En définissant de nouvelles variables :

$$\begin{aligned} x_j^+ &= \max\{0, x_j\} \\ x_j^- &= \max\{0, -x_j\} \\ x &= x^+ - x^- \\ z &= (x^+, x^-, \lambda)^t \end{aligned}$$

et une nouvelle fonction :

$$H(z) = \begin{bmatrix} F(x^+ - x^-) + G'^t(x^+ - x^-)\lambda \\ -F(x^+ - x^-) - G'^t(x^+ - x^-)\lambda \\ -G(x^+ - x^-) \end{bmatrix}$$

j'obtiens la formulation de COMPLÉMENTARITÉ NON LINÉAIRE :

$$\begin{aligned} H(z) &\geq 0 \\ z &\geq 0 \\ \langle H(z), z \rangle &= 0. \end{aligned} \quad (31)$$

2.4 Formulation sous forme d'équation

Il est facile de transformer un problème de complémentarité non linéaire en système d'équations non différentiables. Ainsi, $z \in R^{m+2n}$ est solution de (31) si et seulement si :

$$\sum_{j=1}^{m+2n} [\min\{z_j, H_j(z)\}]^2 = 0 \quad (32)$$

ou, de façon équivalente :

$$\sum_{j=1}^{m+2n} [z_j + H_j(z) - \sqrt{z_j^2 + H_j(z)^2}]^2 = 0 \quad (33)$$

De nombreuses autres formulations sous forme de système d'équations ont été proposées et possèdent grosso modo les mêmes propriétés.

2.5 Formulations d'optimisation

Considérons l'avant-dernière formulation de point fixe :

$$\tilde{F}(z, x) = F(x) + A(x)(z - x)$$

où je suppose que la matrice $A(x)$ est symétrique et définie positive. L'inégalité variationnelle $IV(\tilde{F}, X)$ se ramène alors au problème d'optimisation :

$$\max_{y \in X} \langle F(x), x - y \rangle - \frac{1}{2} \|x - y\|_{A(x)}^2, \quad (34)$$

dont nous dénoterons la valeur optimale par $\tilde{g}(x)$. La fonction \tilde{g} est non négative (il suffit de poser $y = x$) et l'on a que $\tilde{g}(x) = 0$ si et seulement si x est solution de $IV(F, X)$. On peut donc ramener $IV(F, X)$ au problème d'optimisation *globale* :

$$\min_{x \in X} \tilde{g}(x). \quad (35)$$

La FONCTION DE MÉRITE \tilde{g} (parfois appelée 'gap function') est presque toujours non convexe et parfois non différentiable, même si F l'est. Nous réservons la notation g à la fonction de mérite :

$$g(x) = \max_{y \in X} \langle F(x), x - y \rangle. \quad (36)$$

Je rappelle que :

$$g_\alpha(x) = \max_{y \in X} \langle F(x), x - y \rangle - \frac{1}{2\alpha} \|x - y\|^2 \quad (37)$$

$$g_\alpha^K(x) = \max_{y \in X} \langle F(x), x - y \rangle - \frac{1}{2\alpha} \|x - y\|_K^2 \quad (38)$$

Il est possible de transformer $IV(F, X)$ en un problème d'optimisation *sans contraintes* $\min \psi(x)$ en posant :

$$\psi(x) = g_{\alpha_1}(x) - g_{\alpha_2}(x) \quad (39)$$

avec $\alpha_1 > \alpha_2$. On étend la définition des fonctions g_α à tout R^n . Il se peut alors que $g_\alpha(x)$ soit négatif. Par contre, $\psi(x)$ est toujours positif ou nul, et l'on a :

$$\begin{aligned} \psi(x) &\geq \langle F(x), x - p_{\alpha_2}(x) \rangle - \frac{1}{2\alpha_1} \|x - p_{\alpha_2}(x)\|^2 \\ &\quad - \left[\langle F(x), x - p_{\alpha_2}(x) \rangle - \frac{1}{2\alpha_2} \|x - p_{\alpha_2}(x)\|^2 \right] \\ &= \left(\frac{1}{2\alpha_2} - \frac{1}{2\alpha_1} \right) \|x - p_{\alpha_2}(x)\|^2 \\ &= 0 \quad \Leftrightarrow \quad x = p_{\alpha_2}(x). \end{aligned}$$

Si $x \notin X$, on ne peut avoir $x = p_{\alpha_2}(x)$. Donc l'égalité ne peut se produire que si x est solution de l'inégalité variationnelle.

Si F est MONOTONE (voir section suivante), un point x est solution de $IV(F, X)$ si et seulement si il satisfait également :

$$\langle F(y), x - y \rangle \leq 0 \quad \forall y \in X. \quad (40)$$

En définissant la fonction de mérite :

$$\bar{g}(x) = \max_{y \in X} \langle F(y), x - y \rangle \quad (41)$$

j'obtiens la formulation d'optimisation (globale) :

$$\min_{x \in X} \bar{g}(x) \quad (42)$$

ou, de façon équivalente, le problème de programmation linéaire semi-infini, c'est-à-dire impliquant une infinité de contraintes linéaires :

$$\begin{aligned} &\min_{x \in X, z} z \quad (43) \\ \text{sujet à} \quad &z \geq \langle F(y), x - y \rangle \quad \forall y \in X. \end{aligned}$$

3. EXISTENCE ET UNICITÉ

Si la fonction F est continue sur l'ensemble compact X (ce que je supposerai), l'existence d'une solution à l'inégalité variationnelle $IV(F, X)$ nous est garantie par le théorème de Brouwer énoncé en introduction. En effet, la fonction qui associe à $x \in X$ la projection $p(x) \in X$ étant continue, celle-ci doit admettre au moins un point fixe. Or tout point de X satisfaisant l'égalité $p(x) = x$ est solution de l'inégalité variationnelle.

Avant de traiter la question de l'unicité de la solution, nous allons introduire quelques définitions.

Une fonction $F : R^n \rightarrow R^n$ est LIPSCHITZIENNE sur l'ensemble X s'il existe une constante L (CONSTANTE DE LIPSCHITZ) telle que :

$$\|F(x) - F(y)\| \leq L\|x - y\| \quad \forall x, y \in X. \quad (44)$$

Une fonction F est MONOTONE sur l'ensemble convexe X si

$$\langle F(x) - F(y), x - y \rangle \geq 0 \quad \forall x, y \in X. \quad (45)$$

La fonction F est STRICTEMENT MONOTONE si l'inégalité ci-dessus est stricte pour tout $x \neq y$, FORTEMENT MONOTONE de module β s'il existe un nombre positif β tel que :

$$\langle F(x) - F(y), x - y \rangle \geq \beta\|x - y\|^2 \quad \forall x, y \in X \quad (46)$$

et CO-COERCIVE (de module β) s'il existe un nombre positif β tel que :

$$\langle F(x) - F(y), x - y \rangle \geq \beta\|F(x) - F(y)\|^2 \quad \forall x, y \in X. \quad (47)$$

On a trivialement :

$$\text{FORTE MONOTONIE} \Rightarrow \text{STRICTE MONOTONIE} \Rightarrow \text{MONOTONIE.}$$

Quant à la co-coercivité, cette notion, comme la stricte monotonie, est intermédiaire entre la monotonie et la forte monotonie. Cependant, elle n'implique ni n'est impliquée par la stricte monotonie. De fait, une fonction F est co-coercive si et seulement si son inverse F^{-1} (qui peut être une MULTIFONCTION, c'est-à-dire que $F(x)$ est un sous-ensemble de R^n) est fortement monotone et lipschitzienne, au sens des multifonctions.

Sous l'une ou l'autre de ces conditions, l'ensemble des solutions de $IV(F, X)$ est convexe. Si F est strictement monotone, l'ensemble des solutions se réduit à un point unique. D'autres notions dites de MONOTONIE GÉNÉRALISÉE, plus faibles, ont été introduites et étudiées dans [KaSch90] et [KarSC93] entre autres.

Le gradient d'une fonction (strictement, fortement) convexe est (strictement, fortement) monotone ainsi que co-coercif. Si F est monotone et continûment différentiable sur X , alors sa matrice jacobienne $F'(x)$ est semidéfinie positive sur X , c'est-à-dire :

$$\langle F'(x)(x - y), x - y \rangle \geq 0 \quad \forall y \in X. \quad (48)$$

Si F est fortement monotone sur X , alors $F'(x)$ est définie positive sur X . Par contre, je ne peux rien affirmer de particulier si F est strictement monotone.

Je conclus cette section par une preuve de l'équivalence des formes (5) et (40), lorsque la fonction F est monotone. On a en effet :

$$\begin{aligned} \langle F(y), x - y \rangle &= \langle F(x), x - y \rangle - \langle F(y) - F(x), y - x \rangle \\ &\leq \langle F(x), x - y \rangle \end{aligned}$$

si F est monotone sur x . On en déduit que $\langle F(y), x - y \rangle \leq 0$ pour tout $y \in X$, dès que x est solution de (5).

Réciproquement, soit x une solution de (40). Puisque l'inégalité est satisfaite pour tout $y \in X$ et $t \in [0, 1]$, je peux y substituer $x + t(y - x) \in X$ à y pour obtenir :

$$\langle F(x + t(y - x)), x - (x + t(y - x)) \rangle \leq 0$$

ou, en divisant par t :

$$\langle F(x + t(y - x)), x - y \rangle \leq 0.$$

Il suffit maintenant de faire tendre t vers zéro pour conclure, grâce à la continuité de F , que x satisfait $IV(F, X)$.

4. ALGORITHMES

Les processus d'ajustement continus ou discrets menant à un équilibre, basés sur de hypothèses de comportement des agents du système, sont souvent la base d'algorithmes de résolution. Les systèmes dynamiques induits par des processus d'ajustement continus constituent un champ d'étude en eux-mêmes. Le lecteur intéressé consultera l'ouvrage de Aubin et Cellina [**AuCel84**]. Nous nous limiterons à la présentation d'algorithmes itératifs d'un point de vue 'discret'. Il est cependant tout à fait légitime de considérer un algorithme itératif comme une approximation numérique d'un phénomène continu (voir Rosen [**Rosen65**]).

J'ai regroupé les algorithmes en trois catégories : les algorithmes de point fixe et de descente, basés, respectivement, sur les formulations de point fixe et de descente, comme on aurait pu s'en douter, puis les algorithmes d'hyperplans de coupe, basés sur la formulation (40) de l'inéquation variationnelle. Mentionnons que d'autres techniques de résolution, très semblables à celles proposées en optimisation différentiable, ont été considérées, telles que les méthodes de pénalité et de lagrangien augmenté (voir Auslender [**Ausle76**]). Nous ne considérerons pas explicitement les algorithmes basés sur la reformulation de l'inégalité variationnelle en problème de complémentarité non linéaire.

Dans cette section, nous supposons que F est monotone et continûment différentiable sur l'ensemble compact X . Certains des résultats sont valides sous des hypothèses plus faibles que celles mentionnées dans le présent document, le prix à payer étant des démonstrations plus techniques.

4.1 Algorithmes de point fixe

L'algorithme itératif archétypique est l'algorithme de projection, défini par l'équation de récurrence :

$$x^{k+1} = p(x^k) = \text{proj}_X(x^k - \alpha F(x^k)), \quad (49)$$

où α est un paramètre positif qui reste à déterminer. On suppose que F est fortement monotone de module β et lipschitzienne de module L .

Puisque l'opérateur de projection est non expansif, je peux écrire :

$$\begin{aligned}
\|p(x) - p(y)\|^2 &= \|\text{proj}_X(x - \alpha F(x)) - \text{proj}_X(y - \alpha F(y))\|^2 \\
&\leq \|(x - \alpha F(x)) - (y - \alpha F(y))\|^2 \\
&= \|x - y\|^2 - 2\alpha \langle x - y, F(x) - F(y) \rangle + \alpha^2 \|F(x) - F(y)\|^2 \\
&\leq \|x - y\|^2 - 2\alpha\beta \|x - y\|^2 + \alpha^2 L^2 \|x - y\|^2 \\
&= (1 - 2\alpha\beta + \alpha^2 L^2) \|x - y\|^2.
\end{aligned}$$

Dès que $\alpha < 2\beta/L^2$, l'opérateur p est contractant. Par exemple, si $\alpha = \beta/L^2$, j'obtiens :

$$\|p(x) - p(y)\| \leq \rho \|x - y\|,$$

où $\rho^2 = 1 - \beta^2/L^2$ est inférieur à 1. On en conclut que la suite $\{x^k\}$ converge géométriquement vers l'unique solution x^* de l'inégalité variationnelle, c'est-à-dire :

$$\|x^k - x^*\| \leq \rho \|x^{k-1} - x^*\| \leq \rho^k \|x^0 - x^*\|.$$

Le même raisonnement s'applique à l'algorithme de projection utilisant une norme définie par une matrice symétrique et définie positive K :

$$p_K(x) = \text{proj}_X^K(x - \alpha K^{-1}F(x)).$$

En effet :

$$\begin{aligned}
\|p_K(x) - p_K(y)\|_K^2 &= \|\text{proj}_X^K(x - \alpha K^{-1}F(x)) - \text{proj}_X^K(y - \alpha K^{-1}F(y))\|^2 \\
&\leq \|(x - \alpha K^{-1}F(x)) - (y - \alpha K^{-1}F(y))\|_K^2 \\
&= \|x - y\|_K^2 - 2\alpha \langle x - y, K^{-1}(F(x) - F(y)) \rangle_K \\
&\quad + \alpha^2 \|K^{-1}(F(x) - F(y))\|_K^2 \\
&\leq \|x - y\|_K^2 - 2\alpha\beta \|x - y\|^2 + \alpha^2 L^2 \|x - y\|_{K^{-1}}^2 \\
&\leq \|x - y\|_K^2 \left(1 - \frac{2\alpha\beta}{\lambda_{\max}(K)} + \frac{\alpha^2 L^2}{\lambda_{\min}^2(K)}\right),
\end{aligned}$$

où $\lambda_{\min}(K)$ désigne la plus petite valeur propre de la matrice K , et $\lambda_{\max}(K)$ la plus grande. On voit que, pour α suffisamment faible :

$$\alpha < \frac{2\beta}{L^2} \frac{\lambda_{\min}^2(K)}{\lambda_{\max}(K)}$$

l'opérateur p_K est contractant.

4.2 Algorithmes d'approximation

Soit $\tilde{F}(x, y)$ une approximation de la fonction F , telle que définie à la section précédente. A l'itération $k + 1$, x^{k+1} satisfait l'inégalité variationnelle :

$$\langle \tilde{F}(x^{k+1}, x^k), x^{k+1} - y \rangle \leq 0 \quad \forall y \in X. \quad (50)$$

Afin d'obtenir un algorithme bien défini, et également pour des raisons d'ordre théorique, on choisit de préférence des fonctions \tilde{F} qui soient strictement monotones en leur première variable. Notons que l'algorithme basé sur l'approximation de Newton (22) possède un taux de convergence quadratique, sous des hypothèses classiques. Une étude de ce schéma itératif se trouve dans [Dafer83].

Je vais maintenant analyser la méthode du problème auxiliaire résultant de l'approximation :

$$\tilde{F}(z, x) = F(x) + \frac{1}{\alpha}(h'(z) - h'(x)),$$

ou, de façon équivalente :

$$\tilde{F}(z, x) = \alpha F(x) + h'(z) - h'(x).$$

où F est co-coercive de coefficient γ , h est une fonction fortement convexe, de coefficient de forte convexité b et son gradient est lipschitzien de coefficient B . L'on fera appel aux résultats classiques suivants :

$$h(y) - h(x) \geq \langle h'(x), y - x \rangle + \frac{1}{2}b\|y - x\|^2 \quad \forall x, y \in X. \quad (51)$$

$$h(y) - h(x) \leq \langle h'(x), y - x \rangle + \frac{1}{2}B\|y - x\|^2 \quad \forall x, y \in X. \quad (52)$$

A l'itération $k + 1$ on a :

$$\langle \alpha F(x^k) + h'(x^{k+1}) - h'(x^k), x^{k+1} - y \rangle \leq 0 \quad \forall y \in X. \quad (53)$$

Soit x^* une solution arbitraire de l'inégalité variationnelle. Je pose :

$$\Lambda(x) = h(x^*) - h(x) - \langle h'(x), x^* - x \rangle.$$

Puisque h est fortement convexe, on a :

$$\Lambda(x) \geq \frac{b}{2}\|x - x^*\|^2. \quad (54)$$

Maintenant :

$$\begin{aligned} \Lambda(x^k) - \Lambda(x^{k+1}) &= h(x^{k+1}) - h(x^k) - \langle h'(x^k), x^{k+1} - x^k \rangle \\ &\quad + \langle h'(x^{k+1}) - h'(x^k), x^* - x^{k+1} \rangle \\ &\geq \frac{b}{2}\|x^k - x^{k+1}\|^2 \\ &\quad + \alpha \langle F(x^k), x^{k+1} - x^* \rangle, \end{aligned} \quad (55)$$

où le deuxième terme de l'inégalité est obtenu en posant $y = x^*$ dans l'équation (53). Puisque

$$\langle -F(x^*), x^{k+1} - x^* \rangle \leq 0,$$

on obtient, à partir de (55) :

$$\begin{aligned}\Lambda(x^k) - \Lambda(x^{k+1}) &\geq \frac{b}{2}\|x^k - x^{k+1}\|^2 + \alpha\langle F(x^k) - F(x^*), x^{k+1} - x^* \rangle \\ &= \frac{b}{2}\|x^k - x^{k+1}\|^2 + T.\end{aligned}$$

Or :

$$\begin{aligned}T &= \alpha\langle F(x^k) - F(x^*), x^{k+1} - x^* \rangle \\ &= \alpha\langle F(x^k) - F(x^*), x^k - x^* \rangle + \alpha\langle F(x^k) - F(x^*), x^{k+1} - x^k \rangle.\end{aligned}$$

Comme F est co-coercive de coefficient γ , il vient :

$$\begin{aligned}T &\geq \alpha[\gamma\|F(x^k) - F(x^*)\|^2 + \langle F(x^k) - F(x^*), x^{k+1} - x^k \rangle] \\ &\geq -\frac{\alpha}{4\gamma}\|x^{k+1} - x^k\|^2.\end{aligned}$$

Donc :

$$\Lambda(x^k) - \Lambda(x^{k+1}) \geq \frac{1}{2} \left(b - \frac{\alpha}{2\gamma} \right) \|x^{k+1} - x^k\|^2.$$

Soit $\alpha < 2b\gamma$. Si $x^{k+1} = x^k$, x^k est solution de l'inégalité variationnelle. Sinon, la suite $\{\Lambda(x^k) - \Lambda(x^{k+1})\}_k$ est non négative et :

$$\lim_{k \rightarrow \infty} \|x^{k+1} - x^k\| = 0.$$

De plus, puisque $\|x^k - x^*\|^2 \leq (2/b)\Lambda(x^k)$ et que la suite $\{\Lambda(x^k)\}$ est décroissante, je peux affirmer que la suite $\{x^k\}$ est bornée. Soit \bar{x} un point d'accumulation de la suite $\{x^k\}$. Je passe à la limite dans (50) et conclut que \bar{x} est solution de l'inégalité variationnelle.

L'analyse ci-dessus demeure valide en remplaçant x^* par \bar{x} et Λ par

$$\bar{\Lambda}(x) = h(\bar{x}) - h(x) - \langle h'(x), \bar{x} - x \rangle.$$

La suite $\{\bar{\Lambda}(x^k)\}$ est strictement décroissante et, par (52), on a :

$$\bar{\Lambda}(x^k) - \bar{\Lambda}(\bar{x}) = \bar{\Lambda}(x^k) \leq \frac{1}{2}B\|\bar{x} - x^k\|^2.$$

par conséquent, la suite $\bar{\Lambda}(x^k)$ converge vers zéro. Ce résultat, combiné à l'inégalité

$$\bar{\Lambda}(x^k) \geq \frac{b}{2}\|x^k - \bar{x}\|^2,$$

me permet de conclure que toute la suite $\{x^k\}$ converge vers \bar{x} .

4.3 L'algorithme de l'extragradient

Le principal défaut des méthodes précédentes est de ne converger que sous des hypothèses restrictives, incluant la forte monotonie ou la co-coercivité de la fonction

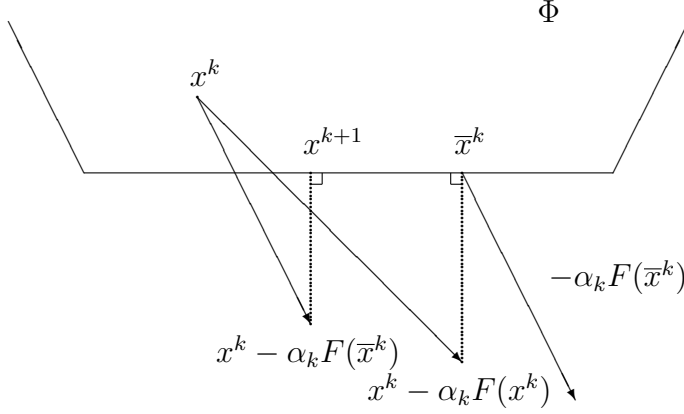


Figure 4: L'EXTRAGRADIANT

F . Je présente maintenant une méthode qui converge sous monotonie simple de F (la solution n'est pas nécessairement unique) et n'exige pas que l'ensemble X soit compact, mais simplement que l'ensemble X^* des solutions soit non vide. La preuve de convergence ne repose plus sur un argument de contraction, mais sur le fait que la distance $\text{dist}(x^k, X^*)$ entre le point courant et l'ensemble des solutions décroît à chaque itération.

L'algorithme de base a été proposé, sous le nom d'ALGORITHME DE L'EXTRAGRADIANT par Korpelevitch (voir [BerTs89]), et fait appel à deux projections à chaque itération :

$$\bar{x}^k = \text{proj}_X(x^k - \alpha_k F(x^k)) \quad (56)$$

$$x^{k+1} = \text{proj}_X(x^k - \alpha_k F(\bar{x}^k)). \quad (57)$$

Soient u et v ($v \in X$) deux vecteurs quelconques de R^n . On a :

$$\|u - v\|^2 \geq \|u - \text{proj}_X(u)\|^2 + \|v - \text{proj}_X(u)\|^2,$$

ou, de façon équivalente :

$$\|\text{proj}_X(u) - v\|^2 \leq \|u - v\|^2 - \|u - \text{proj}_X(u)\|^2.$$

Je pose $u = x^k - \alpha_k F(\bar{x}^k)$, $v = x^*$ ($x^* \in X^*$) et $x^{k+1} = \text{proj}_X(u)$ dans l'inégalité précédente et je réarrange les termes :

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &\leq \|x^k - \alpha_k F(\bar{x}^k) - x^*\|^2 - \|x^k - \alpha_k F(\bar{x}^k) - x^{k+1}\|^2 \\ &= \|x^k - x^*\|^2 - \|x^k - x^{k+1}\|^2 + 2\alpha_k \langle x^* - x^{k+1}, F(\bar{x}^k) \rangle. \end{aligned} \quad (58)$$

On obtient :

$$\begin{aligned} \langle x^* - x^{k+1}, F(\bar{x}^k) \rangle &= \langle x^* - \bar{x}^k, F(\bar{x}^k) \rangle + \langle \bar{x}^k - x^{k+1}, F(\bar{x}^k) \rangle \\ &\leq \langle \bar{x}^k - x^{k+1}, F(\bar{x}^k) \rangle, \end{aligned} \quad (59)$$

puisque x^* est solution de l'inégalité variationnelle et F est monotone. Il s'ensuit, en substituant (59) dans (58) :

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 - \|x^k - x^{k+1}\|^2 + 2\alpha_k \langle \bar{x}^k - x^{k+1}, F(\bar{x}^k) \rangle \\ &= \|x^k - x^*\|^2 - \|x^k - \bar{x}^k\|^2 - \|\bar{x}^k - x^{k+1}\|^2 - 2\langle x^k - \bar{x}^k, \bar{x}^k - x^{k+1} \rangle \\ &\quad + 2\alpha_k \langle \bar{x}^k - x^{k+1}, F(\bar{x}^k) \rangle \\ &= \|x^k - x^*\|^2 - \|x^k - \bar{x}^k\|^2 - \|\bar{x}^k - x^{k+1}\|^2 \\ &\quad + 2\langle x^k - \alpha_k F(\bar{x}^k) - \bar{x}^k, x^{k+1} - \bar{x}^k \rangle \\ &= \|x^k - x^*\|^2 - \|x^k - \bar{x}^k\|^2 - \|\bar{x}^k - x^{k+1}\|^2 \\ &\quad + 2\langle x^k - \alpha_k F(x^k) - \bar{x}^k, x^{k+1} - \bar{x}^k \rangle \\ &\quad + 2\alpha_k \langle F(x^k) - F(\bar{x}^k), x^{k+1} - \bar{x}^k \rangle \\ &\leq \|x^k - x^*\|^2 - \|x^k - \bar{x}^k\|^2 - \|\bar{x}^k - x^{k+1}\|^2 \\ &\quad + 2\alpha_k \|F(x^k) - F(\bar{x}^k)\| \cdot \|x^{k+1} - \bar{x}^k\| \end{aligned} \quad (60)$$

puisque $\langle \text{proj}_X(u) - u, \text{proj}_X(u) - v \rangle \leq 0$ si $v \in X$. En utilisant l'inégalité :

$$\|x^{k+1} - \bar{x}^k\|^2 + \alpha_k^2 \|F(x^k) - F(\bar{x}^k)\|^2 \geq 2\alpha_k \|F(x^k) - F(\bar{x}^k)\| \cdot \|x^{k+1} - \bar{x}^k\|$$

dans (60) il vient :

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 - \|x^k - \bar{x}^k\|^2 - \|\bar{x}^k - x^{k+1}\|^2 \\ &\quad + \|x^{k+1} - \bar{x}^k\|^2 + \alpha_k^2 \|F(x^k) - F(\bar{x}^k)\|^2 \\ &\leq \|x^k - x^*\|^2 - \|x^k - \bar{x}^k\|^2 \left\{ 1 - \alpha_k^2 \frac{\|F(x^k) - F(\bar{x}^k)\|^2}{\|x^k - \bar{x}^k\|^2} \right\}. \end{aligned}$$

Si F est lipschitzienne de constante L , il suffit de choisir $\alpha_k < 1/L$ pour avoir décroissance de la distance $\|x^k - x^*\|$, à moins que x^k ne coïncide avec \bar{x}^k , mais alors x^k est solution de l'inégalité variationnelle. Comme le choix de x^* est arbitraire, j'en déduis également que la distance entre le point courant x^k et *tout* point de l'ensemble des solutions décroît à chaque itération. En particulier, si l'ensemble des solutions est une variété affine de R^n , la suite $\{x^k\}$ converge vers la projection de x^0 sur l'ensemble des solutions. On remarque que le paramètre α_k peut être ajusté à chaque itération, en fonction de la constante de Lipschitz dans la direction $\bar{x}^k - x^k$. Ainsi la constante de Lipschitz n'a pas besoin d'être connue explicitement. Malheureusement, le taux de convergence de la méthode n'est pas terrible. Appliquée à un problème d'optimisation, l'algorithme ne fait que des pas timides vers une solution.

Une famille d'algorithmes plus générale, basée sur le même principe, a par la suite été proposée par Konnov [Konno93]. On définit :

$$\begin{aligned} p &= \text{proj}_X(x - \lambda F(x)) & \lambda &> 0 \\ y &= x + \theta(p - x) & \alpha &> 0 \quad (\text{à déterminer}) \\ x^+ &= \text{proj}_X(x - \alpha F(y)) & 0 &< \theta < \frac{1}{\lambda(1+L)}. \end{aligned}$$

En posant $\alpha = \lambda$ et $\theta = 1$ on retrouve la méthode de l'extragradiant. On écrit :

$$\begin{aligned}\|p - x\|^2 &= \langle p - (x - \lambda F(x)), p - x \rangle - \lambda \langle F(x), p - x \rangle \\ &\leq \lambda \langle F(x), x - p \rangle,\end{aligned}$$

d'où l'on tire :

$$\begin{aligned}\langle F(y), x - p \rangle &= \langle F(x), x - p \rangle + \langle F(y) - F(x), x - p \rangle \\ &\geq \frac{\|p - x\|^2}{\lambda} - L\theta \|p - x\|^2 \\ &= \|p - x\|^2 \left(\frac{1}{\lambda} - L\theta \right) \\ &\geq \theta \|p - x\|^2 \quad \text{puisque } \theta < 1/\lambda(1 + L) \\ &> 0 \quad \text{si } x \notin X^*\end{aligned}\tag{61}$$

et :

$$\begin{aligned}\langle F(y), x - x^* \rangle &= \langle F(y), y - x^* \rangle + \langle F(y), x - y \rangle \\ &\geq \langle F(y), x - y \rangle \\ &= \theta \langle F(y), x - p \rangle \\ &\geq \theta^2 \|p - x\|^2 \quad \text{par (61)} \\ &> 0 \quad \text{si } x \notin X^*.\end{aligned}\tag{62}$$

Maintenant :

$$\begin{aligned}\|x^+ - x^*\|^2 &\leq \|x - \alpha F(y) - x^*\|^2 \quad \text{puisque 'proj' est non expansif} \\ &= \|x - x^*\|^2 - 2\alpha \langle F(y), x - x^* \rangle + \alpha^2 \|F(y)\|^2 \\ &\leq \|x - x^*\|^2 - 2\alpha \langle F(y), x - y \rangle + \alpha^2 \|F(y)\|^2.\end{aligned}$$

Si $F(y) = 0$, on a gagné (y est solution de l'inégalité variationnelle). Sinon on obtient, en posant $\alpha = \langle F(y), x - y \rangle / \|F(y)\|^2$:

$$\|x^+ - x^*\|^2 \leq \|x - x^*\|^2 - \frac{\langle F(y), x - y \rangle^2}{\|F(y)\|^2}.\tag{63}$$

Considérons une suite $\{x^k\}$ produite par ce schéma itératif, et les suites $\{\alpha_k\}$ et $\{p^k\}$ correspondantes. Si $F(y^k)$ tend vers zéro, on démontre facilement (la suite $\{x^k\}$ est bornée) que tout point d'accumulation \bar{y} de la suite $\{y^k\}$ est solution de l'inégalité variationnelle, et qu'en fait toute la suite $\{x^k\}$ converge vers \bar{y} . Si $F(y^k)$ ne tend pas vers zéro, soient \bar{x} et \bar{p} des points d'accumulation des suites $\{x^k\}$ et $\{p^k\}$, respectivement. Puisque $\langle F(y^k), x^k - y^k \rangle$ doit tendre vers zéro (voir (63)), on obtient :

$$\langle F(y^k), x^k - y^k \rangle \geq \theta^2 \|p^k - x^k\|^2 \rightarrow 0,$$

d'où, par continuité : $\|\bar{p} - \bar{x}\| = 0$, c'est-à-dire que \bar{x} est solution de l'inégalité variationnelle. Comme auparavant, il est facile de démontrer que *toute* la suite converge vers \bar{x} .

4.4 Les algorithmes d'optimisation

Nous allons considérer des algorithmes de minimisation de la fonction \tilde{g} définie par (34). Cette fonction est en général non convexe. Il semble assez naturel d'utiliser comme direction de descente $d^k = y^k - x^k$ où $y^k \in \tilde{\Gamma}(x^k)$. Etudions d'abord la fonction g dont l'évaluation $g(x)$ au point x n'exige que la résolution du programme linéaire :

$$\max_{y \in X} \langle F(x), x - y \rangle.$$

Notons que g est convexe dans le cas particulier où F est affine. La dérivée directionnelle d'une fonction 'max' nous est donnée par la règle de Danskin :

$$\begin{aligned} g'(x; \bar{y} - x) &= \max_{y \in \Gamma(x)} \langle F(x), \bar{y} - x \rangle + \langle F'^t(x)(x - y), \bar{y} - x \rangle \\ &= -g(x) - \min_{y \in \Gamma(x)} \langle y - x, F'(x)(\bar{y} - x) \rangle. \end{aligned} \quad (64)$$

On remarque que si $\Gamma(x)$ est un singleton, la dérivée directionnelle est inférieure ou égale à $-g(x)$. Ce sera le cas si l'ensemble X est **FORTEMENT CONVEXE**, c'est-à-dire si, pour tout couple de points x, y dans X , le segment $[x, y]$ est contenu dans l'intérieur de l'ensemble X , à l'exception possible des extrémités x et y . Un ensemble fortement convexe ne comporte pas de faces. Un polyèdre ne peut être fortement convexe. Sous l'hypothèse de forte convexité de X , g est différentiable et l'on démontre facilement la convergence d'un algorithme de descente basé sur l'utilisation de la direction de descente $d = \bar{y} - x$ (voir [Ausle76]).

Dans le cas général, on démontre l'existence d'un élément $\bar{y} \in \Gamma(x)$ tel que $g'(x; \bar{y} - x) \leq -g(x)$. Je présente une preuve constructive de ce résultat, dans le cas où X est polyédral (le cas général est traité dans [Marco85]). Soit $\{y^p\}_{p \in P}$ l'ensemble des points extrémaux de $\Gamma(x)$. On pose :

$$\begin{aligned} \bar{y} &= \sum_{p \in P} \lambda_p y^p & (\sum_{p \in P} \lambda_p = 1 \quad \lambda_p \geq 0) \\ d^p &= y^p - x \\ \bar{d} &= \sum_{p \in P} \lambda_p d^p \end{aligned}$$

et $A = F'(x)$, ce qui permet d'écrire :

$$\begin{aligned} g'(x; \bar{d}) &= -g(x) - \min_{p \in P} \langle d^p, A\bar{d} \rangle \\ &= -g(x) - \min_{p \in P} \sum_{q \in P} \lambda_q \langle d^p, Ad^q \rangle \\ &= -g(x) - \min_{p \in P} \sum_{q \in P} b_{pq} \lambda_q \\ &= -g(x) - \min_{p \in P} (B\lambda)_p, \end{aligned}$$

où B est la matrice semidéfinie positive dont les éléments sont les $b_{pq} = \langle d^p, Ad^q \rangle$. Pour obtenir une direction de descente, il suffit donc de choisir le vecteur de coefficients λ

de telle sorte que le vecteur $B\lambda$ ne comporte que des composantes positives ou nulles. Considérons les programmes linéaires duaux (e représente un vecteur dont toutes les composantes sont égales à 1) :

$$\begin{array}{ll} \max_{\lambda} & e\lambda \\ & B\lambda \geq 0 \\ & \lambda \geq 0 \end{array} \qquad \begin{array}{ll} \min_{\mu} & 0 \\ & -\mu B \geq e \\ & \mu \geq 0 \end{array} .$$

L'ensemble réalisable du problème dual (à droite) est vide. En effet, μ ne pouvant être nul, j'obtiens, si cet ensemble est non vide :

$$-\mu B\mu \geq e\mu > 0,$$

en contradiction avec le fait que la matrice B^t est semidéfinie positive. Par dualité linéaire, j'en déduis que le primal (dont l'ensemble réalisable est trivialement non vide) est non borné. Par normalisation, on en déduit l'existence du vecteur λ recherché.

En pratique, on ne connaît pas explicitement l'ensemble des points extrémaux de $\Gamma(x)$. Il est cependant possible de construire une direction de descente pour g en faisant appel à l'ensemble perturbé :

$$\Gamma_{\epsilon}(x) = \{y : \langle F(x), x - y \rangle \geq g(x) - \epsilon\}.$$

L'idée générale est la suivante : on calcule un élément quelconque y^1 de $\Gamma(x)$ et on 'teste' la direction $d = y^1 - x$ en s'y engageant timidement. Si $y^1 \notin \Gamma(x + \delta(y^1 - x))$ pour δ 'petit', on enrichit $\Gamma_{\epsilon}(x)$ d'un élément quelconque y^2 de $\Gamma(x + \delta(y^1 - x))$. On détermine alors une 'meilleure' direction de descente obtenue par combinaison linéaire appropriée des directions $d^1 = y^1 - x$ et $d^2 = y^2 - x$ et on tente derechef un pas timide, etc. On démontre qu'après un nombre fini k de tentatives infructueuses, on trouve une direction de descente d^k 'acceptable'. Le choix du paramètre δ dépend évidemment du choix de ϵ et l'algorithme converge en temps fini vers une solution ϵ -optimale. Cette technique est très semblable à l'algorithme de descente en optimisation non différentiable convexe où chaque tentative infructueuse de descente mène à l'enrichissement du ϵ -sous-différentiel de la fonction à minimiser, et éventuellement à la détermination d'une direction de descente (voir [LeHir93]). Le choix du paramètre δ dépend évidemment du choix de ϵ et l'algorithme converge en temps fini vers une solution ϵ -optimale.

Si l'ensemble X n'est pas compact, il faut s'attendre à ce que le problème linéaire $\min_{y \in X} \langle F(x), y \rangle$ n'admette pas de solution bornée. Il est alors plus approprié de minimiser la fonction g_{α}^K . Considérons, de façon naturelle, la direction :

$$d = p_{\alpha}^K(x) - x \tag{65}$$

et soit $x_t = x + t(p_{\alpha}^K(x) - x)$ un point du segment $[x, p_{\alpha}^K(x)]$. On définit la fonction r :

$$r(u) = \min_{z \in X} \frac{1}{2} \|z - u\|_K^2 \tag{66}$$

$$= \frac{1}{2} \|\text{proj}_X^K(u) - u\|_K^2. \tag{67}$$

La fonction r est convexe et continûment différentiable sur X , de gradient :

$$r'(u) = K(u - \text{proj}_X^K(u)),$$

et satisfait l'inégalité des fonction convexes :

$$r(v) - r(u) \geq \langle v - u, K(u - \text{proj}_X^K(u)) \rangle.$$

J'applique l'inégalité précédente à la fonction

$$\xi(x) = \frac{1}{2} \|p_\alpha^K(x) - [x - \alpha K^{-1}F(x)]\|_K^2$$

pour obtenir :

$$\xi(y) - \xi(x) \geq \langle y - x - \alpha K^{-1}F(y) + \alpha K^{-1}F(x), K[x - \alpha K^{-1}F(x) - p_\alpha^K(x)] \rangle.$$

Puisque

$$g_\alpha^K(x) = \frac{\alpha}{2} \|F(x)\|_{K^{-1}}^2 - \frac{1}{\alpha} \xi(x)$$

on en déduit :

$$\begin{aligned} g_\alpha^K(y) - g_\alpha^K(x) &\leq \frac{\alpha}{2} \|F(y)\|_{K^{-1}}^2 - \frac{\alpha}{2} \|F(x)\|_{K^{-1}}^2 + \langle F(x), y - x \rangle \\ &\quad + \langle F(y) - F(x), x - p_\alpha^K(x) \rangle - \frac{1}{\alpha} \langle y - x, K(x - p_\alpha^K(x)) \rangle \\ &\quad + \alpha \langle F(x) - F(y), K^{-1}F(x) \rangle \\ &= \frac{\alpha}{2} \|F(y) - F(x)\|_{K^{-1}}^2 + \langle F(x), y - x \rangle \\ &\quad + \langle F(y) - F(x), x - p_\alpha^K(x) \rangle - \frac{1}{\alpha} \langle y - x, K(x - p_\alpha^K(x)) \rangle. \end{aligned}$$

Je remplace maintenant y par x_t dans l'inégalité ci-dessus, ce qui donne, après avoir divisé par t :

$$\begin{aligned} \frac{1}{t} (g_\alpha^K(x_t) - g_\alpha^K(x)) &\leq \left[\frac{\alpha}{2t} \|F(x_t) - F(x)\|_{K^{-1}}^2 - \frac{1}{t^2} \langle F(x_t) - F(x), x_t - x \rangle \right] \\ &\quad + \left[\langle F(x), p_\alpha^K(x) - x \rangle + \frac{1}{\alpha} \|x - p_\alpha^K(x)\|_K^2 \right] \\ &= T_1 + T_2. \end{aligned}$$

Si F est fortement monotone de coefficient β et lipschitzienne de constante L sur X , je peut écrire :

$$T_1 \leq \left(\frac{\alpha L^2 \|K^{-1}\|_2^2}{2} t - \beta \right) \|x - p_\alpha^K(x)\|_K^2.$$

D'autre part, en posant $y = x$ dans (15) et en réarrangeant les termes, j'obtiens :

$$\langle F(x), x - p_\alpha^K(x) \rangle \geq (1/\alpha) \|x - p_\alpha^K(x)\|_K^2,$$

ce qui implique que T_2 est négatif ou nul. Je déduis finalement :

$$(g_\alpha^K)'(x; p_\alpha^K(x) - x) = \lim_{t \rightarrow 0} \frac{1}{t} (g_\alpha^K(x_t) - g_\alpha^K(x)) \quad (68)$$

$$\leq -\beta \|x - p_\alpha^K(x)\|_2^2. \quad (69)$$

La direction d est donc une direction de descente admissible pour la fonction g_α^K au point x , et l'on peut construire un algorithme convergent autour de cette direction, pourvu que F soit fortement monotone. Si F n'est que monotone, mais que l'ensemble X est compact, on peut démontrer que d est toujours une direction de descente, pourvu que α soit suffisamment grand. Plus précisément, d sera une direction de descente dès que les ensembles de contraintes actives en x et $p_\alpha^K(x)$ sont distincts. Si X est polyédral, ceci revient à affirmer que x et $p_\alpha^K(x)$ se trouvent sur des faces différentes de X (voir [ZhuMa93]).

Une autre direction d'intérêt est la DIRECTION DE NEWTON $d_N(x) = w_N(x) - x$ où $w_N(x)$ est solution de l'inégalité variationnelle linéarisée :

$$\langle F(x) + F'(x)(w_N(x) - x), w_N(x) - y \rangle \leq 0 \quad \forall y \in X. \quad (70)$$

En posant $x^{k+1} = w_N(x^k)$ on obtient l'algorithme de Newton, dont le taux de convergence local, sous des hypothèses assez faibles, est quadratique. On peut rendre la méthode globalement convergente en notant que $d_N(x)$ est une direction de descente pour la fonction g (voir [MarDu87]) :

$$\begin{aligned} g'(x; d_N(x)) &= \max_{y \in \Gamma(x)} \langle F(x) + F'(x)(x - y), d_N(x) \rangle \\ &= -\langle d_N(x), F'(x)d_N(x) \rangle \\ &\quad + \max_{y \in \Gamma(x)} [\langle F(x) + F'(x)(w_N(x) - x), w_N(x) - y \rangle + \langle F(x), y - x \rangle] \\ &\leq -g(x) \\ &< 0 \end{aligned}$$

si x n'est pas solution de l'inégalité variationnelle.

Lorsque X est polyédral, une méthode qui se rapproche de la méthode de Newton est obtenue en utilisant la formulation primale-duale de la section 2.3. Je pose $X = \{x \in R^n : Bx \leq b\}$. On peut alors transformer la formulation primale-duale (30) en problème de minimisation globale :

$$\begin{aligned} \min_{x \in X, \lambda} & \quad \langle b - Bx, \lambda \rangle \\ \text{sujet à} & \quad F(x) + \lambda B = 0 \\ & \quad \lambda \geq 0, \end{aligned}$$

qu'on peut réécrire :

$$\begin{aligned} \min_{x \in X, \lambda} & \quad \langle F(x), x \rangle + \langle b, \lambda \rangle \\ \text{sujet à} & \quad F(x) + \lambda B = 0 \\ & \quad \lambda \geq 0 \end{aligned} \quad (71)$$

qui devient, en utilisant la dualité linéaire :

$$\min_{x \in X, \lambda} \max_{y \in X} \langle F(x), x - y \rangle. \quad (72)$$

On retrouve ainsi la fonction g ! Marcotte et Dussault [MarDu89] ont proposé un algorithme où une direction de descente pour g est obtenue en linéarisant l'objectif et les contraintes (non convexes !) de (71). La direction obtenue est localement identique à celle de Newton, sous une hypothèse de régularité faible. La méthode peut donc être vue comme une globalisation de l'algorithme de Newton. Si F est affine, ce schéma se réduit à l'algorithme de Frank et Wolfe appliqué à la résolution du programme mathématique convexe (71).

Enfin, je mentionne qu'il est possible de combiner la direction provenant d'une approximation \tilde{F} avec une fonction de descente qui ne lui est pas reliée. Ainsi, comme il a été vu auparavant, la direction de Newton est une direction de descente pour la fonction de mérite g . Ce cadre très général a été étudié dans [ZhuMa94].

Remarque : Dans tout algorithme d'optimisation (en fait dans *tout* algorithme), il est utile de savoir si x^k est proche de l'ensemble des solutions. Une fonction ν , définie sur X et satisfaisant :

$$\text{dist}(x, X^*) \leq \nu(x) \quad \nu(x) = 0 \Leftrightarrow x \in X^* \quad (73)$$

est appelée une BORNE D'ERREUR (voir [LuoTs92] lorsque F est affine). Dans le cas où F est fortement monotone de coefficient β , nous allons démontrer que g et g_α sont des bornes d'erreur.

En ce qui concerne g , ceci est facile. En effet :

$$\begin{aligned} g(x) &= \max_{y \in X} \langle F(x), x - y \rangle \\ &\geq \langle F(x), x - x^* \rangle \\ &= \langle F(x) - F(x^*), x - x^* \rangle + \langle F(x^*), x - x^* \rangle \\ &\geq \beta \|x - x^*\|^2, \end{aligned}$$

ce qui implique :

$$\|x - x^*\| \leq \sqrt{\frac{1}{\beta} g(x)}. \quad (74)$$

Je tente maintenant d'obtenir une relation semblable pour g_α . On a :

$$\begin{aligned} g_\alpha(x) &\geq \langle F(x), x - x^* \rangle - \frac{1}{2\alpha} \|x - x^*\|^2 \\ &= \langle F(x) - F(x^*), x - x^* \rangle + \langle F(x^*), x - x^* \rangle - \frac{1}{2\alpha} \|x - x^*\|^2 \\ &\geq \left(\beta - \frac{1}{2\alpha}\right) \|x - x^*\|^2. \end{aligned} \quad (75)$$

Si α est assez grand, par exemple $\alpha = 1/\beta$, on obtient, comme précédemment, une borne d'erreur. Si α est 'petit', on a, pour $t \geq 1$:

$$\begin{aligned} g_{t\alpha}(x) &= \langle F(x), x - p_{t\alpha}(x) \rangle - \frac{1}{2\alpha t} \|x - p_{t\alpha}(x)\|^2 \\ tg_{t\alpha}(x) &= t \left(\langle F(x), x - p_{t\alpha}(x) \rangle - \frac{1}{2\alpha} \|x - p_{t\alpha}(x)\|^2 \right) \\ &\quad + \frac{t-1}{2\alpha} \|x - p_{t\alpha}(x)\|^2 \end{aligned}$$

Or, puisque $\langle p_\alpha(x) - x + \alpha F(x), p_\alpha(x) - x \rangle \leq 0$, on a, comme auparavant :

$$g_\alpha(x) \geq \frac{1}{2\alpha} \|p_\alpha(x) - x\|^2,$$

d'où l'on déduit :

$$\begin{aligned} \frac{1}{2t\alpha} \|x - p_{t\alpha}(x)\|^2 &\leq g_{t\alpha}(x) \\ tg_{t\alpha}(x) &\leq tg_\alpha(x) + (t-1)g_{t\alpha}(x) \\ g_{t\alpha}(x) &\leq tg_\alpha(x) \end{aligned}$$

et finalement :

$$g_\alpha(x) \geq \frac{1}{t} g_{t\alpha}(x). \quad (76)$$

Si $\alpha \geq 1/\beta$, on a, d'après (75) :

$$g_\alpha(x) \geq (\beta/2) \|x - x^*\|^2.$$

Sinon, en posant $t = 1/\alpha\beta$ dans (76) :

$$g_\alpha(x) \geq \alpha\beta g_{\frac{1}{\beta}}(x) \geq \frac{\alpha\beta^2}{2} \|x - x^*\|^2.$$

En combinant les deux résultats on obtient :

$$g_\alpha(x) \geq \frac{\beta}{2} \min\{1, \alpha\beta\} \|x - x^*\|^2$$

et :

$$\|x - x^*\| \leq \sqrt{\frac{1}{\gamma} g_\alpha(x)} \quad (77)$$

avec : $\gamma = (\beta/2) \min(1, \alpha\beta)$.

On note que si $X = R^n$, alors :

$$\begin{aligned} p_\alpha(x) &= x - \alpha F(x) \\ g_\alpha(x) &= \alpha \|F(x)\|^2 \\ g_\alpha(x) &= \frac{1}{t} g_{t\alpha}(x), \end{aligned}$$

plutôt que simplement l'inégalité $g_\alpha(x) \geq (1/t)g_{t\alpha}(x)$.

4.5 Algorithmes de plans coupants

On a vu que, si F est monotone sur X , x est solution de l'inégalité variationnelle si et seulement si :

$$\langle F(y), x - y \rangle \leq 0 \quad \forall y \in X.$$

Si $y \in X$, on peut donc limiter notre recherche d'un point d'équilibre à l'ensemble $X \cap \{x \in R^n : \langle F(y), x - y \rangle \leq 0\}$. Ceci suggère plusieurs stratégies de recherche. Soit par exemple E^k une ellipsoïde contenant au moins une solution de l'inégalité variationnelle, et x^k le centre de E^k . On choisit x^{k+1} comme le centre de l'ellipsoïde de volume minimal contenant $E^k \cap \{x \in R^n : \langle F(x^k), x - x^k \rangle \leq 0\}$. Si F est fortement monotone sur X , la suite $\{x^k\}$ converge vers la solution de l'inégalité variationnelle (voir Lüthi [Lüthi85]).

Une autre méthode, inspirée de l'algorithme de Kelley en optimisation convexe, consiste à résoudre une inégalité variationnelle en ne tenant pas compte des contraintes $G(x) \leq 0$. Si une contrainte G_k est violée à l'itération k , on rajoute la contrainte linéaire :

$$G_k(x^k) + \langle G'_k(x^k), x - x^k \rangle \leq 0$$

et on résout de nouveau l'inégalité variationnelle.

Une autre technique a été suggérée par Zuhovickii, Polyak et Primak (consulter [Auslen76]) dans le contexte des jeux concaves-convexes. Supposons que F soit fortement monotone de coefficient β et que l'ensemble X soit compact. A l'itération k , on résout le problème

$$\max_{x \in X} \min_{l=1..k} \langle F(y^l), y^l - x \rangle \quad (78)$$

dont la solution est notée y^{k+1} . Je pose :

$$\phi_k = \max_{x \in X, z} z \quad (79)$$

$$\text{sujet à } z \leq \langle F(y^l), y^l - x \rangle \quad l = 1..k. \quad (80)$$

Soit :

$$\|y^{k+1} - x^*\| = \min_{l=1..k} \|y^l - x^*\|.$$

On a :

$$\phi_k \geq \min_{l=1..k} \langle F(y^l), y^l - x^* \rangle \geq \beta \min_{l=1..k} \|y^l - x^*\|^2 = \beta \|y^{k+1} - x^*\|^2. \quad (81)$$

Si $y^{k+1} = y^l$ pour $l \in [1..k]$ alors $\phi_k = 0$ et $y^l \in X^*$. Sinon soit $\bar{x} \in X$ la limite d'une sous-suite $\{y^k\}_{k \in S}$, m et k des indices dans S , tels que $m < k$. Puisque :

$$\phi_{k-1} = \min_{l=1..k-1} \langle F(y^l), y^l - y^k \rangle,$$

on a :

$$\phi_{k-1} \leq \langle F(y^m), y^m - y^k \rangle.$$

On a donc : $\lim_{k \in S} \phi_{k-1} \leq \langle F(\bar{x}), \bar{x} - \bar{x} \rangle = 0$ et, puisque $\{\phi_k\}$ est une suite décroissante :

$$\lim_k \phi_{k-1} = 0.$$

Or :

$$\phi_k \geq \beta \|y^{l_k} - x^*\|^2$$

implique que

$$\lim_{k \rightarrow \infty} y^{l_k} = x^*.$$

Je suppose maintenant que F est strictement monotone sur l'ensemble X (on peut affaiblir quelque peu cette dernière hypothèse) et que l'intérieur de ce dernier n'est pas vide. Un schéma algorithmique plus 'moderne' est obtenu en calculant, à partir du point x^k , le 'centre' x^{k+1} de l'ensemble.

$$X^{k+1} = X^k \cap \{x \in R^n : \langle F(x^k), x^k - y \rangle \leq 0\},$$

avec :

$$X^0 = X = \{Ax \leq b\} = \{\langle a^i, x \rangle \leq b_i, \quad i = 1..m\}.$$

Le centre de gravité d'un ensemble est difficile à déterminer. On utilise plutôt le CENTRE ANALYTIQUE, le point qui maximise l'écart par rapport aux contraintes linéaires qui définissent les ensembles X^k . Ainsi, le centre analytique de l'ensemble initial X est l'unique solution du programme strictement concave :

$$\max_{x \in X} \prod_{i=1}^m (b_i - \langle a^i, x \rangle). \quad (82)$$

Puisqu'il est impossible de calculer exactement le centre analytique, on se contentera d'une approximation. L'algorithme prend alors une forme très simple :

Etape 0 (initialisation)

$$k = 0, A^k = A, b^k = b, X^k = \{x : A^k x \leq b^k\},$$

Etape 1 (calcul d'un centre analytique approximatif)

Soit x^k un centre analytique approximatif de X^k .

Etape 2 (critère d'arrêt)

$$\text{Evaluer } g(x^k) = \max_{x \in X^k} \langle F(x^k), x^k - x \rangle$$

si $g(x^k) \leq \epsilon$ **alors** STOP

sinon ALLER A Etape 3.

Etape 3 (génération d'un plan coupant)

$$A^{k+1} = \begin{pmatrix} A^k \\ F(x^k) \end{pmatrix} \quad b^{k+1} = \begin{pmatrix} b^k \\ F(x^k)x^k \end{pmatrix}$$

$H^k = \{x : \langle F(x^k), x - x^k \rangle = 0\}$ est le nouveau plan coupant.

Augmenter k de un.

RETOUR

A Etape 1.

On démontre que la suite $\{x^k\}$ produite par l'algorithme converge vers une solution x^* , sous l'hypothèse que la fonction F est strictement monotone (en fait une condition un peu plus faible). De plus, le nombre d'itérations requis est une fonction polynomiale du nombre de bits de précision désirés, qui est proportionnel à $\log(1/\epsilon)$. Les détails sont donnés dans [GofMZ96].

Après avoir considéré plusieurs techniques de résolution, on peut se demander s'il est plus difficile de résoudre une inégalité variationnelle ou de minimiser une fonction. D'une part, résoudre une inéquation variationnelle dont la fonction est un gradient est 'facile', dans le sens où tout algorithme de descente 'raisonnable' convergera vers un point stationnaire du problème de minimisation équivalent, et il est donc plus simple de résoudre le problème variationnel que le problème d'optimisation. Dans le cas où F n'est pas un gradient, il faut distinguer deux cas. Dans le cas où F est monotone, les deux problèmes sont 'équivalents'. De fait, il est plus simple de résoudre le problème d'optimisation, qui possède des propriétés structurelles plus fortes. Dans le cas non monotone, le problème variationnel est difficile, et équivalent à un problème d'optimisation globale. Le problème d'optimisation est également difficile. Par contre, on peut aisément déterminer un optimum local, ou à tout le moins un point stationnaire.

5. APPLICATIONS

Les inéquations variationnelles représentent le cadre mathématique approprié pour modéliser des situations d'équilibre. Dans cette dernière partie du document, j'en donnerai quelques exemples, principalement reliés à un contexte 'économique' ou 'compétitif' sur des réseaux de transports. D'autres modèles du même type sont présentés dans le livre de Nagurney [Nagur93].

5.1 Équilibre oligopolistique

Soit un ensemble de n firmes en compétition pour la vente d'un bien sur un marché unique. On suppose que le coût de production est une fonction croissante de la quantité produite (la production du bien n'est pas soumise à des économies d'échelle). Chaque firme $i \in I = \{1, \dots, n\}$ est caractérisée par un coût de production $c_i(x_i)$ qui dépend de sa propre production x_i . Une FONCTION DE DEMANDE INVERSE p relie le prix du marché $p(Q)$ à la production totale $Q = \sum_{i \in I} x_i$. Un ÉQUILIBRE DE COURNOT-NASH est un vecteur x^* qui satisfait aux conditions :

$$x_i^* \in \arg \max_{x_i \geq 0} f_i(x_i) = x_i p(x_i + \sum_{j \neq i} x_j^*) - c_i(x_i) \quad \forall i \in I. \quad (83)$$

À l'équilibre, aucune firme ne peut augmenter son profit, une fois fixées les quantités produites par les firmes concurrentes. Cette notion d'équilibre constitue un cas particulier de la notion d'équilibre de Nash introduite à la section 1.1.

Des hypothèses classiques sur les fonctions c_i et p sont :

Hypothèse A1 La fonction p est non négative et décroissante sur R_+ .

Hypothèse A2 La fonction $Qp(Q)$ est concave.

Hypothèse A3 Les fonctions c_i sont positives, convexes et croissantes.

Hypothèse A4 Pour tout i dans I , il existe une quantité non négative \bar{x}_i telle que $\bar{x}_i p(\bar{x}_i) - c_i(\bar{x}_i)$ soit négative.

Hypothèse A5 Les fonctions sont toutes continûment différentiables sur R_+ .

Sous les hypothèses A1–A4, chaque fonction de profit f_i est concave par rapport à x_i et l'ensemble des solutions d'équilibre est non vide, convexe et compact. Si l'hypothèse A5 est satisfaite, la fonction f_i atteint son maximum en un point x_i satisfaisant la condition d'optimalité du premier ordre :

$$\begin{aligned} p\left(\sum_{j \in I} x_j\right) + x_i p'\left(\sum_{j \in I} x_j\right) - c'_i(x_i) &\leq 0 \\ x_i \left[p\left(\sum_{j \in I} x_j\right) + x_i p'\left(\sum_{j \in I} x_j\right) - c'_i(x_i) \right] &= 0 \\ x_i &\geq 0. \end{aligned} \tag{84}$$

Ce problème de complémentarité non linéaire peut être mis sous forme variationnelle $IV(F, R_+^n)$ en posant :

$$F_i(x) = - \left[p\left(\sum_{j \in I} x_j\right) + x_i p'\left(\sum_{j \in I} x_j\right) - c'_i(x_i) \right]. \tag{85}$$

Sous les hypothèses A1–A5, la fonction F n'est pas nécessairement monotone. Une condition suffisante assurant la monotonie de F est que la fonction de demande inverse p soit convexe. Dans ce cas, des algorithmes classiques peuvent être implantés. De tels PROCESSUS DE TÂTONNEMENT ont été proposés. Ils requièrent, à chaque itération k , de résoudre n programmes mathématiques, un par firme :

$$\tilde{x}_i^k \in \arg \max_{x_i \geq 0} x_i p\left(x_i + \sum_{\substack{j \in I \\ j \neq i}} x_j^{k-1}\right) - c_i(x_i).$$

A l'itération $k + 1$ on pose $x^{k+1} = t\tilde{x}^k + (1 - t)x^k$ où t ($0 < t \leq 1$) est un paramètre de sous-relaxation.

Dans le cas de l'oligopole, la formulation variationnelle obscurcit la structure inhérente au problème. Il est en effet préférable de construire un algorithme de résolution autour d'un problème unidimensionnel auxiliaire. En effet, si la production totale Q à l'équilibre était connue a priori, il serait aisé d'obtenir les productions individuelles x_i à partir des formules :

$$\begin{aligned} x_i &= 0 && \text{si } p(Q) - c'_i(0) \leq 0 \\ p(Q) + x_i p'(Q) - c'_i(x_i) &= 0 && \text{sinon.} \end{aligned} \tag{86}$$

Ces conditions sont les conditions d'optimalité du premier ordre du problème de maximisation concave paramétré en Q :

$$\max_{x \geq 0} p(Q) \sum_{i \in I} x_i + \frac{1}{2} p'(Q) \sum_{i \in I} x_i^2 - \sum_{i \in I} c_i(x_i). \quad (87)$$

Puisque je recherche des solutions des programmes ci-dessus qui soient compatibles avec la production totale Q , il est naturel d'étudier le programme mathématique :

$$\begin{aligned} \max_{x \geq 0} \quad & p(Q) \sum_{i \in I} x_i + \frac{1}{2} p'(Q) \sum_{i \in I} x_i^2 - \sum_{i \in I} c_i(x_i) \\ \text{sujet à} \quad & \sum_{i \in I} x_i = Q. \end{aligned} \quad (88)$$

Si la fonction de demande est strictement décroissante, la solution $x(Q)$ est unique. Soit $\lambda(Q)$ la variable duale optimale associée à l'unique contrainte du programme (88). Murphy, Serali et Soyster [**MurSS82**] ont démontré que $\lambda(Q)$ est une fonction continue et non croissante de Q et que Q est une production totale d'équilibre si et seulement si $\lambda(Q) = 0$. Les auteurs ont proposé un algorithme de fouille dichotomique pour localiser l'intervalle contenant la valeur optimale de Q . Si les fonctions c_i sont strictement convexes, ou si $Qp(Q)$ est strictement concave, alors le vecteur d'équilibre est unique.

Une approche différente, qui ne requiert pas la différentiabilité des fonctions c_i et p , est basée sur la CORRESPONDANCE DE RÉACTION INVERSE

$$B(Q) = \sum_{i \in I} x_i(Q),$$

où $x_i(Q)$ dénote l'intervalle des solutions optimales de (88). J'ai démontré [**Marco88**] que Q est une quantité d'équilibre si et seulement si $0 \in \Phi(Q) = B(Q) - Q$. Réciproquement, si $0 \in \Phi(Q)$, tout ensemble de productions $\{x_i(Q)\}_{i \in I}$ satisfaisant l'équation de compatibilité

$$\sum_{i \in I} x_i(Q) = Q$$

constitue un équilibre de Cournot-Nash. On peut localiser un zéro de la correspondance $\Phi(Q)$ par fouille dichotomique, comme précédemment. Malheureusement, la fonction $\Phi(Q)$ n'est pas forcément décroissante sur R^+ . Cependant, j'ai pu montrer que Φ est décroissante sur l'intervalle $[Q_{\min}, Q_{\max}]$, où Q_{\min} satisfait :

$$\begin{aligned} Q_{\min} &= \max_{i \in I} \{x_i^{\min}\} \\ x_i^{\min} &\in \arg \max_{x_i \geq 0} x_i p(x_i) - c_i(x_i) \end{aligned} \quad (89)$$

et Q_{\max} est la production totale correspondant à des productions individuelles au coût marginal :

$$\begin{aligned} Q_{\max} &= \sum_{i \in I} x_i^{\max} \\ x_i^{\max} &\in \arg \max_{x \geq 0} \int_0^{\sum_{i \in I} x_i} p(t) dt - \sum_{i \in I} c_i(x_i). \end{aligned} \quad (90)$$

Ces résultats, intimement reliés à la structure du problème, ne peuvent être obtenus à partir de la formulation variationnelle.

5.2 Équilibres sur des réseaux de transport

Dans cette section, nous reprenons le modèle d'équilibre du trafic en zone urbaine brièvement introduit à la section 1.2. La situation à modéliser est le choix d'itinéraires des usagers d'un réseau routier urbain à l'heure de pointe. Le réseau comprend des arcs, des sommets 'origines', des sommets 'destinations' ainsi que des sommets intermédiaires. Soit P_{kl} l'ensemble des chemins reliant l'origine k à la destination l , et d_{kl} la demande de transport entre les sommets k et l . Le temps de transport sur un arc a est fonction des débits sur les arcs du réseau (principalement le débit sur l'arc a et sur les arcs voisins), exprimés en véhicules/minute. Le temps de transport sur le chemin j est la somme des temps de transport sur les arcs constituant le chemin j . Une AFFECTATION D'ÉQUILIBRE est obtenue lorsque, pour tout couple origine-destination (k, l) , les temps de transport sur les chemins utilisés sont tous égaux, et inférieurs aux temps de parcours des chemins non utilisés (PRINCIPE D'ÉQUILIBRE DE WARDROP). Le problème d'affectation d'équilibre a été abondamment étudié ([Patrik94], [HeFlo96]).

Je dénote par v_a le débit sur l'arc a , x_j le débit sur le chemin j , $C_a(v)$ le temps de parcours de l'arc a , d_{kl} la demande de transport entre k , l , A la matrice d'incidence arc-chemin, $F_j(x)$ le temps de parcours du chemin j et

$$X = \{x \geq 0 : \sum_{j \in P_{kl}} x_j = d_{kl} \quad \forall (k, l)\} \quad (91)$$

l'ensemble des vecteurs de débit admissibles. On a les relations :

$$\begin{aligned} v &= Ax \\ F(x) &= A^t C(Ax). \end{aligned}$$

Le principe de Wardrop peut se reformuler de la façon suivante : un vecteur x est un vecteur d'équilibre si et seulement s'il n'existe pas un vecteur y de coût inférieur par rapport au vecteur de coût $F(x)$, c'est-à-dire que x est un vecteur de plus courts chemins par rapport à $F(x)$:

$$\langle F(x), x \rangle \leq \langle F(x), y \rangle \quad \forall y \in X.$$

Ceci correspond à l'inégalité variationnelle :

$$\langle F(x), x - y \rangle \leq 0 \quad \forall y \in X.$$

Notons que, si C est fortement convexe sur X , alors F est co-coercive sur X , ce qui permet l'application d'algorithmes de projection, entre autres. Les débits d'équilibre sur les arcs sont alors uniques, bien que les débits sur les chemins ne le soient pas.

La structure de réseau permet de résoudre efficacement le problème linéaire

$$\min_{y \in X} \langle F(x), y \rangle$$

qui se décompose en autant de problèmes de plus court chemin qu'il y a de couples origine-destination. Si C (et par conséquent F) est le gradient d'une certaine fonction f , cette structure est fort bien exploitée par l'algorithme de linéarisation de Frank et Wolfe, qui est en fait à la base de tous les algorithmes de calcul d'équilibre sur des réseaux de transport. Si F n'est pas un gradient, on peut minimiser la fonction g à l'aide de l'algorithme construisant une direction de descente $d = y - x$ où $y \in \Gamma_\epsilon(x)$ (noter que la direction de Frank-Wolfe utilise un point *quelconque* de $\Gamma(x)$). Une stratégie populaire est la stratégie de restriction dite de DÉCOMPOSITION SIMPLI-CIALE. Étant donné un ensemble de points extrémaux $\{y^l\}_{l=1..k}$ de X on résoud, de façon approximative, le problème variationnel obtenu en substituant à X l'enveloppe convexe des points y^l . Soit x^k la solution ainsi obtenue. À l'itération subséquente, l'ensemble $\{y^l\}$ est enrichi du point extrémal y^{k+1} solution du problème linéaire :

$$\min_{y \in X} \langle F(x^k), y \rangle.$$

À l'étape k , on se contente donc de résoudre un problème variationnel de très petite taille (k variables non négatives et une seule contrainte de convexité) par l'un ou l'autre des algorithmes étudiés précédemment. Il semble qu'en pratique, moins de 20 itérations soient requises pour obtenir une approximation fort valable d'un équilibre, ce qui permet de calculer des affectations pour des réseaux de très grandes tailles (plusieurs milliers de sommets et d'arcs).

Plusieurs variantes du modèle de base ci-dessus ont été proposées. Mentionnons l'affectation d'équilibre à DEMANDE VARIABLE ou DEMANDE ÉLASTIQUE où la demande d_{kl} est fonction du vecteur des temps de parcours u^* (unique à l'équilibre) des chemins utilisés reliant origines et destinations : $d^* = D(u^*)$. Si cette FONCTION DE DEMANDE D est inversible, le vecteur u^* des coûts de transport à l'équilibre doit satisfaire la relation :

$$u^* = D^{-1}(d^*).$$

La formulation variationnelle du problème avec demande variable est alors, en éten-dant de façon évidente la définition de X :

$$\langle F(x^*), x^* - x \rangle - \langle D^{-1}(d^*), d^* - d \rangle \leq 0 \quad \forall (x, d) \in X \times \{d \geq 0\}. \quad (92)$$

Un autre modèle classique est le problème MULTIMODAL où plusieurs modes de transport (automobile, camion, autobus, taxi, bicyclette) partagent un réseau, en tout ou en partie. À chaque mode r on associe un vecteur de débit x^r , un ensemble admissible X^r ainsi qu'un vecteur de temps de transport $F^r(x)$, où $x = (x^r)_r$. La fonction F^r prend souvent la forme $\bar{F}(\sum_r \alpha_r x^r)$, les coefficients positifs α_r représentant la contribution du mode r à la congestion.

Le mode r est en équilibre lorsque :

$$\langle F^r(x), x^r - y^r \rangle \leq 0 \quad \forall y^r \in X^r. \quad (93)$$

On retrouve IV(F, X) en posant $F(x) = (F^r(x))_r$. Les stratégies de résolution de Jacobi et de Gauss-Seidel, qui décomposent le problème par mode, sont tout indiquées.

Mentionnons également un modèle d'oligopole sur un réseau de transport qui constitue une extension du modèle d'équilibre oligopolistique de la section précédente ([Marco87]) à plusieurs firmes et marchés reliés par un réseau de transport sujet à des effets de congestion. Il est à noter que, dans ce cas, la formulation variationnelle n'est pas monotone et les contraintes de réseau détruisent la structure qui avait permis de développer les algorithmes efficaces de la section précédente. Comme dans le cas multimodal, les algorithmes de décomposition de Jacobi et Gauss-Seidel donnent de bons résultats.

5.3 Équilibre bicritère

Certains choix économiques ou sociaux sont influencés par plusieurs critères. Ainsi, le choix d'un trajet du domicile au lieu de travail peut dépendre du temps de trajet, de son coût, de sa distance, etc. De fait, un usager donné du réseau tentera de minimiser sa propre fonction économique, ou COÛT GÉNÉRALISÉ DE TRANSPORT, qui prendra la forme d'une combinaison linéaire positive des différents critères. Je considère un cas particulier n'impliquant que deux critères, par exemple le temps réel de parcours et le coût monétaire de parcours, dans le contexte de l'équilibre du trafic urbain. On divise les usagers en K classes. Le coût généralisé associé à la classe c est donné par l'expression :

$$F(x) + \alpha_c G(x), \quad (94)$$

où $F(x)$ représente la valeur du premier critère et $G(x)$ celle du second. Soit x^c le vecteur associé à la classe c . Le vecteur de demande b est réparti parmi les classes d'usagers suivant la formule :

$$b^c = h_c b,$$

où (h_c) est un vecteur de probabilités. On dénote :

$$\begin{aligned} X^c &= \{x^c \geq 0 : Ax^c = b^c\} \\ X &= \prod_c X^c \\ x &= (x^c)_c \\ \bar{x} &= \sum_c x^c. \end{aligned}$$

La classe c sera en équilibre lorsque :

$$\langle F(\bar{x}) + \alpha_c G(\bar{x}), x^c - y^c \rangle \leq 0 \quad \forall y^c \in X^c.$$

En combinant toutes ces inégalités variationnelles, on obtient l'inégalité variationnelle $IV(H, X)$ avec :

$$H(x) = \begin{pmatrix} F(\bar{x}) + \alpha_1 G(\bar{x}) \\ \vdots \\ F(\bar{x}) + \alpha_K G(\bar{x}) \end{pmatrix}. \quad (95)$$

La fonction H associée à cette inégalité variationnelle est-elle monotone? La réponse est en général 'non', même si F et G sont monotones. Considérons en effet le cas

unidimensionnel impliquant deux classes, et supposons de plus que les fonctions F et G soient linéaires. La fonction associée à (95) est :

$$\begin{pmatrix} F \cdot (x^1 + x^2) + \alpha_1 G \cdot (x^1 + x^2) \\ F \cdot (x^1 + x^2) + \alpha_2 G \cdot (x^1 + x^2) \end{pmatrix},$$

où F et G sont des nombres réels positifs, et dont la matrice jacobienne est :

$$\begin{pmatrix} F + \alpha_1 G & F + \alpha_1 G \\ F + \alpha_2 G & F + \alpha_2 G \end{pmatrix} = F \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + G \begin{pmatrix} \alpha_1 & \alpha_1 \\ \alpha_2 & \alpha_2 \end{pmatrix}.$$

La première matrice est semidéfinie positive alors que la seconde est définie *négative* (le déterminant de la matrice symétrisée est égal à $-(\alpha_1 - \alpha_2)^2/4$), sauf dans le cas dégénéré où $\alpha_1 = \alpha_2$. En fait, une condition nécessaire et suffisante pour que l'inégalité variationnelle soit monotone est que F et G soient monotones et que G soit *constante*. Dans la suite des opérations, je supposerai que ces conditions sont satisfaites.

Considérons maintenant le cas où la fonction de masse est remplacée par une densité de probabilité continue $h(\alpha)$ sur l'intervalle $[0, \alpha_{\max}]$, où α représente maintenant l'indice d'une classe. On note par $x(\alpha)$ le vecteur associé à la classe α et l'on pose :

$$\bar{x} = \int_0^{\alpha_{\max}} x(\alpha) d\alpha. \quad (96)$$

Les variables de décision $x(\alpha)$ sont maintenant des fonctions, et l'analogie de l'inégalité variationnelle $IV(H, X)$ est située dans l'espace L^2 de dimension infinie des fonctions de carré intégrable, le produit scalaire de deux fonctions vectorielles Φ et Ψ étant défini comme :

$$\begin{aligned} \langle \Phi, \Psi \rangle_{L^2} &= \sum_{j=1}^n \int_0^{\alpha_{\max}} \Phi_j(\alpha) \Psi_j(\alpha) d\alpha \\ &= \int_0^{\alpha_{\max}} \langle \Phi(\alpha), \Psi(\alpha) \rangle d\alpha. \end{aligned}$$

Soit $\bar{X} = \{\bar{x} : A\bar{x} = b, \bar{x} \geq 0\}$, $X(\alpha) = h(\alpha)\bar{X}$ et $X = \prod_{\alpha} X(\alpha)$. Chaque classe α est en équilibre lorsque :

$$\langle F(\bar{x}) + \alpha G, x(\alpha) - y(\alpha) \rangle \leq 0 \quad \forall y(\alpha) \in X(\alpha), \quad \forall \alpha \in [0, \alpha_{\max}].$$

En intégrant, on obtient :

$$\langle F(\bar{x}) + \alpha G, x - y \rangle_{L^2} \leq 0 \quad \forall y \in X. \quad (97)$$

A première vue, ce problème semble difficile. Cependant on remarque que, une fois fixé \bar{x} , le problème :

$$\max_{y \in X} \langle F(\bar{x}) + \alpha G, x - y \rangle_{L^2} \quad (98)$$

n'est rien d'autre qu'un problème de programmation linéaire paramétré par α . Or, les bases de $X(\alpha)$ étant identiques (à une constante multiplicative près) à celles de

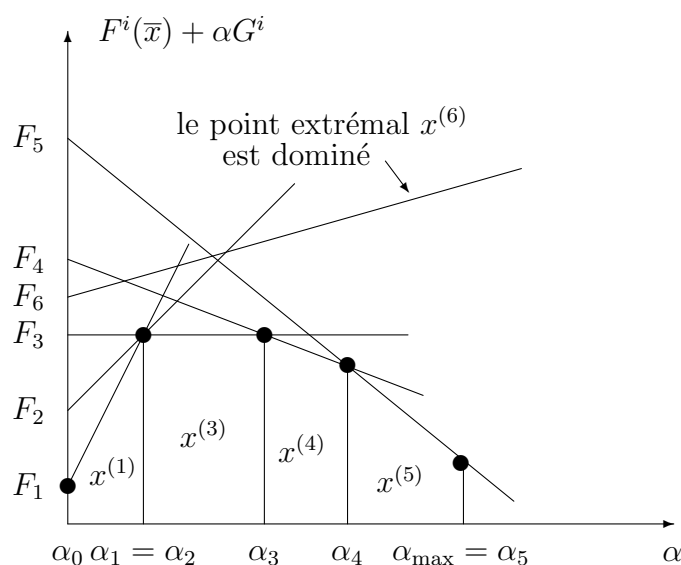


Figure 5: SOLUTION DU PL PARAMÉTRISÉ

X , elles sont en nombre fini, et l'on peut résoudre le problème en dimension infinie en un nombre *fini* d'étapes.

Soient $\{x^{(i)}\}_{i=1}^l$ les points extrémaux du polyèdre compact $\{x : Ax = b, x \geq 0\}$ et posons :

$$\begin{aligned} F^i(\bar{x}) &= \langle F(\bar{x}), x^{(i)} \rangle \\ G^i &= \langle G, x^{(i)} \rangle, \end{aligned}$$

avec $G^1 \geq G^2 \geq G^3 \geq \dots$. Pour une valeur donnée du débit total \bar{x} , la solution du programme linéaire paramétrisé prend la forme (illustrée à la figure 5) :

$$y(\alpha)(\bar{x}) = h(\alpha)x^{(i)} \quad \alpha \in [\alpha_{i-1}, \alpha_i]. \quad (99)$$

Les points critiques α_i correspondent aux intersections des droites $F^i(\bar{x}) + \alpha G^i$ (on pose $\alpha_{l+1} = \alpha_{\max}$). Ainsi, une fois les droites 'dominées' éliminées, on détermine :

$$\alpha_i = -\frac{F^{i+1}(\bar{x}) - F^i(\bar{x})}{G^{i+1} - G^i} \quad \text{si } G^{i+1} \neq G^i.$$

Si les G^i sont tous distincts, ce que j'appellerai le cas 'non dégénéré', alors la solution $(y(\alpha))(\bar{x})$ du problème linéarisé en α est unique, sauf aux points critiques, qui sont en nombre fini, c'est-à-dire que cette solution est unique au sens de la théorie de la mesure. Il s'ensuit que la direction $d(\bar{x}) = y(\bar{x}) - \bar{x}$ est une direction de descente pour la fonction g associée à l'inégalité variationnelle. On peut même démontrer [MarZh95] que, si la fonction F est fortement monotone, les algorithmes utilisant cette direction de descente possèdent un taux de convergence linéaire, ce qui constitue une amélioration considérable par rapport à l'algorithme de Frank et Wolfe classique,

dont le taux de convergence est en général sous-linéaire. De plus, au point d'équilibre x , on a : $y(\bar{x}) = \bar{x}$!

Une formulation en dimension finie du problème d'affectation bicritère a été proposée par Leurent [**Leure93**]. Je présente ci-dessus une 'famille' de formulations en dimension finie. Pour simplifier l'exposition, je suppose, sans perte de généralité, que l'ensemble \bar{X} est le simplexe unité :

$$\bar{X} = \{x \in R^n : \sum_{j=1}^n x_j = 1, x \geq 0\}.$$

On a alors $F^i(\bar{x}) = F_i(\bar{x})$ et $G^i = G_i$, simplement.

A chaque valeur du vecteur \bar{x} correspond un ensemble de valeurs critiques :

$$\alpha_0 = 0 \leq \alpha_1 \leq \dots \leq \alpha_{l-1} \leq \alpha_l = \alpha_{\max}$$

obtenues en résolvant le problème linéaire paramétrisé (98). A partir des points critiques, on obtient $(y(\alpha))(\bar{x})$, puis :

$$\bar{y}(\bar{x}) = \int_0^{\alpha_{\max}} (y(\alpha))(\bar{x})(\alpha) d\alpha.$$

Un équilibre correspond à un point fixe de la correspondance \bar{y} , c'est-à-dire que l'on cherche un vecteur \bar{x} tel que $\bar{x} = \bar{y}(\bar{x})$. Dans le cas où la correspondance \bar{y} n'est pas unique, et si l'on dénote par $\bar{y}(\bar{x})$ l'ensemble des vecteurs totaux associées à l'ensemble des solutions du programme (98), un vecteur d'équilibre total doit satisfaire la relation : $\bar{x} \in \bar{y}(\bar{x})$. Une fois déterminé \bar{x} , il est aisé de reconstruire le vecteur fonctionnel d'équilibre $x(\alpha)$.

Examinons de plus près la figure 5. Posons : $\bar{\alpha} = (\alpha_1, \dots, \alpha_l)$. Les valeurs critiques des α_i (qui correspondent aux intersections des droites 'consécutives') satisfont la relation :

$$\int_0^{\alpha_{\max}} \min_i (F_i(\bar{x}) + \alpha G_i) d\alpha = \min_{\bar{\alpha} \in S} \sum_{i=1}^l \int_{\alpha_{i-1}}^{\alpha_i} (F_i(\bar{x}) + \alpha G_i) d\alpha,$$

avec : $S = \{\bar{\alpha} : 0 = \alpha_0 \leq \alpha_1 \leq \alpha_2 \dots \leq \alpha_l = \alpha_{\max}\}$. On peut donc déterminer les α_i en résolvant le programme mathématique quadratique convexe :

$$\begin{aligned} \min_{\bar{\alpha} \in S} p(\bar{\alpha}) &= \sum_{i=1}^l \int_{\alpha_{i-1}}^{\alpha_i} (F_i(\bar{x}) + \alpha G_i) d\alpha \\ &= \sum_{i=1}^l [(F_i(\bar{x}) - F_{i+1}(\bar{x}))\alpha_i + \frac{1}{2}(G_i - G_{i+1})\alpha_i^2], \end{aligned}$$

avec $F_{l+1}(\bar{x}) = G_{l+1} = 0$ par définition. Les conditions d'optimalité de ce programme sont :

$$\langle p'(\bar{\alpha}), \bar{\alpha} - \bar{\beta} \rangle \leq 0 \quad \forall \bar{\beta} \in S,$$

avec :

$$p'(\bar{\alpha}) = \begin{pmatrix} F_1(\bar{x}) - F_2(\bar{x}) \\ F_2(\bar{x}) - F_3(\bar{x}) \\ \vdots \\ F_i(\bar{x}) - F_{i+1}(\bar{x}) \\ \vdots \\ F_l(\bar{x}) \end{pmatrix} + \begin{pmatrix} (G_1 - G_2)\alpha_1 \\ (G_2 - G_3)\alpha_2 \\ \vdots \\ (G_i - G_{i+1})\alpha_i \\ \vdots \\ G_l\alpha_l \end{pmatrix} .$$

Soit $\phi(\alpha) = \int_0^\alpha h(t) dt$. En remplaçant \bar{x} par

$$\left\{ \int_{\alpha_{k-1}}^{\alpha_k} h(\alpha) d\alpha \right\}_k = \{\phi(\alpha_k) - \phi(\alpha_{k-1})\}_k,$$

on obtient la formulation variationnelle en dimension finie :

$$\langle P(\bar{\alpha}), \bar{\alpha} - \bar{\beta} \rangle \leq 0 \quad \forall \bar{\beta} \in S,$$

avec :

$$P_i(\bar{\alpha}) = F_i \begin{pmatrix} \phi(\alpha_1) \\ \phi(\alpha_2) - \phi(\alpha_1) \\ \vdots \\ \phi(\alpha_l) - \phi(\alpha_{l-1}) \end{pmatrix} - F_{i+1} \begin{pmatrix} \phi(\alpha_1) \\ \phi(\alpha_2) - \phi(\alpha_1) \\ \vdots \\ \phi(\alpha_l) - \phi(\alpha_{l-1}) \end{pmatrix} + (G_i - G_{i+1})\alpha_i.$$

Malheureusement, cette formulation n'est pas monotone, ni même symétrique lorsque F est symétrique. On peut cependant refaire le même boulot en minimisant cette fois l'aire pondérée sous la courbe linéaire par morceaux illustrée à la figure 5. Si l'on pondère par la densité $h(\alpha)$, on obtient le programme mathématique :

$$\min_{\bar{\alpha} \in S} \sum_{i=1}^l \int_{\alpha_{i-1}}^{\alpha_i} (F_i(\bar{x}) + \alpha G_i) h(\alpha) d\alpha \quad (100)$$

qui mène à une formulation variationnelle non monotone, mais symétrique si F est symétrique. Si, par la suite, on effectue un changement de variable judicieux, on obtient une formulation ayant toutes les propriétés requises, et ne faisant intervenir que le vecteur de variables totales \bar{x} . Nous allons retrouver ce résultat, mais d'une façon un peu différente. Si F est le gradient d'une fonction convexe f , l'inégalité variationnelle (97) se ramène au programme convexe :

$$\min_{x \in X} f(\bar{x}) + \sum_{i=1}^l G_i \int_0^{\alpha_{\max}} \alpha x_i(\alpha) d\alpha, \quad (101)$$

qu'on reformule en les variables α_i :

$$\min_{\bar{\alpha}} q(\bar{\alpha}) = f(\phi(\alpha_1), \phi(\alpha_2) - \phi(\alpha_1), \dots, \phi(\alpha_l) - \phi(\alpha_{l-1})) + \sum_{i=1}^l G_i [\sigma(\alpha_i) - \sigma(\alpha_{i-1})], \quad (102)$$

avec $\sigma(\alpha) = \int_0^\alpha th(t) dt$. On remarque, si la fonction ϕ est inversible, que :

$$\begin{aligned} \bar{x}_i &= \phi(\alpha_i) - \phi(\alpha_{i-1}) & i &= 1..l \\ \alpha_i &= \phi^{-1}(\bar{x}_1 + \dots + \bar{x}_i) & i &= 1..l \end{aligned} \quad (103)$$

On peut donc réécrire le problème d'optimisation (101) comme :

$$\min_{\bar{x} \in X} f(\bar{x}) + \sum_{i=1}^l (G_i - G_{i+1}) \mu(\bar{x}_1 + \dots + \bar{x}_i) \quad (104)$$

où $\mu = \sigma \circ \phi^{-1} = \int \phi^{-1} + \text{constante}$. Or, puisque $G_i \geq G_{i+1}$ et que ϕ^{-1} est croissante, on en déduit que cette formulation en est une d'optimisation convexe. Si la fonction F n'est pas un gradient, on démontre que la formulation variationnelle correspondant à (101) est monotone.

5.4 Équilibres stratégiques

Dans le modèle classique d'affectation d'équilibre, les effets de congestion sont représentés à l'aide de fonctions de délais. Ces fonctions, croissantes, ont comme argument les débits sur les arcs ou les chemins. En pratique, cependant, les débits ne peuvent dépasser les capacités physiques du réseau. En fait, un débit faible peut correspondre soit à une circulation fluide, soit à une circulation extrêmement congestionnée! Si l'on introduit explicitement des capacités dans le modèle, on n'est plus assuré de l'existence d'un équilibre au sens de Wardrop. Je considère donc l'extension suivante du principe de Wardrop. Pour simplifier l'exposition, je supposerai que l'ensemble X est le simplexe unité muni des contraintes de capacités : $x_j \leq u_j$. Je dirai qu'un vecteur x de débits sur les chemins est un vecteur d'équilibre si les conditions suivantes sont satisfaites :

$$\begin{aligned} F_j(x) - \alpha + \beta_j &\geq 0 & \forall j \\ x_j(F_j(x) - \alpha + \beta_j) &= 0 & \forall j \\ \beta_j(x_j - u_j) &= 0 & \forall j \\ \beta_j &\geq 0 & \forall j \\ \sum_j x_j &= 1 \\ x_j &\geq 0 & \forall j \\ x_j &\leq u_j & \forall j \end{aligned} .$$

Il n'est pas difficile de voir que les conditions ci-dessus coïncident avec les conditions de 'Kuhn-Tucker' de l'inégalité variationnelle :

$$\langle F(x), x - y \rangle \leq 0 \quad \forall y \in X.$$

Les variables duales β_j peuvent être interprétées comme des délais dus à des phénomènes d'attente.

Une autre approche, dite 'stratégique', peut être envisagée. Je vais d'abord présenter le concept d'une STRATÉGIE DE DÉPLACEMENT dans un réseau. Ce concept a été utilisé avec succès dans le contexte des transports en commun urbains, où les usagers peuvent effectuer des correspondances. A chaque sommet de correspondance, un

choix de lignes (qui peut se réduire à une seule) se présente à l'utilisateur. Ce dernier doit donc effectuer un choix. Pour simplifier, je supposerai que 2 lignes de fréquences respectives φ_1 et φ_2 relient l'origine du trajet à sa destination, et que les arrivées des autobus constituent un processus sans mémoire (processus de Poisson), c'est-à-dire que l'espérance du temps d'attente est indépendante du temps passé à l'arrêt. Puisque le processus est sans mémoire, il n'est pas nécessaire de considérer les stratégies qui tiennent compte du temps d'attente écoulé. Soient T_1 et T_2 ($T_1 < T_2$) les durées des trajets des lignes 1 et 2. Trois choix, ou stratégies, se présentent à l'utilisateur :

- Monter à bord d'un autobus de la ligne 1, exclusivement.
- Monter à bord du premier autobus se présentant à l'arrêt.
- Monter à bord d'un autobus de la ligne 2, exclusivement.

La troisième stratégie étant clairement dominée par les deux premières, nous ne la considérerons pas. La deuxième stratégie sera optimale si :

$$T_2 \leq \frac{1}{\varphi_1} + T_1,$$

c'est-à-dire que la durée de trajet plus courte de la ligne 1 ne compense pas pour l'espérance du temps d'attente $1/\varphi_1$. La stratégie optimale correspondra au minimum des quantités $(1/\varphi_1)(1 + \varphi_1 T_1)$ et $(1/(\varphi_1 + \varphi_2))(1 + \varphi_1 T_1 + \varphi_2 T_2)$. Ce résultat se généralise à n lignes communes 1, 2, ..., n de fréquences $\varphi_1, \dots, \varphi_n$ et durées de trajet $T_1 < T_2 < \dots < T_n$. On identifie la stratégie incorporant les lignes 1, 2, ..., k à l'indice k . La stratégie optimale correspond alors à :

$$k \in \arg \min_{1 \leq k \leq n} \frac{1}{\sum_{l=1}^k \varphi_l} \left(1 + \sum_{l=1}^k \varphi_l T_l\right). \quad (105)$$

Dans un réseau complexe, il est possible de déterminer les stratégies optimales par des algorithmes très semblables aux algorithmes de calcul des plus courts chemins, tels que ceux de Dijkstra, Ford-Bellman, etc (voir [GaLNP93]). On note que, alors que le temps de transport relié à l'utilisation d'une stratégie est aléatoire, la stratégie elle-même est déterministe.

Je vais appliquer cette notion de stratégie aux réseaux avec capacités rigides, tels que les réseaux de transport en commun interurbain où le nombre de places disponibles à l'intérieur des véhicules est limité. Lorsque x usagers tentent d'accéder à un arc du réseau dont la capacité est u , on fait l'hypothèse que la probabilité d'accès est égale à :

$$\min \left\{ 1, \frac{u}{x} \right\}.$$

Un usager n'ayant pu accéder à un arc saturé doit reporter son choix sur un second arc, puis éventuellement sur un troisième, etc. Une stratégie correspondra dans ce cas, en chaque sommet du réseau, à un ordre de préférence sur les arcs incidents au sommet considéré. Cet ordre dépendra évidemment de la destination de l'utilisateur

arc	(1,3)	(1,4)	(2,3)	(2,6)	(3,4)	(3,6)	(4,5)	(4,6)	(5,6)
temps	15	50	10	55	10	150	30	10	30
capacité	∞	∞	∞	∞	10	∞	∞	10	∞

demande : 10 unités de 1 à 6 et 10 unités de 2 à 6.

Table 2: LES DONNÉES DE L'EXEMPLE

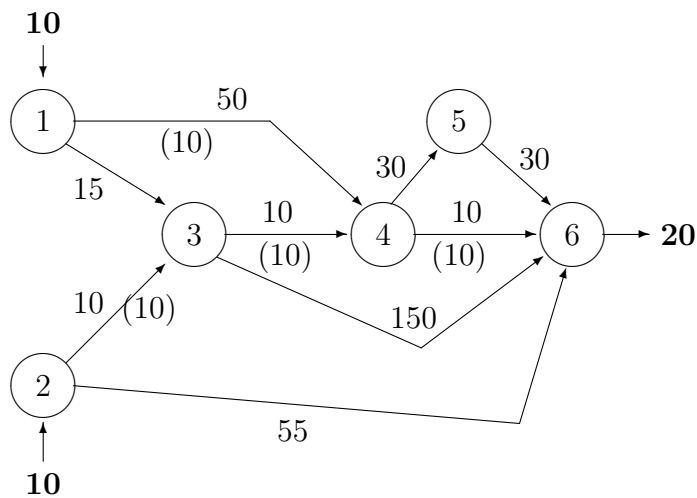


Figure 6: UN RÉSEAU AVEC CAPACITÉS RIGIDES

ayant atteint le sommet en question. On note que, contrairement à la situation précédente, les probabilités d'accès dépendent des stratégies de tous les usagers, et non pas seulement des choix individuels des usagers.

Un exemple permettra d'éclaircir la situation. Considérons la situation illustrée à la figure 6 ainsi que les temps, capacités (entre parenthèses) et demandes données à la table 2 et à la figure 6.

Nous cherchons à déterminer une répartition des stratégies parmi les usagers telle que, pour chaque couple origine-destination, les temps (espérés) de transport soient identiques (pour chaque couple origine-destination) et inférieurs ou égaux aux temps des stratégies non utilisées, comme l'aurait souhaité Wardrop. Un échantillon de stratégies (et elles sont en grand nombre, même pour un si petit réseau!) est donné à la table 3.

Les 12 premières lignes de cette table correspondent aux stratégies associées au couple 1–6, et les 11 dernières lignes sont associées au couple 2–6. Chaque colonne contient les successeurs d'un sommet, en ordre de préférence décroissant. Une case vide correspond à un sommet qui ne peut être atteint à partir de l'origine de la stratégie

	sommets :						
stratégies	1	2	3	4	5	6	
s_1	[4]	[]	[]	[5]	[6]	[]	dominée par s_3 non admissible
s_2	[4]	[]	[]	[6]	[]	[]	
s_3	[4]	[]	[]	[6,5]	[6]	[]	équivalente à s_1
s_4	[4]	[]	[]	[5,6]	[6]	[]	
s_5	[3]	[]	[6]	[]	[]	[]	
s_6	[3]	[]	[4]	[6]	[]	[]	
s_7	[3]	[]	[4,6]	[6,5]	[6]	[]	
s_8	[3]	[]	[4,6]	[5,6]	[6]	[]	
s_9	[3]	[]	[6,4]	[6,5]	[6]	[]	
s_{10}	[3]	[]	[6,4]	[5,6]	[6]	[]	
s_{11}	[3]	[]	[4,6]	[5]	[6]	[]	
s_{12}	[3]	[]	[6,4]	[5]	[6]	[]	
s_{13}	[]	[6]	[]	[]	[]	[]	
s_{14}	[]	[3]	[6]	[]	[]	[]	
s_{15}	[]	[3]	[4]	[6]	[]	[]	
s_{16}	[]	[3]	[4]	[6,5]	[6]	[]	
s_{17}	[]	[3]	[4]	[5,6]	[6]	[]	
s_{18}	[]	[3]	[4,6]	[6,5]	[6]	[]	
s_{19}	[]	[3]	[4,6]	[5,6]	[6]	[]	
s_{20}	[]	[3]	[6,4]	[6,5]	[6]	[]	
s_{21}	[]	[3]	[6,4]	[5,6]	[6]	[]	
s_{22}	[]	[3]	[4,6]	[5]	[6]	[]	
s_{23}	[]	[3]	[6,4]	[5]	[6]	[]	

Table 3: UN ÉCHANTILLON DE STRATÉGIES

	sommets : 1 2 3 4 5 6					
stratégies						
s_3	[4]	[]	[]	[6,5]	[6]	[]
s_7	[3]	[]	[4,6]	[6,5]	[6]	[]
s_{13}	[]	[6]	[]	[]	[]	[]
s_{18}	[]	[3]	[4,6]	[6,5]	[6]	[]

Table 4: LES QUATRE STRATÉGIES INTÉRESSANTES

correspondante.

Une stratégie est ADMISSIBLE si la probabilité d'atteindre éventuellement la destination en utilisant cette stratégie est égale à 1, quelle que soit la demande et les choix stratégiques sur le réseau. En d'autres termes, une stratégie est admissible si et seulement si elle contient un chemin de capacité infinie, un sentier pédestre par exemple. Ainsi les stratégies s_2 , s_6 , s_{15} et s_{16} ne sont pas admissibles. Par la suite, nous ne considérerons que les stratégies admissibles.

Une stratégie s est DOMINÉE s'il existe une stratégie s^* telle que, quelles que soient les valeurs des débits sur le réseau, le temps espéré de s^* est inférieur à celui de s . Dans notre exemple, la stratégie s_1 est dominée par s_3 , la stratégie s_5 est dominée par s_1 (et aussi par d'autres stratégies), s_{11} est dominée par s_7 ; s_{14} et s_{22} sont dominées par s_{13} . Enfin, certaines stratégies sont ÉQUIVALENTES si les probabilités d'accès aux arcs qui leurs sont associées sont égales, en tout sommet du réseau. En ce sens, la stratégie s_4 est équivalente à s_1 . Certaines stratégies redondantes, telles que $([4, 3], [], [], [5], [6], [])$ n'ont pas été incluses dans la table 3. On ne conservera, parmi les stratégies redondantes, que celle dont les listes de préférence sont les plus courtes.

Une fois éliminées les stratégies dominées et redondantes, les seules stratégies intéressantes sont s_3 , s_7 , s_{13} et s_{18} (voir la table 4).

Puisque la stratégie s_{13} est formée du chemin unique 2–6, son délai de 55 est indépendant des débits. Ce nombre fournit donc une borne supérieure sur le temps de transport pour ce couple origine-destination. Je prétends qu'à l'équilibre les deux stratégies s_{13} et s_{18} sont utilisées et que par conséquent le temps à l'équilibre doit être égal à 55. En effet, si la stratégie s_{18} est utilisée, son temps doit être inférieur ou égal à 55, et le temps de la stratégie s_7 doit être inférieur à 50, puisqu'elles partagent une sous-stratégie commune à partir du sommet 3. Donc la stratégie s_3 ne sera pas utilisée à l'équilibre. Ceci implique que les usagers ayant comme origine le sommet 2 ne seront pas tous affectés à la stratégie s_{18} , car ceci résulterait en un temps de $(10 + (20 + 150)/2) > 55$ pour ceux-ci. A l'équilibre, les usagers provenant du sommet 2 sont répartis entre les stratégies s_{13} et s_{18} , et les usagers provenant du sommet 1 sont tous affectés à la stratégie s_7 . Il reste à spécifier le nombre x d'usagers provenant

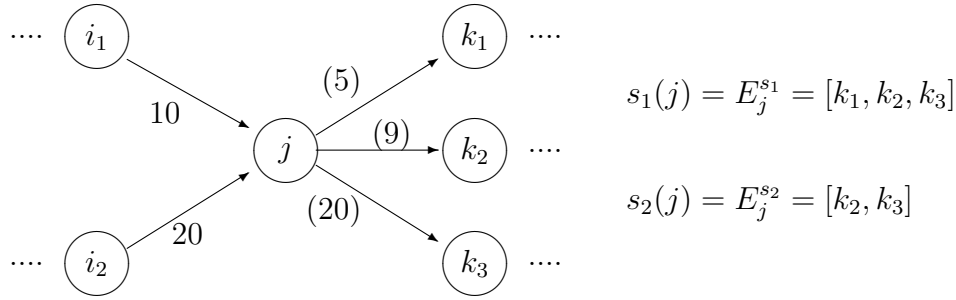


Figure 7: EXEMPLE DE CHARGEMENT À UN SOMMET

		arc	(j, k_1)	(j, k_2)	(j, k_3)
Itération 1 :	capacité résiduelle		5	9	∞
	débit		$5(s_1)$	$5(s_2)$	0
Itération 2 :	capacité résiduelle		0	4	∞
	débit		$5(s_1)$	$5(s_2)+1(s_1)+3(s_2)$	0
Itération 3 :	capacité résiduelle		0	0	∞
	débit		$5(s_1)$	$1(s_1)+8(s_2)$	$4(s_1)+12(s_2)$

probabilités : $\pi_{jk_1}^{s_1} = \frac{5}{10}$ $\pi_{jk_2}^{s_1} = \frac{1}{10}$ $\pi_{jk_3}^{s_1} = \frac{4}{10}$ $\pi_{jk_2}^{s_2} = \frac{8}{20}$ $\pi_{jk_3}^{s_2} = \frac{12}{20}$

Table 5: LE CHARGEMENT

du sommet 2 qui utilisent la stratégie s_{18} . En suivant les raisonnements précédents, on dérive l'équation :

$$10 + \frac{10}{10+x}(10+10) + \frac{x}{10+x}150 = 55,$$

dont la solution est : $x = 50/21$.

Afin d'obtenir une formulation variationnelle de ce problème d'affectation, il est nécessaire de définir le temps espéré associé à un vecteur de stratégies. Cette fonction de temps est déterminé à l'aide d'un ALGORITHME DE CHARGEMENT. A un sommet j du réseau, les usagers sont d'abord affectés à leur premier choix, suivant un taux uniforme. Dès que le premier choix d'un groupe d'usager est saturé, on remet à jour sa liste de préférence et on répète le processus. Un exemple de chargement est illustré à la figure 7 et à la table 5, où 10 usagers utilisant la stratégie s_1 ont atteint le sommet j et doivent être affectés aux arcs ayant j comme sommet initial. Ce nombre est égal à 20 pour la stratégie s_2 . Une fois affectées 5 unités des stratégies s_1 et s_2 à leur premier choix respectif (j, k_1) et (j, k_2) , l'arc (j, k_1) est saturé. Les débits résiduels (5 et 15) sont alors affectés dans la proportion 1 à 3 à l'arc (j, k_2) . Une fois celui-ci saturé, les unités résiduelles sont affectées à l'arc (j, k_3) .

L'étude de l'inégalité variationnelle résultante (unicité de la solution, monotonie) est évidemment très complexe, et toute approche de résolution est forcément heuristique.

L'extension de ce modèle au cas où des priorités sont présentes est encore plus difficile. Or, ce cas est d'un intérêt pratique car les usagers se trouvant déjà à bord d'un autobus ont évidemment priorité sur les usagers effectuant une correspondance.

5.54 Équilibre dynamique

Dans cette section, je présente un modèle dynamique d'affectation du trafic urbain, formulé comme une inégalité variationnelle dans un espace de dimension infinie. Ce modèle est dû à Friesz et al. [FBSTW93]. Deux aspects principaux distinguent ce modèle du modèle statique :

- les variables de décision ne sont plus des débits sur les arcs ou chemins du réseau, mais plutôt des volumes (nombres réels) de véhicules;
- il est difficile de déduire les volumes sur les arcs et les chemins, et par conséquent les délais sur les arcs, à partir des 'désirs' de déplacement des usagers du réseau.

Le dernier point constitue le grand défi que doit relever tout modèle dynamique. En effet, afin d'obtenir les volumes sur les arcs à chaque instant, il est nécessaire d'effectuer un chargement des véhicules sur le réseau, un peu comme à la section précédente. Ce qui amène la question suivante : les usagers ont-ils pleine et entière connaissance des conditions sur le réseau, à *tout instant*? Peuvent-ils s'ajuster aux conditions locales? Le modèle présenté en est un en BOUCLE FERMÉE, c'est-à-dire que les usagers ont une connaissance parfaite des conditions du réseau et des effets ultérieurs (parfois pervers!) qu'auront leurs décisions présentes ainsi que celles des autres automobilistes.

Je dénote par $G(N, A)$ le graphe physique, formé d'un ensemble A d'arcs et d'un ensemble N de sommets. Soit P_{rs} l'ensemble des chemins reliant l'origine r à la destination s , et $P = \cup_{rs} P_{rs}$. Soit $[0, T]$ la période de départs considérée et $[0, \bar{T}]$ une période suffisamment étalée pour permettre à tous les automobilistes quittant leur origine à l'instant T de compléter leur trajet avant l'instant \bar{T} . Je pose :

$$\begin{aligned} x_a(t) &= \text{volume sur l'arc } a \text{ à l'instant } t, \\ x_a^p(t) &= \text{volume du chemin } p \text{ sur l'arc } a \text{ à l'instant } t. \end{aligned}$$

Un élément δ_{ap} de la matrice d'incidence sommet-arc Δ satisfait :

$$\delta_{ap} = \begin{cases} 1 & \text{si } a \in p \\ 0 & \text{sinon} \end{cases} \quad (106)$$

et l'on a :

$$x_a(t) = \sum_{p \in P} \delta_{ap} x_a^p(t) \quad \forall a \in A. \quad (107)$$

On fait l'hypothèse que les volumes sont nuls au début de la période étudiée :

$$x_a^p(0) = 0, \quad \forall a, p. \quad (108)$$

Soit $u_a(t)$ le débit à l'entrée de l'arc a à l'instant t et v_a le débit de sortie. A tout instant t , l'équation différentielle suivante doit être satisfaite :

$$(x_a^p)'(t) = u_a^p(t) - v_a^p(t) \quad \forall a, p. \quad (109)$$

Les volumes à l'instant t sont donnés par la relation :

$$x_a^p(t) = \int_0^t [u_a^p(s) - v_a^p(s)] ds \quad \forall a, p. \quad (110)$$

Le temps de parcours de l'arc a pour les usagers parvenant à l'arc a à l'instant t , noté D_a , est déterminé par le volume sur l'arc a à l'instant t . L'instant de sortie (fonction) de l'arc a est noté τ_a et l'on a :

$$\tau_a(t) = t + D_a(x_a(t)). \quad (111)$$

Les équations de conservation des volumes au sommet j ($j \neq r, s$) prennent la forme :

$$\sum_{a \in B(j)} v_a^p(t) = \sum_{a \in A(j)} u_a^p(t) \quad \forall j \quad \forall t \quad (112)$$

où $B(j)$ représente l'ensemble des arcs ayant j comme sommet initial et $A(j)$ l'ensemble des arcs ayant j comme sommet terminal. Les équations de conservation à un sommet origine r relient les débits de départ (exogènes) h_p aux volumes d'entrée des arcs ayant r comme sommet initial. On note Q_{rs} le nombre total d'usagers voyageant entre r et s pendant la période $[0, T]$ et $I_p(t)$ le nombre de véhicules ayant accédé au premier arc du chemin p entre les instants 0 and t . Soit $h_p(t)$ le débit à l'entrée du premier arc du chemin p à l'instant t , c'est-à-dire :

$$h_p(t) = I_p'(t) \quad (113)$$

En utilisant ces notations, on peut exprimer les contraintes de satisfaction de la demande de transport :

$$\sum_{p \in P_{rs}} I_p(T) = \sum_{p \in P_{rs}} \int_0^T h_p(s) ds = Q_{rs} \geq 0 \quad \forall r, s. \quad (114)$$

Enfin soit :

$$\Lambda = \{(h_p)_{p \in P} : h_p \geq 0, h_p \in (L^2[0, T])^{|P|}, \sum_{p \in P_{rs}} \int_0^T h_p(s) ds = Q_{rs}\}. \quad (115)$$

Les équations (106)-(113) et (115), combinées aux contraintes de non négativité sur les volumes, définit le modèle dynamique qui décrit le chargement du réseau correspondant aux choix de chemin des usagers. Les variables de décision sont les volumes sur les chemins $h_p(t)$, alors que les variables x_a^p , u_a^p , v_a^p et τ_a sont les variables d'état du système. Si le système satisfait la condition PAPS (PREMIER ARRIVÉ PREMIER

SERVI), c'est-à-dire qu'il n'y a pas de dépassements sur les arcs du réseau, alors les quantités x_a , u_a , v_a et τ_a sont bien définies en fonction de h . La condition PAPS n'est pas une conséquence de la définition du système dynamique, et ceci constitue une difficulté non négligeable de la modélisation dynamique. Des conditions suffisantes garantissant la condition PAPS ont été données par [FBSTW93] et [ZhuMa96]. On note que la condition PAPS est satisfaite si et seulement si la fonction τ_a est inversible.

A partir du système dynamique ci-dessus, il est possible de formuler le problème d'affectation d'équilibre comme une inégalité variationnelle. Pour l'instant, les algorithmes de résolution n'ont pu traiter que des problèmes de tailles faibles (voir [WuCFX95]).

6. PROGRAMMATION À DEUX NIVEAUX

Cette section ne traite pas explicitement d'inégalités variationnelles mais plutôt de modèles où elles tiennent une place importante comme problème interne d'un PROGRAMME MATHÉMATIQUE À DEUX NIVEAUX. Mais étudions d'abord un exemple simple. Retournons au problème d'équilibre oligopolistique impliquant $n + 1$ firmes, où les n premières firmes (suiveurs) se comportent 'à la Cournot-Nash', mais où la firme $n + 1$ (meneur) maximise son revenu en fonction de la réaction des n autres firmes à son niveau de production. On obtient ainsi le programme mathématique :

$$\max_{x_{n+1} \geq 0} x_{n+1} p(x_{n+1} + \sum_{j=1}^n x_j(x_{n+1})) - f_{n+1}(x_{n+1}), \quad (116)$$

où $\{x_j(x_{n+1})\}_{j=1..n}$ représente un équilibre de Cournot-Nash paramétrisé en x_{n+1} . L'équilibre obtenu est un ÉQUILIBRE DE STACKELBERG-NASH-COURNOT (consulter [SheSM83]) et se réduit, dans le cas d'un suiveur unique ($n = 1$) au concept classique d'équilibre de Stackelberg. La fonction objectif du meneur étant non convexe, on ne peut guère faire mieux qu'une recherche exhaustive de la valeur optimale de x_{n+1} .

On peut généraliser ce modèle en introduisant plusieurs meneurs qui se comportent entre eux à la Cournot-Nash et envers les suiveurs à la Stackelberg. Ce cas est malheureusement plus complexe, et des conditions fortes doivent être imposées pour garantir l'existence d'une solution d'équilibre.

Le modèle de Stackelberg-Nash-Cournot constitue un cas particulier de PROGRAMME MATHÉMATIQUE AVEC CONTRAINTES D'ÉQUILIBRE (ou MPEC pour 'Mathematical Program with Equilibrium Constraints' [LuoPR96]). Une bibliographie récente en a été effectuée par Vicente et Calamai [ViCal94]. Ces modèles représentent adéquatement des situations où un gouvernement doit tenir compte de l'influence de ses prises de décision sur l'économie, les entreprises, les électeurs, etc. La formulation mathématique générale d'un programme mathématique à deux niveaux est :

$$\begin{aligned} & \min_{(x,y) \in X} && f(x, y) \\ \text{sujet à} &&& \langle F(x, y), y - z \rangle \leq 0 \quad \forall z \in Y(x). \end{aligned} \quad (117)$$

On requiert habituellement, afin de garantir l'existence d'une solution, que l'inégalité variationnelle du suiveur (ou SECOND NIVEAU) soit monotone. Si les contraintes de second niveau sont données explicitement par le système d'inégalités régulier : $G(x, y) \leq 0$, il est alors possible de substituer au problème du second niveau ses conditions de Kuhn-Tucker, pour obtenir un programme à un seul niveau. Soit G_y la matrice jacobienne de la fonction G par rapport au vecteur du second niveau y . Je réécris (117) sous la forme :

$$\begin{aligned} \min_{x,y,\lambda} \quad & f(x, y) \\ \text{sujet à} \quad & (x, y) \in X \\ & F(x, y) + \lambda G(x, y) = 0 \\ & \lambda \geq 0 \\ & \lambda G(x, y) = 0. \end{aligned} \tag{118}$$

Une erreur funeste, commise à plusieurs reprises, consiste à dériver des conditions nécessaires d'optimalité pour le programme mathématique (118) en écrivant de nouveau les conditions de Kuhn et Tucker. En effet, les contraintes de (118), sauf situation exceptionnelle, ne sont jamais régulières, c'est-à-dire ne satisfont aucune qualification des contraintes.

Afin d'apprécier la difficulté d'un tel programme, je vais considérer le cas le plus 'simple' possible, où f est linéaire, F est constante, $Y(x)$ est un polyèdre convexe et il n'y a pas de contraintes au premier niveau :

$$\begin{aligned} \min_{x,y} \quad & ax + by \\ \min_y \quad & cx + dy \\ & Ax + By \geq s, \end{aligned} \tag{119}$$

qu'on peut réécrire, comme auparavant dans le cas général et après avoir noté qu'on peut poser $c = 0$, sans perte de généralité :

$$\begin{aligned} \min_{x,y,\lambda} \quad & ax + by \\ \text{sujet à} \quad & Ax + By \geq s \\ & \lambda B = d \\ & \lambda \geq 0 \\ & \lambda(Ax + By - s) = 0. \end{aligned} \tag{120}$$

S'il existe une solution optimale de ce problème, il existe une solution extrême optimale, comme en programmation linéaire. Il en résulte qu'un programme à deux niveaux linéaire est en fait un problème combinatoire. Il a été démontré que ce problème est NP-COMPLET, c'est-à-dire qu'il fait partie d'une classe de problèmes combinatoires 'difficiles'. La difficulté provient des contraintes de complémentarité

$$\lambda(Ax + By - s) = 0.$$

Il a été proposé de pénaliser ces contraintes, ce qui mène au problème bilinéaire :

$$\begin{aligned} \min_{x,y,\lambda} \quad & ax + by + M\lambda(Ax + By - s) \\ \text{sujet à} \quad & Ax + By \geq s \\ & \lambda B = d. \\ & \lambda \geq 0 \end{aligned} \tag{121}$$

Cette pénalisation est EXACTE dans le sens où il existe une valeur finie de M telle que les solutions du problème original et du problème pénalisé coïncident.

Le problème étant NP-complet, il n'est pas surprenant que les algorithmes les plus efficaces soient des algorithmes d'énumération implicite de type 'Branch-and-Bound', le branchement étant effectué sur les contraintes de complémentarité (voir [HanJS92]).

Le problème de taxation optimale (optimale dans le sens de maximisation des revenus) se prête bien à une formulation biniveau. Soit x un vecteur de taxes affectant un vecteur de produits et activités y_1 et soit y_2 un vecteur de produits et activités non taxés. Je considère le programme bilinéaire à deux niveaux :

$$\begin{aligned} \max_{x,y_1,y_2} \quad & xy_1 \\ \min_{y_1,y_2} \quad & (c+x)y_1 + dy_2 \\ & A_1y_1 + A_2y_2 \geq s. \end{aligned} \tag{122}$$

Ce problème a été étudié par Labbé, Marcotte et Savard [LaMaS96] qui ont démontré que certains cas particuliers reliés à la détermination de péages optimaux sur un réseau routier peuvent être résolus en temps polynomial.

Mentionnons pour terminer une application dans le contexte de la déréglementation du transport en commun urbain ([FerMa92]) qui tient à la fois de l'équilibre de Nash et de l'équilibre de Stackelberg. On considère un ensemble fixe de trajets desservis par des lignes d'autobus. Chaque opérateur est propriétaire de son véhicule et l'entrée dans le système est libre. On s'attend ainsi à ce que les lignes les plus fréquentées aient des fréquences plus élevées. On cherche alors un équilibre entre l'offre de service sur les lignes du réseau et la demande de transport sur ces lignes. On suppose que les usagers adoptent le comportement stratégique décrit à la section 5.4.

La particularité de ce modèle tient de la relation entre la variation de fréquence sur une ligne et la demande sur cette même ligne. Plutôt que d'ajuster les fréquences aux débits courants, l'on suppose que la demande aux sommets de correspondance est reliée à l'offre par le biais des probabilités d'accès aux lignes attrayantes, c'est-à-dire les lignes faisant partie des stratégies des usagers. Les opérateurs intègrent ainsi partiellement dans leur processus de décision le comportement stratégique des usagers.

BIBLIOGRAPHIE

- AuCel84** AUBIN, J.-P. ET CELLINA, A., Differential inclusions, *Springer-Verlag*, Berlin, 1984.
- Ausle76** AUSLENDER, A., Optimisation. Méthodes numériques, *Masson*, Paris, 1976.
- Bazar94** BAZAARA, M. S., SHERALI, H. D. ET SHETTY, C. M., Nonlinear Programming. Theory and algorithms, *Wiley*, New York, 1993.
- BerTs89** BERTSEKAS, D. P. ET TSITSIKLIS, Parallel and distributed computation. Numerical methods, *Prentice-Hall*, Englewood Cliffs, 1989.
- Borde85** BORDER, K., Fixed point theorems with applications to economics and game theory, *Cambridge University Press*, Cambridge (1985).
- Cohen88** COHEN, G., Auxiliary problem principle extended to variational inequalities, *JOTA* **59** (1988) 325–333.
- Dafer83** DAFERMOS, S. C., An iterative scheme for variational inequalities, *Mathematical Programming* **26** (1983) 40–47.
- FBSTW93** FRIESZ, T.L., BERNSTEIN, D., SMITH, T.E., TOBIN, R.L. AND WEI, B.W., A variational inequality formulation of the dynamic network user equilibrium problem, *Operations Research* **41** 179-191 (1993).
- FerMa92** FERNÁNDEZ, E. ET MARCOTTE, P., Operators-users equilibrium model in a partially regulated market, *Transportation Science* **26** (1992) 93–105.
- GaLNP93** GALLO, G., LONGO, G., NGUYEN, S. ET PALLOTTINO, S., Directed Hypergraphs and applications, *Discrete Applied Mathematics* **42** (1993) 177–201.
- GloLT76** GLOWINSKY, R., LIONS, J. L. ET TRÉMOLIÈRES, R., *Analyse numérique des inéquations variationnelles*, Bordas, Paris (1976).
- GofMZ96** GOFFIN, J.-L., MARCOTTE, P. ET ZHU, D. L., An analytic center cutting plane method for pseudomonotone variational inequalities, *Operations Research Letters* (1996) – .
- HanJS92** HANSEN, P., JAUMARD, B. ET SAVARD, G., New branch-and-bound rules for linear bilevel programming, *SIAM Journal on Scientific and Statistical Computing* **13** (1992) 1194–1217.
- HarPa90** HARKER, P. ET PANG, J.-S., Finited-dimensional variational inequality and complementarity problems, *Mathematical Programming* **48** (1990) 161–220.

- HeFlo96** HEARN, D. W. ET FLORIAN, M., Network equilibrium models and algorithms, NETWORK ROUTING. HANDBOOKS IN OPERATIONS RESEARCH AND MANAGEMENT SCIENCE, VOLUME 8, NORTH-HOLLAND, Amsterdam, 1996.
- KarSC93** KARAMARDIAN, S., SCHAIBLE, S. AND CROUZEIX, J.-P., Characterizations of Generalized Monotone Maps, *JOTA* **76** (1993) 399–413.
- KaSch90** KARAMARDIAN, S. AND SCHAIBLE, S., Seven kinds of monotone maps, *JOTA* **66** (1990) 37–46.
- Konno93** KONNOV, I., Combined relaxation methods for finding equilibrium points and solving related problems, *Izvestiya VUZ, Matematika* **37** (1993) 44–51.
- LaMaS96** LABBÉ, M., MARCOTTE, P. ET SAVARD, G., A bilevel model of taxation and its application to optimal highway pricing, Publication 96–14, Centre de recherche sur les transports, Université de Montréal, 1996.
- LeHir93** LEMARÉCHAL, C. ET HIRRIART-URRUTY, J.-B., Convex analysis and minimization algorithms, volumes I et II, *Springer-Verlag*, Berlin, 1993.
- Leure93** LEURENT, F., Cost versus time equilibrium over a network, *EJOR* **71** (1993) 205–221.
- LuoPR96** LUO, Z-Q, PANG, J. S. ET RALPH, D., Mathematical programs with equilibrium constraints, Cambridge University Press, Cambridge, 1996.
- LuoTs92** LUO, Z-Q ET TSENG, P., On a global error bound for a class of monotone affine variational inequality problems, *Operations Research Letters* **11** (1992) 159–165.
- Lüthi85** LÜTHI, H.-J., On the solution of variational inequalities by the ellipsoid method, *Mathematics of Operations Research* **10** (1985) 515–522.
- MarDu87** MARCOTTE, P. ET DUSSAULT, J.-P., A note on a globally convergent Newton method for solving monotone variational inequalities, *Operations Research Letters* **6** (1987) 35–42.
- Marco85** MARCOTTE, P., A new algorithm for solving variational inequalities, with application to the traffic assignment problem', *Mathematical Programming* **33** (1985) 339–351.
- Marco87** MARCOTTE, P., Algorithms for the network oligopoly problem, *Journal of the Operational Research Society* **38** (1987) 1051–1065.
- Marco88** MARCOTTE, P., Un algorithme général de calcul de l'état d'équilibre d'un oligopole, *R.A.I.R.O. Recherche Opérationnelle* **22** (1988) 225–241.

- MarZh95** MARCOTTE, P., ET ZHU, D. L., Equilibria with infinitely many differentiated classes of customers, forthcoming in the proceedings of the *International conference on complementarity problems*, Johns Hopkins University, Baltimore, 1–5 Novembre 1995.
- MurSS82** MURPHY, F. H., SHERALI, H. D. ET SOYSTER, A. L., A mathematical programming approach for determining oligopolistic market equilibria, *Mathematical Programming* **24** (1982) 92–106.
- Nagur93** NAGURNEY, A., Network economics, A variational inequality approach, Kluwer, Dordrecht, 1993.
- Oweng82** OWEN, G., Game Theory. Second edition, *Academic Press*, Orlando, 1982.
- PangC82** PANG, J.-S. ET CHAN, D., Iterative methods for variational and complementarity problems, *Mathematical Programming* **24** (1982) 284–313.
- Patrik94** PATRIKSSON, M., The traffic assignment problem. Models and methods, *VSP Utrecht*, 1994.
- Robin79** ROBINSON, S. M., Generalized equations and their applications, part I: basic theory, *Mathematical Programming Study* **10** (1979) 128–141.
- Rosen65** ROSEN, J. B., Existence and uniqueness of equilibrium for concave N -person games, *Econometrica* **33** (1965) 520–534.
- SheSM83** SHERALI, H. D., SOYSTER, A. L. ET MURPHY, F. H., Stackelberg-Nash-Cournot equilibria; characterizations and computations, *Operations Research* **31** (1983) 253–276.
- Tseng91** TSENG, P., Applications of a splitting algorithm to decomposition in convex programming and variational inequalities, *SIAM Journal on Control and Optimization* **29** (1991) 119–138.
- ViCal94** VICENTE, L. N. ET CALAMAI, P. H., Bilevel and multilevel programming: a bibliography review, *JOTA* **5** (1994) 291–306.
- WuCFX95** WU, J. H., CHEN, Y., FLORIAN, M. ET XU, Y. W., A projection algorithm for the dynamic network equilibrium problem, Publication 95–64, Centre de recherche sur les transports, Université de Montréal, 1995.
- ZhuMa93** ZHU, D. L. ET MARCOTTE, P., Modified descent methods for solving the monotone variational inequality problem, *Operations Research Letters* **14** (1993) 111–120.
- ZhuMa94** ZHU, D. L. ET MARCOTTE, P., An extended descent framework for variational inequalities, *JOTA* **80** (1994) 349–366.

ZhuMa96 ZHU, D. L. ET MARCOTTE, P., On the existence of solutions to the dynamic user equilibrium problem, *preprint*.