الجمهوريـة الجزائـريـة الديمقراطيـة الشعبيـة
République Algérienne Démocratique et Populaire
وزارة التعليـم العالـي والبحـث العلمـي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

N° Réf :……………

Centre Universitaire
Abd Elhafid Boussouf Mila

Institut des Sciences et Technologie

Département de Mathématiques et Informatique

Mémoire préparé en vue de l'obtention du diplôme de Master

En: Mathématiques

Spécialité : Mathématiques Appliquées

Résolution d'un problème d'optimisation non linéaire : méthodes primales.

Préparé par : Bouadj Amira.
Benhaddad Amel.

Les nembres de Jury:

Benhabiless Hanane	M. A.A	C.U.Abd Elhafid Boussouf	Président
Fadel Wahida	M.A.A	C.U.Abd Elhafid Boussouf	Rapporteur
Bouzekria Fahima	M.A.B	C.U.Abd Elhafid Boussouf	Examinateur

Année Universitaire : 2019/2020


بسم الله الرحمن الرحيم

REMERCIENT

Nous remercions**''ALLAH''**Clément et Miséricordieux, tout puissant d'avoir guidé nos pas

vers les portes du savoir tout en illuminant notre chemin, et nous avoir donné suffisamment de courage et de persévérance pour mener à terme ce travail.

Nous tenons à remercier notre encadreur madame **Fadel wahida**, qui nous en couragé en

nous faisant part d'observations constructives et pour ses précieux conseils.

Nos vifs remerciements les membres de jury pour l'intérêt qu'ils ont porté à notre travail

en acceptant d'examiner. Leurs avis et leurs remarques ne feront qu'apporter des idées

nouvelles pour les études futures.

Nous remercions nos parents, pour tous les sacrifices qu'ils ont consentis pour nous per-mettre de suivre nos études dans les meilleures conditions et de nous avoir encouragé tout

au long de ces années.

En fin, Nous tenons à remercier nos familles et nos amis au département de mathématiques

et informatiques, et ailleurs, pour leur disponibilité et leur soutien.

DÉDICACE

Je dédie ce modeste travail

A cet source de tendresse, de patience, de courage et de générosité, à ma mère Zohra et mon pére Hocine.

A mes frères Abd Rahim, Abd Nour, et Rida et A ma sœurs Cherifa et son marie Tarek.

A ma belle famille.

A mes amis Ahlam,Sonia,Amel et Karima .

A mon binôme Amel.

A ma promo Mathématique appliquée 2020

A tous ceux qui, par un mot, m'ont donné la force de continuer...

Amira

DÉDICACE

Je dédie ce travail …

A mes chers parents ma mère Dalila et mon père

Mouhamed.

Aucune dédicace ne saurait exprimer mon respect, mon amour éternel et ma considération pour les sacrifices que vous avez consenti pour mon instruction et mon bien être.

Puisse Dieu, le Très Haut, vous accorder santé, bonheur et longue vie et faire en sorte que jamais je ne vous déçoive.

A mon cher frère hamza et mes chères sœurs Rokia, Kanza et Rayane.

Je vous souhaite une vie pleine de bonheur et de succès et que Dieu, le tout puissant, vous protège et vous garde.

A mes amis Sarra,Amira et faten .

A toute la famille Benhaddad.

A mon binôme Amira.

A ma promo Mathématique appliquée 2020

Amel

RÉSUME

L'objectif principal de ce mémoire est de **résoudre de problème d'optimisation non linéaire avec contraintes**, pour cela, en premiers lieu, nous avons donné des notions et des définitions de base. En deuxième lieu, il est essentiel de présenter **les conditions nécessaires et suffisantes d'optimalité dans le cas des contraintes d'égalités et d'inégalités** et encore **des contraintes mixte**, qui sont utiles par la suite. A la fin nous nous sommes focalisés sur **les méthodes primales** de résolution du problème étudié : **méthode de direction réalisable, méthode de Frank-Wolfe et méthode de gradient réduit**.
Mots clés : **optimisation non linéaire, méthode de direction réalisable, méthode de Frank-Wolfe, méthode de gradient réduit**.

ABSTRACT

The main objective of this memory is to solve non-linear optimization problem with constraints. for that, in the first place, we have given main and important definitions. secondly, it is necessarily necessary to present the necessary and sufficient conditions to type equality and inequality and extend these results to problems with mixed constraints, which are the main ones to solve the studied problem.

In the end we are focused on primals methods for problem solving, these methods is decompose to methods transferring the non-linear problem to a linear problem(achievable direction methods and Franke-Wolfe method). and a method transfer a problem with constraint has a problem without constraint (reduced gradient method).

Key words : **non-linear optimization, achievable direction method, Franke-Wolfe method and reduced gradient method**.

الملخص

الهدف الرئيسي من هذه المذكرة هو **حل المشاكل التحسينية غير الخطية بقيود**. من أجل هذا في المقام الأول تعريفات رئيسية و هامة.

في المقام الثاني، من الضروري للغاية تقديم الشروط اللازمة والكافية للتحسين لنوعي القيود بالمساواة وعدم المساواة، وتوسيع نطاق هذه النتائج مع قيود مختلطة والتي هي مهمة لحل المشكل المدروس.

في النهاية نحن ركزنا على الطرق الأولية لحل المشكلات،هذه الطرق تنقسم إلى طريقة تحويل برمجة غير خطية إلى برمجة خطية **(طريقة الاتجاه القابل للتحقيق، خواريزمية فرانك-وولف).**

وأخرى تحول المشكلة ذات شروط إلى مشكلة بدون شروط **(طريقة انخفاض التدرج)**.

الكلمات المفتاحية : البرمجة الغير خطية، طريقة الاتجاه القابل للتحقيق ، خواريزمية فرانك-وولف ، طريقة انخفاض التدرج

TABLE DES MATIÈRES

TABLE DES FIGURES

INTRODUCTION GÉNÉRALE

Dans la vie courante, nous sommes fréquemment confrontés à des problèmes "d'optimisation" plus ou moins complexes. Ces problèmes peuvent être exprimés sous la forme générale d'un "problème d'optimisation". L'optimisation peut être définie comme la science qui détermine la meilleure solution à certains problèmes mathématiquement définie, qui sont souvent des modèles de physique réelle. C'est une technique qui permet de "quantifier" les compromis entre des critères parfois non commensurables.
L'optimisation est une branche des mathématiques cherchant à analyser et à résoudre analytiquement ou numériquement les problèmes, qui consistent à déterminer le meilleur élément d'un ensemble, afin de maximiser ou minimiser un critère qualitatif qui mesure la qualité de la décision.
Le mot optimisation vient de latin « option » qui signifie la meilleur. L'optimisation joue un rôle très important en recherche opérationnelle (en économie et en microéconomie), en mathématiques appliquées, en analyse numérique, et en statistique.Il existe deux grandes classes d'optimisation, celle sans contraintes et avec contraintes.
Ce mémoire est structuré en l'introduction générale, trois chapitres et conclusion générale. Dans ce mémoire, on s'intéresse aux méthodes primales pour résoudre un problème d'optimisation non linéaire avec contraintes.
Dans le premier chapitre, nous avons donné le bagage mathématique nécessaire pour le traitement de problème considérés dans le mémoire.

Dans le deuxième chapitre, nous avons présenté les conditions d'optimalités (nécessaire et suffisantes), dans le cas des contraintes égalité et inégalité.
Dans le troisième chapitre, nous avons étudié trois méthodes primales de résolution appelés méthode de direction réalisable, méthode de gradient réduit et méthode de Frank-Wolfe.

CHAPITRE 1

GÉNÉRALITÉS

Nous présentons dans ce premier chapitre le bagage mathématique nécessaire pour le traitement du problème considéré dans ce mémoire.
Il s'agit de quelque définitions sur les ensembles, programmation mathématique, la différentiabilité et des notions de convexité.

1.1 Ensemble ouvert et ensemble fermé et borné

Définition 1.1.1. *(point intérieur)*
Soit $X \subseteq \mathbb{R}^n$ et $x \in X$:
On dit que x est intérieur à X s'il existe un voisinage de x contenu dans X.
D'une manière équivalente, x est intérieur à X s'il existe $\varepsilon > 0$ tel que :

$$\forall y \in X, \|x - y\| \leq \varepsilon.$$

L'ensemble des points intérieur de X est appelé intérieur de X et noté $Int(X)$ ou X^0.

[16]

Définition 1.1.2. *(Ensemble ouvert)*

Un ensemble $X \subset \mathbb{R}^n$ est un ouvert si et seulement si $X^0 = X$ [15].

Définition 1.1.3.

Soit $x \in \mathbb{R}^n$ et $v(x)$ un sous ensemble ouvert de $\mathbb{R}^n$ contient x, on dit que $v(x)$ est un voisinage de x.

Définition 1.1.4. *(L'adhérence)*

On dit que $x \in \mathbb{R}^n$ est un point d'adhérence d'un sous ensemble $X \in \mathbb{R}^n$:

$$\forall v(x), v(x) \cap X \neq \phi.$$

L'ensemble de tous les points d'adhérence d'un ensemble est appelé fermeture ou clôture noté $cl(X)$ ou $\overline{X}$.

Définition 1.1.5. *(Ensemble fermé)*

$X \subset \mathbb{R}^n$ est un ensemble fermé si et seulement si $X = \overline{X}$ [15].

Définition 1.1.6. *(Ensemble compact)*

Soit S un sous-ensemble d'un espace métrique. S est compact si pour toute suite $(x_k)_k$ d'élément de S, il existe une sous-suite convergente vers un élément de S. Si l'espace métrique est de dimension finie (comme $\mathbb{R}^n$), S est compact si et seulement si S est fermé et borné.

Définition 1.1.7. *(Ensemble borné)*

Soit $S \subset \mathbb{R}^n$ est un ensemble borné si : $\exists\ M > 0$ tel que $\forall\ x \in S$, $||x|| \leq M$.

[1]

1.2 Différentiabilité

Définition 1.2.1. *(Fonction continue)*

Soit une application $f : \mathbb{R}^n \to \mathbb{R}^m$ pour deux entiers $n \geq 1$ et $m \geq 1$. La fonction f est continue en $x \in \mathbb{R}^n$ si $\forall\ \varepsilon > 0$, $\exists\ \delta(x) > 0$ tel que :

$$\forall y, \textit{tel que} : ||y - x||_{\mathbb{R}^n} < \delta(x), ||f(y) - f(x)||_{\mathbb{R}^m} < \varepsilon.$$

La fonction f est continue dans $U \subset \mathbb{R}^n$ si f est continue en tout point de U [1].

Définition 1.2.2. *(Gradient)*

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction continue. La fonction notée $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ est appelée le gradient de f et est définie par :

$$\nabla f(x) = \begin{pmatrix} \dfrac{\partial f(x)}{\partial x_1} \\ \dfrac{\partial f(x)}{\partial x_2} \\ \vdots \\ \dfrac{\partial f(x)}{\partial x_n} \end{pmatrix} \tag{1.1}$$

Définition 1.2.3. *(Hessien)*

$$\nabla^2 f(x) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \cdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \dfrac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{pmatrix} \tag{1.2}$$

La matrice hessienne est toujours symétrique [1].

Définition 1.2.4. *(Dérivé directionnelle)*

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction continue. Soient $x \in \mathbb{R}^n$ et $d \in \mathbb{R}^n$.

La dérivée directionnelle de f en x dans la direction d est donnée par :

$$\lim_{\alpha \to 0} \frac{f(x + \alpha d) - f(x)}{\alpha}. \tag{1.3}$$

Si la limite existe. De plus, lorsque le gradient existe, la dérivée directionnelle est le produit scalaire entre le gradient de f et la direction d, c'est-à-dire :

$$\nabla f(x)^T d. \tag{1.4}$$

Définition 1.2.5. *(Fonction différentiable)*

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction continue. Si pour tout $d \in \mathbb{R}^n$ la dérivée directionnelle de f dans la direction d existe, alors la fonction f est dite différentiable [1].

1.3 Notions de convexité

La convexité joue un rôle très important dans la théorie classique de l'optimisation. Elle est un outil indispensable pour la recherche des conditions à la fois nécessaires et suffisantes d'optimalité.

Définition 1.3.1. *(Ensemble convexe)*
On dit que l'ensemble $X \subset \mathbb{R}^n$ est convexe si :

$$\forall (x, y) \in X \times X, \forall \lambda \in [0, 1], \lambda x + (1 - \lambda) y \in X.$$

Autrement dit, X est convexe s'il contient tout "segment" reliant deux quelconques de ses points [16].

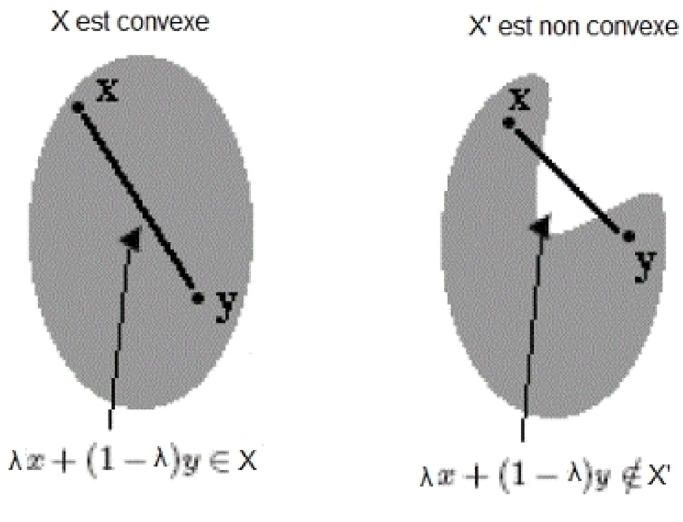

FIGURE 1.1 – L'ensemble convexe et non convexe

Définition 1.3.2. *(Fonction convexe)*

On dit que la fonction $f : X \longrightarrow \mathbb{R} \cup \{+\infty\}$ est convexe si X est convexe et si :

$$\forall (x, y) \in X \times X, \forall \lambda \in [0, 1], f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y). \tag{1.5}$$

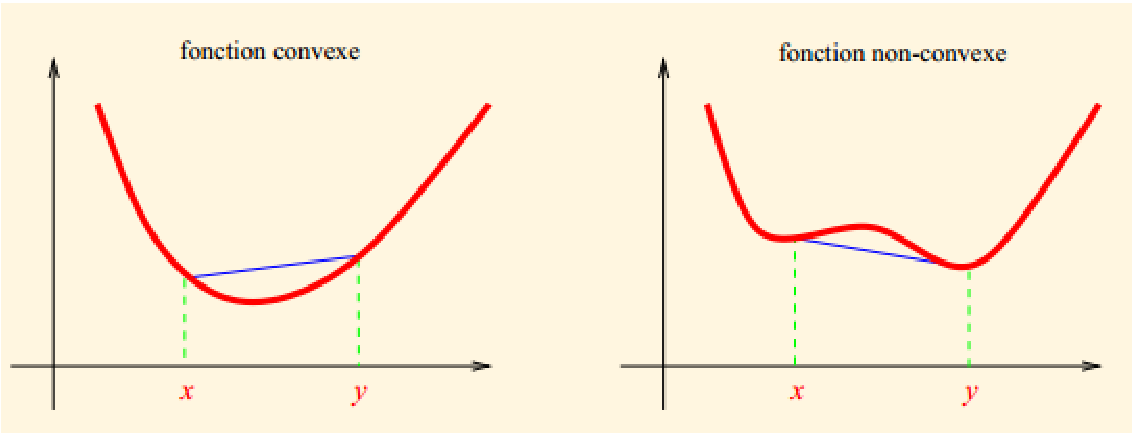

FIGURE 1.2 – Fonction convexe et non convexe

Définition 1.3.3. *(Fonction strictement convexe)*
On dit que la fonction $f : X \longrightarrow \mathbb{R} \cup \{+\infty\}$ est strictement convexe si X est convexe et si :

$$\forall (x, y) \in X \times X, avec, x \neq y, \forall \lambda \in]0, 1[, f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y). \tag{1.6}$$

Définition 1.3.4. *(Fonction concave)*
une fonction $f : X \longrightarrow \mathbb{R}$ est concave si $-f$ est une fonction convexe, c'est-à-dire si pour tout $(x, y) \in X \times X$ et pour tout $\lambda \in [0, 1]$, on a :

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y). \tag{1.7}$$

[1]

Test 1.3.1. *(Tests de convexité et de concavité)*

- *Supposons que f est une fonction deux fois dérivable d'une seule variable :*
- *f est convexe ssi :* $\dfrac{d^2 f(x)}{dx^2} \geq 0$, $\forall x$.

- *f est strictement convexe ssi :* $\dfrac{d^2 f(x)}{dx^2} > 0,\ \forall x.$
- *f est concave ssi :* $\dfrac{d^2 f(x)}{dx^2} \leq 0,\ \forall x.$
- *f est strictement concave ssi :* $\dfrac{d^2 f(x)}{dx^2} < 0,\ \forall x.$

Théorème 1.3.1. *(convexité par le gradient)*
Soit $f : X \to \mathbb{R}$ une fonction différentiable sur un ensemble convexe ouvert X. f est convexe sur X si et seulement si :

$$f(y) - f(x) \geq (y - x)^T f(x), \forall (x, y) \in X. \tag{1.8}$$

f est strictement convexe sur X si et seulement si :

$$f(y) - f(x) > (y - x)^T f(x), \forall (x, y) \in X. \tag{1.9}$$

Définition 1.3.5.
Un ensemble $X \subseteq \mathbb{R}^n$ est dit affine si et seulement si :

$$\forall (x, y) \in X, \lambda x + (1 - \lambda) y \in X, \forall \lambda \in \mathbb{R}.$$

Définition 1.3.6. *(Fonction coercive)*
Soient $X \subseteq \mathbb{R}^n$ et $f : X \longrightarrow \mathbb{R}$ une fonction à valeurs réelles. f est dite coercive si pour tout suite $(x_k)_k$ d'élément de X telle que $\|x_k\| \to \infty$ pour une norme quelconque, on a :

$$\lim_{k \to \infty} (x_k) = +\infty.$$

[1]

1.4 Programmation mathématique

1.4.1 Programme d'optimisation

En général, un programme mathématique est définit comme suit :

$$(P)\begin{cases} minf(x) \\ x \in X \subseteq \mathbb{R}^n \end{cases}.$$

Où $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ continue, $X \subseteq \mathbb{R}^n$ l'ensemble des contraintes présentées souvent comme :

$$X = \{x \in \mathbb{R}^n / h(x) = 0, g(x) \leq 0 \text{ et } x \in X\}.$$

avec :

$$h, g : \mathbb{R}^n \longrightarrow \mathbb{R}.$$

Définition 1.4.1. *(Solution réalisable)*
On appelle solution réalisable de (P) tout point x^0 vérifiant les contraintes (i.e, $x^0 \in X$) [7]
.

1.4.2 Minimum global et minimum local

Définition 1.4.2. *(Minimum local)*
Le vecteur $x^* \in X$ est appelé minimum local du problème (P) s'il existe $\varepsilon > 0$ tel que :

$$f(x^*) \leq f(x), \forall x \in X \textit{tel que} :\parallel x - x^* \parallel < \varepsilon.$$

Définition 1.4.3. *(Minimum local strict)*
Le vecteur $x^* \in X$ est appelé minimum local strict du problème (P) s'il existe $\varepsilon > 0$ tel que :

$$f(x^*) < f(x), \forall x \in X, x \neq x^* \textit{tel que} :\parallel x - x^* \parallel < \varepsilon.$$

Définition 1.4.4. *(Minimum global)*

Le vecteur $x^* \in X$ est appelé minimum global du problème (P) si :

$$f(x^*) \leq f(x), \forall x \in X.$$

Définition 1.4.5. *(Minimum global strict)*
Le vecteur $x^* \in X$ est appelé minimum global strict du problème (P) si :

$$f(x^*) < f(x), \forall x \in X, x \neq x^*.$$

[1]

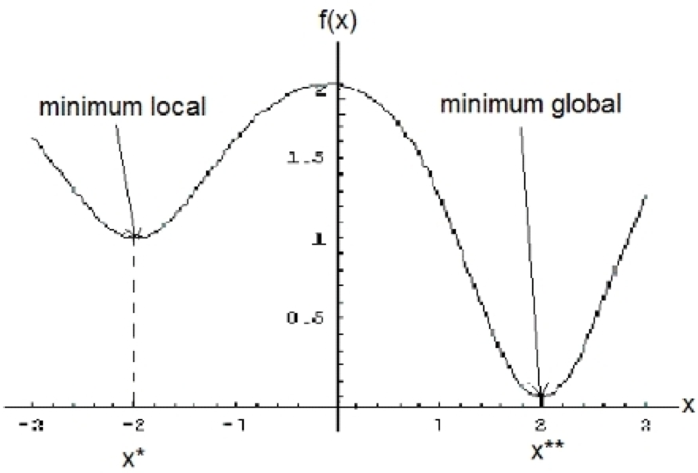

FIGURE 1.3 – Optimum local et global

1.5 Existence et l'unicité d'un optimum

Théorème 1.5.1. *(Weierstrass)*
Si X est compact dans $\mathbb{R}^n$ et f est continue sur X ; alors (P) admet au moins une solution optimale globale $x^* \in X$ c-a-d :

$$\exists x^* \in X, \forall x \in \mathbb{R}^n, f(x^*) \leqslant f(x).$$

Corollaire 1.5.1.

Si X est non vide et fermé, et f est continue et coercive sur X, alors (P) admet une solution optimale globale.

Théorème 1.5.2. *(Unicité)*
Soit $f : \mathbb{R}^n \to \mathbb{R}$ strictement convexe. Alors le problème (P) admet au plus une solution [7].

CHAPITRE 2

CONDITIONS D'OPTIMALITÉ POUR UN PROBLÈME D'OPTIMISATION NON LINÉAIRE AVEC CONTRAINTES

Dans ce chapitre, on s'intéresse au problème d'optimisation non linéaire avec contraintes, dont on va étudier les conditions d'optimalités pour les différents types des contraintes : égalités, inégalités et mixte.

2.1 Formulation d'un problème d'optimisation

Un programme mathématique est un problème d'optimisation sous contraintes dans $\mathbb{R}^n$ de la forme :

$$(P):\begin{cases}\min f(x)\\ \text{Sous les contraintes}\\ g_j(x)\leq 0, j=1,\cdots\cdots,p\\ h_i(x)=0, i=1,\cdots\cdots,q\\ x\in\mathbb{R}^n\end{cases}.$$

Où :

- $f:\mathbb{R}^n\to\mathbb{R}$ est une fonction de plusieurs variables $x=(x_1,\cdots,x_n)$ a valeurs réelles. cette fonction est appelée fonction **Objectif**.
- $g:\mathbb{R}^n\to\mathbb{R}^p$ est une fonction de plusieurs variables $x\in\mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$: elle a p contraintes. On peut écrire :

$$g(x)=(g_1(x),\cdots,g_p(x)).$$

- La fonction $g(x)$ représente **Les contraintes inégalités**.

- $h:\mathbb{R}^n\to\mathbb{R}^q$ est une fonction de plusieurs variables $x\in\mathbb{R}^n$ à valeurs dans $\mathbb{R}^q$ elle a q contraintes. On peut écrire

$$h(x)=(h_1(x),\cdots,h_q(x)).$$

- La fonction $h(x)$ représente **Les contraintes égalités** [5].

Définition 2.1.1. ***(Ensemble admissible)***
On appelle ensemble admissible du problème, l'ensemble X définie par :

$$X=\{x\in\mathbb{R}^n : g(x)\leq 0, h(x)=0\}.$$

[8]

Définition 2.1.2. ***(Direction admissible)***
Soit $X\subset\mathbb{R}^n$ un ensemble et $x^*\in X$. On dit que $d\in\mathbb{R}^n$ est une ***direction admissible*** *pour $x^*\in X$ s'il existe $\lambda'>0$ tel que : $x^*+\lambda d\in X,\forall\lambda\in[0,\lambda']$.*

Lemme 2.1.1.

Soit $\Omega \subset \mathbb{R}^n$ un ouvert, $X \subset \Omega$, $f : \Omega \to \mathbb{R}$ une fonction de classe C^1 et $x^* \in X$ un point de minimum relatif de f sur X. Soit $d \in \mathbb{R}^n$ une direction admissible pour $x^* \in X$, Alors : $\langle \nabla f(x^*), d\rangle \geqslant 0$.

Démonstration 2.1.1.

Soit V un voisinage de x^* en $\mathbb{R}^n$ tel que :

$$f(x) \geqslant f(x^*), \forall x \in X \cap V. \tag{2.1}$$

Soit $\lambda^* > 0$ tel que :

$$x^* + \lambda d \in V, \forall \lambda \in [-\lambda^*, \lambda^*].$$

(pour $d \neq 0$ il suffit de prendre $\lambda^* = \dfrac{r}{2||d||}$ où $r > 0$ est tel que $B(x^*, r) \subset V$ pour $d = 0$ tout $\lambda^{'}$ convient).

D'autre part on a :

$$x^* + \lambda d \in X, \forall \lambda \in [0, \lambda^{'}].$$

Où $\lambda^{'}$ est comme dans la définition de la direction admissible on déduit alors de (2.1)

$$f(x^* + \lambda d) - f(x^*) \geqslant 0, \forall t \in [0, min\lambda^{'}, \lambda^*].$$

En divisant par λ et en passant à la limite pour $\lambda \to 0, \lambda > 0$ on obtient :

$$\frac{\partial f}{\partial d}(x^*) \geq 0.$$

Ce qui est exactement l'inégalité demandée [9]. ■

Théorème 2.1.1. *(Direction admissible : cas linéaire)*

[1] Soit le problème d'optimisation $minf(x)$ sous contraintes $Ax = b$ et $x \geq 0$ et soit un point admissible x^* . Une direction d est admissible en x^* si et seulement si

1) $Ad = 0$.

2) $d_i \geq 0$ *si* $x_i^* = 0$.

Définition 2.1.3. ***(Direction de descente)***

Un vecteur d est une direction de descente s'il existe τ tel que :

$$f(x+td) < f(x), t \in [0, \tau].$$

[10]

Définition 2.1.4. ***(cône tangent)***

Soit $x \in X$ le cône tangent (ou cône des directions admissibles) à X en x, noté $T_x(X)$, est l'ensemble des vecteurs $d \in \mathbb{R}^n$ tel qu'il existe une suite $(d_n)_{n\in\mathbb{N}}$ de limite d et une suite $(\lambda_n)_{n\in\mathbb{N}}$ de réels strictement positifs de limite nulle telle que :

$$x + \lambda_n d_n \in X.$$

Autrement dit le cône admissible est l'ensemble des directions admissibles dans X au point x, aussi que les limites de ces directions. [12].

Définition 2.1.5. ***(Ensemble des contraintes actives)***

$$\Omega = \left\{ x \in \mathbb{R}^n \begin{array}{l} h_i(x) = 0, i = 1 \cdots q \\ g_j(x) \leq 0, j = 1 \cdots p \end{array} \right\} \subseteq \mathbb{R}^n$$

- *Pour tout $x \in \mathbb{R}^n$, l'ensemble des contraintes actives $I(x)$ est l'ensemble des indices des contraintes inégalité satisfait à égalité en x :*

$$I(x) = \{ j \in I : g_j(x) = 0 \}.$$

- ***Attenttion*** *: Même si les contraintes égalité sont toujours actives, elles ne sont pas représentées par $I(x)$ [4].*

2.2 Conditions d'optimalité

Théorème 2.2.1.

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une application différentiable et x^* un minimum local du problème (P) :

$$\forall d \in T_{x^*}(X), \langle \nabla f(x^*), d \rangle \geq 0. \tag{2.2}$$

[12]

Preuve 2.2.1.

Soit d une direction admissible en x^*. Il existe alors $\bar{\alpha} > 0$ tel que $x^* + \alpha d \in X$ pour tout $\alpha \in [0, \bar{\alpha}]$.

Considérons la fonction $\varphi(\alpha) = f(x^* + \alpha d)$.

Comme x^* est un minimum local, il existe $\epsilon > 0$ tel que :

$$f(x^*) \leq f(x), \forall x \in B(x^*, \epsilon) \cap X.$$

Considérons

$$X_{\overline{\alpha}} = \{x^* + \alpha d, \alpha \in [0, \overline{\alpha}]\}.$$

$\forall x \in X_{\overline{\alpha}} \cap B(x^*, \varepsilon) \cap X$, *on aura*

$$f(x^*) \leq f(x).$$

Soit

$$[0, \tilde{\alpha}] = \{\alpha \in [0, \overline{\alpha}] / x^* + \alpha d \in X_{\overline{\alpha}} \cap B(x^*, \varepsilon) \cap X\},$$

et

$$\varphi(\alpha) \geq \varphi(0), \forall \alpha \in [0, \tilde{\alpha}].$$

D'où $\alpha^* = 0$ est un minimum local de $\varphi(\alpha)$ sur $[0, \tilde{\alpha}]$.

On a

$$\varphi(\alpha) = \varphi(0) + \alpha \varphi'_+(0) + \sigma(\alpha)$$

où

$$\varphi_+^{'}(0) = \lim_{\alpha \to 0^+} \frac{\varphi(\alpha) - \varphi(0)}{\alpha}.$$

Si $\varphi_+^{'}(0)$ *était négative, alors pour un nombre* α *assez petit, on aurait en*

$$\varphi(\alpha) < \varphi(0).$$

Ce qui contredirait le fait que φ *possède un minimum local en* $\alpha = 0$. *On a donc bien* $\varphi_+^{'}(0) \geq 0$.

Par conséquent, comme

$$\varphi^{'}(\alpha) = [\nabla f(x^* + \alpha d)]^T d,$$

on obtient

$$\varphi_+^{'}(0) = [\nabla f(x^*)]^T d \geq 0.$$

la relation (2.2) est alors démontrée [14] . ■

2.2.1 Condition nécessaire d'optimalité du premier ordre (condition de Lagrange)

Considérons un problème avec contraintes uniquement d'égalité :

$$(PCE) \begin{cases} \min_{x \in \mathbb{R}^n} f(x) \\ h_i(x) = 0, i = 1, \cdots, q \end{cases}$$

On note :

$$X = \{x \in \mathbb{R}^n : h_i(x) = 0, i = 1, \cdots, q\}.$$

Le domaine admissible. On cherche les conditions d'optimalité associées au problème (PCE). La première étape (la plus difficile) consiste à expliciter le cône tangent.

Étape 1 : Qualification des contraintes : comment calculer le cône tangent à X

Définition 2.2.1. *(Qualification des contraintes d'égalité)*

La contrainte $h(x) = 0$ du problème (PCE) est dite qualifiée en $x \in X$ si h différentiable en x et si :

$$T_x(X) = \{d \in \mathbb{R}^n / \forall i = 1, \cdots, q, \langle \nabla h_i(x), d \rangle = 0\} = kerDh(x).$$

[12]

- Comment interpréter cette définition ? Le cône tangent ne dépend que de l'ensemble X et pas de la fonction h choisie pour le représenter. Au contrainte $KerDh(x)$ dépend directement de h. La qualification des contraintes peut être vue comme un critère.
- permettant de sélectionner de bonnes représentations de X par h.
  En pratique, cette définition de qualification des contraintes est difficile à vérifier ; une hypothèse couramment utilisée pour garantir la qualification des contraintes est qu'elles soient régulières.

Définition 2.2.2. *(contraintes régulières)*

Soit $x \in X$. Supposons h de classe C^1 au voisinage de x. On dit que les contraintes du problème (PCE) sont régulières en $x \in X$, ou que $x \in X$ est régulier, si la jacobienne $Dh(x)$ des contraintes est surjective.

Sous l'hypothèse que h est de classe C^1 au voisinage de $x \in X$, on rappelle que la matrice Jacobienne $Dh(x)$ de h en x est définie par :

$$Dh(x)^T = [\nabla h_1(x)^T / , \cdots , / \nabla h_q(x)^T].$$

La matrice $Dh(x)$ étant de taille $q \times n$. On a alors :

Proposition 2.2.1.

Le point $x \in X$ est régulier ssi les vecteurs $\nabla h_i(x), i = 1, ..., q$ sont linéairement indépendants.

Proposition 2.2.2. *(Condition suffisante de qualification des contraintes)*

Soit $x \in X$. Si x est régulier, alors les contraintes d'égalité $h(x) = 0$ sont qualifiées au *point x i.e. :*

$$T_x(X) = KerDh(x) = \left\{d \in \mathbb{R}^n / \forall i = 1, \cdots, q, \langle \nabla h_i(x), d\rangle = 0\right\}.$$

[12]

Étape 2 : Condition nécessaire d'optimalité

Théorème 2.2.2. *(Théorème de Lagrange. Condition Nécessaire d'optimalité du premier ordre)*
Soit $f : \mathbb{R}^n \to \mathbb{R}$ et $h : \mathbb{R}^n \to \mathbb{R}^q$ différentiables en x^*. Supposons la contrainte $h(x) = 0$ qualifiée en $x^* \in X$. Si x^* est un point de minimum local de f sur X, alors il existe des réels $\lambda_1^*, \cdots, \lambda_q^*$ tels que :

$$\nabla f(x^*) + \sum_{i=1}^{q} \lambda_i^* \nabla h_i(x^*) = 0.$$

Le vecteur $\lambda^* = [\lambda_1^*, \cdots, \lambda_q^*]$ est appelé multiplicateur de Lagrange et est déterminé de façon unique [12].

Preuve 2.2.2.
Considérons une courbe $x(t)$ définie pour $t \geq 0$ vérifiant :

$$\begin{cases} x(t) \in X, \forall t \in [-\alpha, \alpha], \alpha > 0. \\ x(0) = x^*. \end{cases}.$$

Puisque $x(t) \in X$ on a $h_i(x(t)) = 0$ pour $1 \leq i \leq q$ et on peut écrire que :

$$\frac{d}{dt} h_i(x(t)) = \nabla h_i(x(t))^T \dot{x(t)} = 0, 1 \leq i \leq q.$$

On a :

$$f(x(0)) \leq f(x(t)), \forall t \in [-\alpha, \alpha].$$

donc nécessairement :

$$\frac{d}{dt} f(x(t))|_{t=0} = \nabla f(x^*)^T \dot{x(0)} = 0.$$

ce qui signifie que $\nabla f(x^*)$ se trouve dans l'orthogonal de $T(x^*)$ le plan tangent à X en

x^*. Si l'on utilise l'équivalence :

$$T(x^*) = ker\nabla h(x^*)^T \Leftrightarrow T(x^*)^\perp = Im\nabla h(x^*).$$

Il existe donc vecteur $\lambda \in \mathbb{R}^q$ tel que :

$$\nabla f(x^*) = -\nabla h(x^*)\lambda.$$

*L'unicité résulte du fait que $\nabla f(x^*)$ est de **rang** q [6].* ■

Terminologie

- Le vecteur λ^* est aussi appelé solution duale du problème (PCE), x^* solution primale de (PCE) et (x^*, λ^*) solution primal-duale de (PCE).
- On appelle point stationnaire du problème (PCE) tout point x vérifiant les conditions nécessaires d'optimalité du premier ordre :

$$\begin{cases} \nabla f(\overline{x}) + \sum\limits_{i=1}^{q} \overline{\lambda} \nabla h_i(\overline{x}) = 0. \\ h(\overline{x}) = 0. \end{cases}$$

Pour un certain multiplicateur $\overline{\lambda} \in \mathbb{R}^q$.

Réécriture des conditions d'optimalité à l'aide du Lagrangien

En pratique, on retrouve les conditions d'optimalité du problème (PCE) en introduisant le Lagrangien du problème (PCE) :

$$L(x, \lambda) = f(x) + \sum_{i=1}^{q} \lambda_i h_i(x) = 0, x \in \mathbb{R}^n, \lambda \in \mathbb{R}^q.$$

Sous hypothèse de qualification des contraintes, les conditions d'optimalité du problème (PCE) s'écrivent alors :

$$\begin{cases} \nabla_x L(x, \lambda) = 0 \\ h(x) = 0 \end{cases}$$

Ou encore : $\nabla_{(x,\lambda)} L(x, \lambda) = 0$ [12].

2.2.2 Conditions nécessaire et suffisante d'optimalité du second ordre

Nous admettrons sans démonstration que les conditions nécessaire et suffisante d'optimalité du second ordre s'obtiennent en étudiant le Hessien du Lagrangien calculé selon la composante x uniquement :

Condition nécessaire d'optimalité locale du second ordre :

$$\forall d \in T_x(X), \left\langle \nabla^2_{xx}[L](x,\lambda)d, d \right\rangle \geq 0.$$

Condition suffisante d'optimalité locale du second ordre :

$$\forall d \in T_x(X), d \neq 0, \left\langle \nabla^2_{xx}[L](x,\lambda)d, d \right\rangle > 0.(CS1)$$

– Remarquons que si la matrice hessienne $\nabla^2_{xx}[L]$ est définie positive, alors il est inutile de calculer le cône tangent au point x puisque dans ce cas :

$$\forall d \in \mathbb{R}^n, d \neq 0, \left\langle \nabla^2_{xx}[L](x)d, d \right\rangle > 0.$$

Et donc

$$\forall d \in T_x(X), d \neq 0, \left\langle \nabla^2_{xx}[L](x)d, d \right\rangle > 0.$$

La condition suffisante d'optimalité du second ordre est automatiquement satisfaite.

Une autre Condition suffisante d'optimalité locale du second ordre :

$$\forall, d \in \mathbb{R}^n, d \neq 0, \left\langle \nabla^2_{xx}[L](x,\lambda)d, d \right\rangle > 0.(CS2)$$

On a :

$$(CS2) \Longrightarrow (CS1).$$

[12]

2.2.3 Exemple

Soit le problème :

$$\begin{cases} minf(x,y) = x^2 + 2y^2. \\ h(x,y) = -x - y - 3 = 0. \end{cases}.$$

Qualification des contraintes :

$$\nabla h(x,y) = \begin{pmatrix} \frac{\partial h}{\partial x} \\ \frac{\partial h}{\partial y} \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Donc $\nabla h(x,y)$ est libre alors (x,y) est régulier et la contrainte est qualifiée.

On peut remarquer que f et h sont convexes. f est même strictement convexe.

$$L(x,y,\lambda) = f(x,y) + \lambda h(x,y)$$

Condition nécessaire du premier ordre :

$$\begin{cases} \nabla_x L(x,y,\lambda) = 0 \\ \nabla_\lambda L(x,y,\lambda) = 0 \end{cases} \Leftrightarrow \begin{cases} \nabla f(x,y) + \lambda \nabla h(x,y) = 0 \\ \lambda h(x,y) = 0 \\ \lambda \geq 0 \end{cases}.$$

$$2x - \lambda = 0 \tag{2.3}$$

$$4y - \lambda = 0 \tag{2.4}$$

$$\lambda(-x - y - 3) = 0 \tag{2.5}$$

$$\lambda \geq 0 \tag{2.6}$$

On a $\lambda \geq 0$.

• Si $\lambda = 0$:

$$\begin{cases} 2x = 0 \Rightarrow x = 0 \\ 4y = 0 \Rightarrow y = 0 \end{cases}$$

ce qui donne $x = 0, y = 0$ point ne vérifie pas la contrainte (2.4) par conséquent ce cas est impossible.

• Si $\lambda \neq 0$

$$\begin{cases} 2x - \lambda = 0 \Rightarrow \lambda = 2x \\ 4y - \lambda = 0 \Rightarrow \lambda = 4y \end{cases} \Leftrightarrow x = 2y$$

on remplace le x :

$$-x - y - 3 = -2y - y = -3y - 3 = 0 \Rightarrow y = -1.$$

donc $(x^*, y^*) = (-2, -1)$ la solution admissible et $\lambda = 4$.

Condition suffisante :

$$d^T \nabla^2_{xy} L(x^*, y^*, \lambda^*) \geq 0.$$

$$\nabla^2_{xy} L(x^*, y^*, \lambda^*) = \begin{pmatrix} \dfrac{\partial^2 L}{\partial x^2} & \dfrac{\partial^2}{\partial x \partial y} \\ \dfrac{\partial^2}{\partial y \partial x} & \dfrac{\partial^2 L}{\partial y^2} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} = 8 \geq 0$$

donc définie positif.

Donc $(x^*, y^*) = (-2, -1)$ est un point min local de problème (PCE).

2.2.4 Condition nécessaire d'optimalité du premier ordre (condition Karuch-Kuhn-Tuker (KKT))

Intéressons nous maintenant aux problèmes d'optimisation avec contraintes d'inégalité :

$$(PCI) \begin{cases} \min_{x \in \mathbb{R}^n} f(x) \\ g_j(x) \leq 0, j = 1, \cdots, p \end{cases}$$

On note :

$$X = \{x \in \mathbb{R}^n : g_j(x) \leq 0, j = 1, \cdots, p\}.$$

- On dit que la contrainte g est active en x^* si $g(x^*) = 0$.
- On appelle contraintes saturées en x^* l'ensemble des indices j tel que : $g_j(x^*) = 0$.

On note :

$$I(x^*) = \{j/g_j(x^*) = 0, j \in I(x^*)\}. \tag{2.7}$$

Étape 1 : **Qualification des contraintes d'inégalité : comment calculer le cône tangent à** X

Définition 2.2.3. *(Qualification des contraintes d'inégalités)*
La contrainte $g(x) \leq 0$ du problème (PCI) est dite qualifiée au point $x \in X$ si g est différentiable en x et si :
$T_x(X) = \{d \in \mathbb{R}^n$ *pour tout j actif en x,* $\langle \nabla g_j(x), d\rangle \leq 0.$

L'interprétation est la même que dans le cas des contraintes d'égalité : la qualification des contraintes peut être vue comme un critère permettant de sélectionner de "bonnes" représentations de X par g.

Remarque 2.2.1.
La plupart des algorithmes échouent lorsqu'ils doivent résoudre un problème dont les contraintes ne sont pas qualifiées en la solution. Il est dans ce cas préférable de changer la description de l'ensemble X des contraintes avant de chercher à résoudre le problème. Une fois de plus, cette condition est difficile à vérifier en pratique et couramment remplacée par des conditions suffisantes de qualification des contraintes :

Proposition 2.2.3. *(Condition Suffisante de qualification des contraintes)*
Les contraintes $g_j(x) \leq 0, j = 1, ..., p$, sont qualifiées en un point $x \in X$ s'il existe un vecteur $d \neq 0$ tel que pour toute contrainte active $g_j(x) = 0$, on ait :

- *soit $g_j(x)$ est affine et : $\langle \nabla g_j(x), d\rangle = 0$.*
- *soit : $\langle \nabla g_j(x), d\rangle < 0$.*

Proposition 2.2.4.
Les contraintes $g_j(x) \leq 0, j = 1, .., p$ sont qualifiées au point $x \in X$ si les gradients des contraintes actives en x sont linéairement indépendants.

A noter que, dans le cas convexe, ces conditions de qualification des contraintes s'écrivent plus simplement sous la forme :

Proposition 2.2.5.

Supposons les fonctions $g_j(x), j = 1, ..., p$, convexes. Les contraintes $g_j(x) \leq 0, j = 1, ..., p$ sont qualifiées en tout point de X s'il existe un point $x_0 \in X$ tel que pour tout $j = 1, ..., p$, on a : [12]

- *soit $g_j(x_0) < 0$.*
- *soit $g_j(x)$ est affine.*

Étape 2 : condition nécessaire d'optimalité

Si x est solution du problème (PCI) avec contraintes inégalités et si ces contraintes sont qualifiées en x au sens de la Proposition (2.2.3), alors la condition nécessaire d'optimalité s'écrit :

$\forall\ d \in \mathbb{R}^n$, $\forall j$ actif en x, $\langle \nabla g_i(x), d \rangle \leq 0 \Rightarrow \langle \nabla f(x), d \rangle \geq 0$.

En utilisant le lemme de Farkas (ici admis) :

Lemme 2.2.1. *(Lemme de Farkas)*

Soient $d_1, \ldots, d_l$ un ensemble de vecteurs de $\mathbb{R}^n$ et $u \in \mathbb{R}^n$. Alors les propositions suivantes sont équivalentes :

i. $\forall x \in \mathbb{R}^n, [\forall j = 1, \ldots, p, \langle d_j, x \rangle \leq 0] \Rightarrow \langle u, x \rangle \geq 0.$

ii. Il existe des réels $\lambda_1, \ldots, \lambda_p$ positifs tels que : $u = -\sum\limits_{j=1}^{p} \lambda_j d_j.$

On obtient immédiatement la formulation suivante :

Théorème 2.2.3. *(Condition nécessaire d'optimalité du 1^{er} ordre sous contraintes d'inégalités)*

Soit $x^* \in X$. Supposons f et g différentiables en x^* et les contraintes qualifiées en x^*. Si x^* est un point de minimum local de f sur X, alors il existe des réels positifs ou nuls $\lambda_1^*, \ldots, \lambda_p^*$ tels que :

$$\nabla f(x^*) + \sum_{j=1}^{p} \lambda_j^* \nabla g_j(x^*) = 0 \tag{2.8}$$

$$\lambda_j^* \geq 0, j = 1, \ldots, p \tag{2.9}$$

$$\lambda_j^* g_j(x^*) = 0. \tag{2.10}$$

[12]

Preuve 2.2.3.

Les relations (2.8), (2.9) *des résultats directe du théorème du Lagrange il suffit de prendre* $\lambda_j^* = 0$ *pour* $j \notin I(x^*)$ *on peut ensuite montrer* (2.9) *par l'absurde.*
Supposons qu'il existe $k \in I(x^*)$ *tq* $\lambda_k < 0$.
On définit $S_k = \{x/g_j(x) = 0, j \in I(x^*), j \neq k\}$.
On définit $d \in \mathbb{R}^n$ *tel que :*

$$\begin{cases} \nabla g_j(x^*)^T d = 0, j \in I(x^*), j \neq k. \\ \nabla g_k(x^*)^T d = -1. \end{cases}$$

Alors d *est une direction admissible en* x^* *puisque :*

$$\nabla g_j(x^*)^T d \leq 0, j \in I(x^*).$$

et que x^* *est un point régulier, il existe donc une courbe* $x(t) \in S_k$ *et vérifiant de plus* $x(t) \in X$, *pour* $t \in [-\alpha, \alpha]$ *tel que :* $x^*(0) = y$.
On a donc :

$$\begin{aligned} \frac{\partial}{\partial t} f(x(t))|_{t=0} &= \nabla f(x^*)^T d. \\ &= -\sum_{j=1}^{p} \lambda_j^* \nabla g_j(x^*)^T d. \\ &= -\lambda_k \nabla g_k(x^*)^T d. \\ &= \lambda_k < 0. \end{aligned}$$

Ce qui est impossible car f *est minimum en* x^* *[6].* ■

Terminologie :

- Vecteur λ^* est aussi appelé solution duale du problème (PCI), x^* solution primale de (PCI) et (x^*, λ^*) solution primale-duale de (PCI).

- On appelle point stationnaire du problème (PCI) tout point $\overline{x}$ vérifiant les conditions nécessaires d'optimalité du premier ordre :

$$\begin{cases} \nabla f(\overline{x}) + \sum\limits_{j=1}^{p} \lambda_j \nabla g_j(\overline{x}) & = & 0 \\ \lambda_j & \geq & 0, j = 1, \ldots, p. \\ \lambda_j g_j(\overline{x}) & = & 0, j = 1, \ldots, p. \end{cases}$$

  Pour un certain multiplicateur $\lambda \in \mathbb{R}^p$.

- Les relations : $\lambda_j g_j(\overline{x}) = 0$ sont appelées relations de complémentarité et signifie : Soit : $\lambda_j = 0$, soit : $g_j(\overline{x}) = 0$.
  Ces relations sont trivialement satisfaites par toute contrainte j active en $\overline{x}$ et indiquent que pour toute contrainte inactive, le multiplicateur λ_j correspondant est nul. Cela signifie que toute contrainte inactive à l'optimum aurait pu être relaxée.

Réécriture des conditions d'optimalité à l'aide du Lagrangien

En pratique, on retrouve les conditions d'optimalité du problème (PCI) en introduisant le Lagrangien associé au problème (PCI) :

$$L(x, \lambda) = f(x) + \sum_{j=1}^{p} \lambda_j g_j(x), x \in \mathbb{R}^n, \lambda \in \mathbb{R}^p.$$

Sous hypothèse de qualification des contraintes, les conditions nécessaires d'optimalité du premier ordre du problème (PCI) s'écrivent alors :

$$\begin{cases} \nabla_x L(x,\lambda) & = & 0. \\ \lambda_j g_j(x) & = & 0, j = 1,\dots,p. \\ \lambda_j & \geq & 0, j = 1,\dots,p. \end{cases} .$$

[12]

2.2.5 Condition suffisante d'optimalité du second ordre

Théorème 2.2.4.

Soit $x^* \in \mathbb{R}^n$ *et* $\lambda^* \in \mathbb{R}^p$ *vérifiant les conditions :*

$$\begin{cases} \nabla_x L(x^*,\lambda^*) & = & \nabla f(x^*) + \sum\limits_{j=1}^{p} \lambda_j^* \nabla g_j(x^*) = 0. \\ g_j(x^*) & \leq & 0. \\ \lambda_j^* & \geq & 0, j = 1,\dots,p. \\ \lambda_j^* \nabla g_j(x^*) & = & 0, j = 1,\dots,p. \\ d^T \nabla_{xx}^2 L(x^*,\lambda^*) d & \geq & 0, \forall d \in T_{x^*}(X), d \neq 0. \end{cases}$$

Où on a noté $T_{x^*}(X)$ *le plan tangent en* x^* *[6].*

2.2.6 Condition nécessaire d'optimalité de 1^{er} ordre de type mixte (égalité et inégalité)

Considérons un problème d'optimisation avec des contraintes d'égalité et d'inégalité :

$$(P) \begin{cases} min f(x). \\ h_i(x) = 0, i = 1,\dots,q. \\ g_j(x) \leq 0, j = 1,\dots,p. \end{cases} .$$

Définition 2.2.4.

Les contraintes du problème (P) *sont dites qualifiées au point* $x \in X$ *si* g *et* h *sont* *différentiables au point x et si :*

$$T_x(X) = \left\{ \begin{array}{c} d \in \mathbb{R}^n, \forall i = 1, \ldots, q, \langle \nabla h_i(x), d \rangle = 0. \\ \textit{et pour toute contrainte } j \textit{active} : \langle \nabla g_i(x), d \rangle \leq 0 \end{array} \right\}.$$

Proposition 2.2.6. *(Condition Suffisante de qualification des contraintes)*
Les contraintes du problème (P) sont qualifiées au point $x \in X$ si les gradients des contraintes actives en x :
$\{\nabla h_i(x); i = 0, \ldots, q\} \cup \{\nabla g_j(x), j \in \{0, \ldots, p\}$ *active*$\}$ *sont linéairement indépendants.*

• On introduit le Lagrangien associé au problème (P) :
$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{q} \lambda_i h_i(x) + \sum_{j=1}^{p} \mu_j g_j(x), x \in \mathbb{R}^n, \lambda \in \mathbb{R}^q, \mu \in \mathbb{R}^p$ [12].

Définition 2.2.5. *(point régulier)*
On dit qu'un élément x^* de $\mathbb{R}^n$ est régulier pour les contraintes h et g :

• *S'il est réalisable $h(x^*) = 0, g(x^*) \leq 0$.*

• *Les vecteurs $\nabla h_i(x^*), i = 1, \ldots, q$ sont indépendant.*

• *Si on trouver $d \neq 0 \in \mathbb{R}^n$ tel que :*
$\langle \nabla h_i(x^*), d \rangle = 0, i = 1, \ldots, q$ *et* $\langle \nabla g_i(x^*), d \rangle < 0, \forall i \in I(x^*)$.

Théorème 2.2.5. *(condition nécessaire)*
On suppose que f, g, h sont de classe C^1. Soit x une solution du problème (P). On suppose que x est régulier pour les contraintes h et g. Alors il existe $\lambda^* = (\lambda_1^*, \ldots, \lambda_q^*) \in \mathbb{R}^q$; et $\mu^* = (\mu_1^*, \ldots, \mu_p^*) \in \mathbb{R}^p$ tels que :

$$\begin{cases} \nabla_x L(x, \lambda^*, \mu^*) = 0 \\ h_i(x) = 0, i = 1, \ldots, q \\ \mu_j^* g_j(x) = 0, j = 1, \ldots, p \\ \mu_j^* \leq 0, j = 1, \ldots, p \end{cases}$$

[16]

2.2.7 Exemple

On considère le problème suivant :

$$\begin{cases} \min 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2 \\ \qquad x_1^2 + x_2^2 \leq 5 \\ \qquad 3x_1 + x_2 \leq 6 \end{cases}$$

Dans sous cas $x = (x_1, x_2)$, il n'y pas de contraintes en égalité ($p = 0, /q = 2$) et $f(x) = 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2$.

$g = (g_1, g_2)$ avec $g_1(x_1, x_2) = x_1^2 + x_2^2 - 5$ et $g_2(x_1, x_2) = 3x_1 + x_2 - 6$.

vérifiée que les contraintes sont qualifiées :

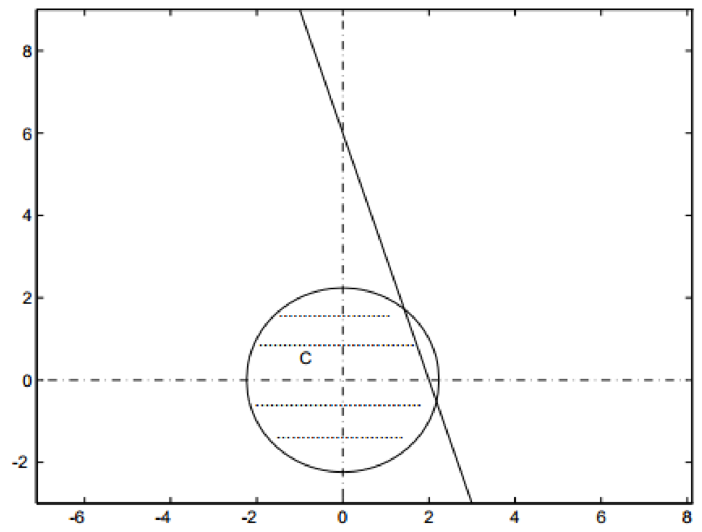

FIGURE 2.1 – ensemble C des contraintes

$$\begin{pmatrix} \dfrac{\partial g_1}{\partial x_1} \\ \dfrac{\partial g_1}{\partial x_2} \end{pmatrix} + \begin{pmatrix} \dfrac{\partial g_2}{\partial x_1} \\ \dfrac{\partial g_2}{\partial x_2} \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} + \begin{pmatrix} 3 \\ 1 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{cases} 2x_1 + 3 \neq 0 \\ 2x_2 + 1 \neq 0 \end{cases} \Rightarrow$$

$$\begin{cases} x_1 \neq \dfrac{-3}{2} \\ x_2 \neq \dfrac{-1}{2} \end{cases}$$

- Les contraintes sont qualifiées si et seulement $(x_1, x_2) \neq (\frac{-3}{2}, \frac{-1}{2})$.

-On peut remarquer que f, h et g sont convexes. f est même strictement convexe.

$$L(x, \mu_1, \mu_2) = f(x) + \mu_1 g_1(x) + \mu_2 g_2(x)$$

Condition nécessaire de premier ordre :

Écrivons les conditions de KKT a priori. On vérifiera la régularité du point obtenu a posteriori.

On obtient

$$\mu_1 \geq 0, \mu_2 \geq 0, \tag{2.11}$$

$$x_1^2 + x_2^2 \leq 5, 3x_1 + x_2 \leq 6, \tag{2.12}$$

$$\mu_1(x_1^2 + x_2^2 - 5) = 0, \mu_2(3x_1 + x_2 - 6) = 0, \tag{2.13}$$

$$4x_1 + 2x_2 - 10 + 2\mu_1 x_1 + 3\mu_2 = 0, \tag{2.14}$$

$$2x_1 + 2x_2 - 10 + 2\mu_1 x_2 + 3\mu_2 = 0. \tag{2.15}$$

Comme il y a deux contraintes en inégalité il y a quatre possibilités :

1. les deux contraintes sont inactives : $x_1^2 + x_2^2 < 5, 3x_1 + x_2 < 6$,
   ce qui donne : $\mu_1 = 0, \mu_2 = 0$

2. g_1 est active et g_2 ne l'est pas, ce qui donne : $x_1^2 + x_2^2 = 5$ et $\mu_2 = 0$.
3. g_1 est inactive et g_2 est active, ce qui donne : $\mu_1 = 0$ et $3x_1 + x_2 = 6$.
4. Les deux contraintes sont actives : $x_1^2 + x_2^2 = 5$ et $3x_1 + x_2 = 6$. Il faut donc tester chacun de ces cas et résoudre les équations de KKT à chaque fois.

1. Premier cas : les deux contraintes sont inactives $\mu_1 = 0$ et $\mu_2 = 0$. Les équations (7) et (8) deviennent :
   $4x_1 + 2x_2 = 10$ et $2x_1 + 2x_2 = 10$,
   ce qui donne $x_1 = 0$ et $x_2 = 5$ Or ce point ne vérifie pas les contraintes (condition de réalisable (5). Par conséquent ce cas est impossible.

2. Deuxième cas : g_1 est active et g_2 ne l'est pas : on obtient alors
   $x_1^2 + x_2^2 = 5, \mu_2 = 0$.

$4x_1 + 2x_2 - 10 + 2\mu_1 x_1 = 0, 2x_1 + 2x_2 - 10 + 2\mu_1 x_2 = 0.$

Ce système a pour solution $x_1 = 1, x_2 = 2$ et $\mu_1 = 1$. Cette solution est admissible : le point trouvé vérifie les conditions de KKT.

3. Troisième cas : g_1 est inactive et g_2 est active ; on obtient

   $4x_1 + 2x_2 - 10 + 3\mu_2 = 0, 2x_1 + 2x_2 - 10 + \mu_2 = 0$ et $x_2 = 6 - 3x_1$.

   C'est un système linéaire de trois équations à trois inconnues dont la solution est $x_1 = \frac{2}{5}, x_2 = \frac{24}{5}$ et $\mu_2 = \frac{-2}{5}$. On voit que μ_2 n'est pas positif, donc ce cas est impossible.

4. Le cas où les deux contraintes sont actives correspond aux points situés à l'intersection de la droite et du cercle. Un rapide calcul montre que ce sont les points : $x_1 = \frac{9}{5} + \frac{\sqrt{14}}{10} \backsimeq 2.17$ ou $x_1 = \frac{9}{5} - \frac{\sqrt{14}}{10} \backsimeq 1.17$

   avec $x_2 = 6 - 3x_1$ et que $\mu_1 = \frac{(4 - 10x_1)}{(20x_1 - 36)}$.

   On voit donc que μ_1 est négatif pour $x_1 = 2.17$. Dans l'autre cas $\mu_1 \backsimeq 1.34$. mais $\mu_2 \backsimeq -0.98$ est négatif. Finalement ce cas est impossible.

Finalement la seule solution possible des équations de KKT est $x^* = (1, 2)$. On peut vérifier a posteriori que ce point est régulier. La seule contrainte active est g_1. Comme $\nabla g_1(x^*) = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$ n'est pas nul, la condition de régularité est donc satisfaite. On pourrait aussi vérifier la condition du second ordre.

<u>Condition nécessaire et suffisante de seconde ordre :</u>

$$d^T \nabla^2_{xx} L(x^*, \mu^*) d \geq 0.$$

$L(x^*, \mu^*) = \begin{pmatrix} 4 + 2\mu_1^* & 2 \\ 2 & 2 + 2\mu_1^* \end{pmatrix} = \begin{pmatrix} 6 & 2 \\ 2 & 4 \end{pmatrix} = 24 > 0$

Donc $(x_1^*, x_2^*) = (1, 2)$ est un point min locale du problème (PCI).

CHAPITRE 3

MÉTHODES PRIMALES

Dans ce chapitre, on va étudier des méthodes appelées méthodes primales.
Ces méthodes focalisent sur un groupe des méthodes appelés méthode de direction réalisable, méthode de gradient réduit et méthode de Frank et Wolfe.
Ces méthodes sont des techniques permettant d'analyser et de résoudre numériquement les problèmes d'optimisation non linéaire avec contraintes.

3.1 Position du problème

Le problème primale de la programmation mathématique est le suivant :

$$Min_{x \in S} f(x) \tag{3.1}$$

tel que $S = \{x \in \mathbb{R}^n, g_i(x) \leq 0, i = 1, \ldots, n\}$
S l'ensemble des contraintes.

3.2 Approche de résolution (méthode itérative)

3.2.1 Itération initiale

Déterminer une solution initiale $x_0 \in S$.

3.2.2 Itération générale $(k+1)$

Déterminer une direction d^k et un pas α_k dans cette direction pour s'éloigner de la solution actuelle x^k en prenant :

$$x^{k+1} = x^k + \alpha_k d^k \in S.$$

3.2.3 Test d'arrêt

jusqu'à ce qu'un critère d'arrêt soit satisfait la direction d^k varie selon la méthode.
Le pas optimale déterminé comme suit :
Étant donné x^k et d^k, dénotons par $\overline{\alpha}_k$ la plus grande valeur que peut prendre le pas α_k de sorte que $x^k + \overline{\alpha}_k d^k \in S$.
Alors le pas optimal α_k^* satisfait la relation :

$$f(x^k + \alpha_k^* d^k) = Min_{0 \leq \alpha_k \leq \overline{\alpha}_k} \left\{ f(x^k + \alpha_k d^k) \right\} \tag{3.2}$$

3.3 Méthode de direction réalisable

Soit le problème avec contraintes d'inégalités linéaires :

$$(P) \begin{cases} min_{x \in \mathbb{R}^n} f(x) \\ A_i x \leq b_i. \end{cases}$$

Supposons que nous dispose d'un point initiale admissible x_0 de (P)
$x_0 \in S = \{x \in \mathbb{R}^n, Ax \leq b\}$. L'ensemble des indices des contraintes actives (saturée) :

$$I_k = \{i/A_i x_0 - b_i = 0\}.$$

3.3.1 Déterminer la direction d^k

Puisque le pas $\alpha_k \geq 0$ alors pour que $x^k + \alpha_k d^k \in S$ il faut que :

$$a_i^T(x^k + \alpha_k d^k) \leq b_i, \forall i \in I_k.$$

$$a_i^T x^k + a_i^T \alpha_k d^k \leq b_i, \forall i \in I_k.$$

On a : $a_i^T x^k = b_i$ Alors :

$$\Rightarrow a_i^T \alpha_k d^k \leq b_i - b_i.$$

i.e :

$$a_i^T \alpha_k d^k \leq 0, \forall i \in I_k.$$

Par conséquent la direction d^k est définie a l'aide d'une solution optimale du problème :

$$(P_k) \begin{cases} min \nabla f(x^k)^T d \\ a_i^T d^k \leq 0, \forall i \in I_k \\ -1 \leq d_j \leq 1, j = 1, \ldots, n. \end{cases}$$

Si la valeur optimale $\nabla f(x^k)^T d < 0$ de (P_k), alors la direction d^k est une direction de f au point x^k et par conséquent la fonction décroit à la direction d^k dans le voisinage de x^k.

3.3.2 Déterminer le pas de déplacement α_k

Étant donné x^k et d^k, dénotons par $\overline{\alpha}_k$ est la plus grande valeur que peut prendre le pas α_k de sorte que $x^k + \overline{\alpha}_k d^k \in S$.
Ainsi, dans le cas de contraintes linéaires, $\overline{\alpha}_k$ est la plus grande valeur que peut prendre

α_k :

$$a_i^T(x^k + \alpha_k d^k) \leq b_i, i = 1, \ldots, m.$$

$$\Rightarrow a_i^T x^k + \alpha_k a_i^T d^k \leq b_i, i = 1, \ldots, m.$$

Or puisque la direction d^k a été choisi de telle sorte

$$a_i^T(x^k + \alpha_k d^k) \leq b_i, i \in I_k.$$

Pour toute valeur de $\alpha_k \geq 0$, alors il suffit de choisir $\overline{\alpha}_k$ comme étant la plus grande valeur que peut prendre le pas α_k de sorte :

$$a_i^T(x^k + \alpha_k d^k) \leq b_i, i \notin I_k.$$

$$\Rightarrow a_i^T x^k + \alpha_k a_i^T d^k \leq b_i, i \notin I_k.$$

puisque $a_i^T x^k < b_i$ lorsque $i \notin I_k$, il s'ensuit que :

- Si : $a_i^T d^k \leq 0$, alors

$$a_i^T x^k + \alpha_k a_i^T d^k < b_i$$

  pour toute valeurs $\alpha_k \geq 0$.

- Si : $a_i^T d^k > 0$, alors

$$a_i^T x^k + \alpha_k a_i^T d^k \leq b_i.$$

$$\Rightarrow \alpha_k a_i^T d^k \leq b_i - a_i^T x^k.$$

$$\Rightarrow \alpha_k \leq \frac{b_i - a_i^T x^k}{a_i^T d^k}.$$

  Par conséquent

$$\overline{\alpha}_k = Min_{i \notin I_k} \left\{ \frac{b_i - a_i^T x^k}{a_i^T d^k} : a_i^T d^k > 0 \right\}$$

  et

$$\alpha_k \in [0, \overline{\alpha}_k].$$

3.3.3 Critère d'arrêt

Le critère d'arrêt est satisfait lorsque $\nabla f(x^k)^T d^k \geq 0$ pour un certain indice k, sinon la méthode s'arrête à la limite quand $k \to \infty$.

Théorème 3.3.1.
Supposons que f est convexe. Si la solution optimale d^k de (p_k) vérifiée $\nabla f(x^k)^T d^k \geq 0$, alors x^k est une solution optimale de (p) [13], [2].

Preuve 3.3.1.
Par contradiction, supposons que x^k n'est pas solution optimale de (p). Alors il existe une solution réalisable $\overline{x}$ de (p) telle que $f(\overline{x}) < f(x^k)$ et puisque f est convexe, l'inégalité du gradient s'applique et

$$f(\overline{x}) \leq f(x^k) + \nabla f(x^k)^T(\overline{x} - x^k).$$

$$\Rightarrow f(\overline{x}) - f(x^k) \geq \nabla f(x^k)^T(\overline{x} - x^k). \tag{3.3}$$

on a

$$f(\overline{x}) < f(x^k) \Rightarrow f(\overline{x}) - f(x^k) < 0.$$

et substituant dans (2.3),il s'ensuit que

$$0 > f(\overline{x}) - f(x^k) \geq \nabla f(x^k)^T(\overline{x} - x^k).$$

D'autre part, si $i \in I_k$, alors $a_i^T x^k = b_i$ et $a_i^T \overline{x} \leq b_i$ et ainsi

$$a_i^T \overline{x} - a_i^T x^k \leq b_i - b_i = 0.$$

$$\Rightarrow a_i^T(\overline{x} - x^k) \leq 0.$$

Donc $\overline{d} = \dfrac{\overline{x} - x^k}{\sum_{j=1}^{n} \mid \overline{x}_j - x_j^k \mid}$ est une solution de (p_k) puisque :

$$a_i^T \left(\frac{\overline{x} - x^k}{\sum\limits_{j=1}^{n} \mid \overline{x}_j - x_j^k \mid} \right) = \frac{1}{\sum\limits_{j=1}^{n} \mid \overline{x}_j - x_j^k \mid} a_i^T(\overline{x} - x^k) \leq 0, i \in I_k.$$

Par conséquent, le fait que $\nabla f(x^k)^T \frac{\overline{x} - x^k}{\sum\limits_{j=1}^{n} \mid \overline{x}_j - x_j^k \mid} < 0$ *contredit l'hypothèse que* d^k *est une solution optimale de* (p_k).■

3.3.4 Exemple

Soit le problème :

$$(p) \begin{cases} Min f(x) = \frac{1}{2}(x_1^2 + x_2^2) \\ -x_1 + x_2 \leq 7 \\ x_1 + x_2 \leq 5 \\ -x_2 \leq -2 \end{cases}$$

minimiser la fonction f par la méthode de direction réalisable avec

$x^{(0)} = \begin{pmatrix} -2 \\ 3 \end{pmatrix}$

Itération 1 :

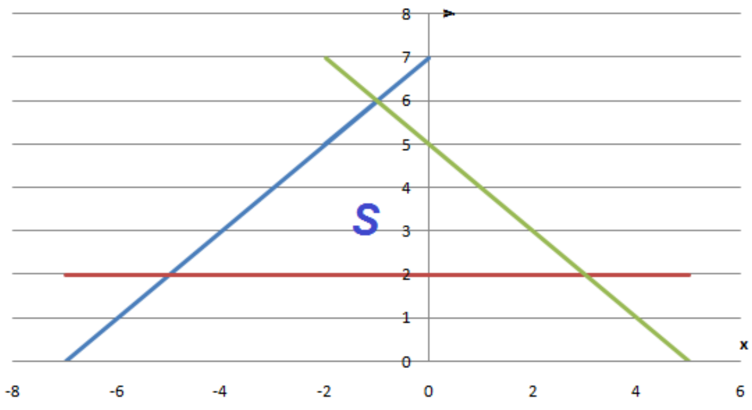

FIGURE 3.1 – S ensemble réalisable

On a $x^{(0)} = \begin{pmatrix} -2 \\ 3 \end{pmatrix}$

calculer $d^{(0)}$:

Trouver la solution optimale du problème suivant :

$$(1)\begin{cases} min \nabla f(x^{(0)})d^{(0)} \\ a_i d^{(0)} \leq 0, i \in I_0. \\ -1 \leq d^{(0)} \leq 1 \end{cases}$$

On a $\nabla f(x) = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

Trouver I_0 **:**

$I_0 = \left\{ i / a_i x^{(0)} - b_i = 0 \right\}$. donc :

$-x_1^{(0)} + x_2^{(0)} - 7 = 2 + 3 - 7 = -2 \neq 0$ n'est pas active en $x^{(0)}$.

$x_1^{(0)} + x_2^{(0)} - 5 = -2 + 3 - 5 = -4 \neq 0$ n'est pas active en $x^{(0)}$.

$-x_2^{(0)} + 2 = 5 + 2 = 7 \neq 0$ n'est pas active en $x^{(0)}$.

donc le problème (1) devient

$$(1)\begin{cases} min(-2d_1^{(0)} + 3d_2^{(0)}) \\ -1 \leq d_{1,2}^{(0)} \leq 1 \end{cases}$$

la solution optimale $d^{(0)} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

<u>Test d'arrêt :</u>

$\nabla f(x^0)d^{(0)} = \begin{pmatrix} -2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -5 < 0$

<u>Itération 2</u>

Calculer α_0 :

On a $\alpha_{0_{max}} = min \left\{ \frac{b_i - a_i x^{(0)}}{a_i d^{(0)}}, a_i d^{(0)} > 0 \right\}$.

Calculer $a_i d^{(0)}$:

On a $a_{ix_1} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, a_{ix_2} = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$ donc

$a_1 d^{(0)} = \begin{pmatrix} -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -2 < 0.$

$a_2 d^{(0)} = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 0.$

$a_3 d^{(0)} = \begin{pmatrix} 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 1 > 0.$

Donc : $\alpha_{0_{max}} = \dfrac{b_3 - a_3 x^{(0)}}{a_3 d^{(0)}} = \dfrac{-2 - \begin{pmatrix} 0 & -1 \end{pmatrix}\begin{pmatrix} -2 \\ 3 \end{pmatrix}}{1} = 1.$

Donc : $\alpha_0 \in [0, \alpha_{0_{max}}]$ alors $\alpha_0 \in [0, 1]$.

Calculer α_0

$\alpha_0 = argmin f(x^{(0)} + \alpha_0 d^{(0)})$.

$$\begin{aligned} f(x^{(0)} + \alpha_0 d^{(0)}) &= f(-2 + \alpha_0, 3 - \alpha_0) \\ &= \frac{1}{2}(-2 + \alpha_0)^2 + \frac{1}{2}(3 - \alpha_0)^2 \\ &= \alpha_0^2 - 5\alpha_0 + \frac{13}{2}. \end{aligned}$$

$$\begin{aligned} \alpha_0 \text{min de} f &\iff f^{'}(-2 + \alpha_0, 3 - \alpha_0) = 0 \\ &\iff 2\alpha_0 - 5 = 0 \\ &\iff \alpha_0 = \frac{5}{2} \notin [0, 1]. \end{aligned}$$

donc : $\alpha_0 = \alpha_{0_{max}} = 1$

Calculer $x^{(1)}$ **:**

$x^{(1)} = x^{(0)} + \alpha_0 d^{(0)} = \begin{pmatrix} -2 \\ 3 \end{pmatrix} + 1 \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}.$

Calculer $d^{(1)}$ **:**

trouver la solution optimale du problème :

$(2) \begin{cases} Min \nabla f(x^{(1)}) d^{(1)} \\ a_i d^{(1)}, i \in I_1 \\ -1 \leq d^{(1)} \leq 1 \end{cases}$

Trouver I_1 **:**

$-x_1^{(1)} + x_2^{(1)} - 7 = -4 \neq 0$

$-1 + 2 - 5 = -4 \neq 0$

$-2 + 2 = 0$ contrainte active

donc le problème (2) devient

$$(2) \Longleftrightarrow \begin{cases} Min(-d_1^{(1)} + 2d_2^{(1)}) \\ \quad -d_2^{(1)} \leq 0 \\ -1 \leq d_{1,2} \leq 1 \end{cases}$$

La solution optimale : $d^{(1)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

<u>Teste d'arrêt</u>

$\nabla f(x^{(1)})d^{(1)} = -1 < 0$

Calculer $\alpha^*_{1_{max}}$

$\alpha^*_{1_{max}} = min\left\{\dfrac{b_i - a_i x^{(1)}}{a_i d^{(1)}}, a_i d^{(1)} > 0\right\}.$

Calculer $a_i d^1$

$a_{1x_1} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, a_{2x_1} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ donc $a_1 d^{(1)} = -1 < 0$ et $a_2 d^{(1)} = 1 > 0$ alors :

$\alpha^*_{1_{max}} = \dfrac{5 - \begin{pmatrix} 1 & 1 \end{pmatrix}\begin{pmatrix} -1 \\ 2 \end{pmatrix}}{1} = 4$

$\alpha_{1_{max}} \in [0, 4]$

$$\begin{aligned} \alpha_1 &= argminf(x + \alpha_1 d^{(1)}) \\ &= argminf(-1 + \alpha_1, 2) \end{aligned}$$

$f(\alpha_1) = \dfrac{1}{2}\alpha_1^2 + \dfrac{5}{2}.$

$f^{'}(\alpha_1) = 0 \Longleftrightarrow \alpha_1 = 1 \in [0, 1].$

Calculer $x^{(2)}$

$$\begin{aligned} x^{(2)} &= x^{(1)} + \alpha_1 d^{(1)} \\ &= \begin{pmatrix} -1 \\ 2 \end{pmatrix} + 1\begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ x^{(2)} &= \begin{pmatrix} 0 \\ 2 \end{pmatrix} \end{aligned}$$

Calculer $d^{(2)}$

Trouver la solution optimale du problème

$$(3)\begin{cases} min\nabla f(x^{(2)})d^{(2)} \\ a_1 d^{(2)} \leq 0, i \in I_2 \\ -1 \leq d_{1,2} \leq 1 \end{cases}$$

Trouver I_3 **:**

$-x_1^{(2)} + x_2^{(2)} - 7 = -5 \neq 0$

$2 - 5 = -3 \neq 0$

$-2 + 2 = 0$ (contrainte active)

donc le problème (3) devient

$$(3) \Longleftrightarrow \begin{cases} min(2d_2^{(2)}) \\ -d_2 \leq 0 \\ -1 \leq d_{1,2} \leq 1 \end{cases}$$

la solution optimale $d^{(2)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

<u>**Test d'arrêt :**</u>

$\nabla f(x^{(2)})d^{(2)} = \begin{pmatrix} 0 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0 \geq 0$

donc $x^{(2)}$ est la solution optimale du problème.

3.4 Méthode de gradient réduit

Il s'agit d'une généralisation de la méthode qui consiste a utiliser les contraintes (linéaires) d'égalités pour éliminer des variables. En utilisant les notions de variable de base et variable hors base d'une façon analogue. Nous ne la présenterons brièvement que dans le cas des contraintes linéaires.

On considère le problème :

$$(P)\begin{cases} minf(x) \\ Ax = b \\ x \geq 0 \end{cases} \tag{3.4}$$

avec

$$x \in \mathbb{R}^n, A \in \mathbb{R}^{m*n}, b \in \mathbb{R}^m, rang(A) = m$$

et f une fonctionnelle deux fois continument différentiable.
Choisissons m colonnes linéairement indépendant de$A$.
Appelons B la matrice forme de ces m colonnes et N la matrice formes des $n-m$ colonnes.
Étant donnée une solution admissibles x, on décompose x en deux sous vecteurs (x_B, x_N) de telle manière que x_B contienne m composante strictement positive et x_N contienne les autres composants (positives ou nulles).
Soit :

- x_B présente les variables de base.
- x_N variables hors-base [2].

donc $x = (x_B, x_N)$ et on posant $A = (B, N)$.
Les contraintes $Ax = b$ peuvent s'écrire :

$$(B, N) \begin{pmatrix} x_B \\ x_N \end{pmatrix} = b$$
$$\Rightarrow Bx_B + Nx_N = b$$

Comme les m première colonnes de A sont linéairement indépendant, la matrice B est inversible donc :

$$B^{-1}Bx_B + B^{-1}Nx_N = B^{-1}b$$
$$x_B = B^{-1}(b - N)x^N \tag{3.5}$$

En adaptant cette convention, considérons le problème d'optimisation avec contrainte d'égalité linéaires :

$$min_{(x_B, x_N)} f \begin{pmatrix} x_B \\ x_N \end{pmatrix}$$

Sous contrainte :

$$Bx_B + Nx_N = b.$$

Il est équivalent au problème sans contraintes [2] :

$$PGR \begin{cases} min_{x_N} f \begin{pmatrix} B^{-1}(b - Nx_N) \\ x_N \end{pmatrix}, x_N \geq 0 \end{cases} \tag{3.6}$$

de façon similaire d partitionné comme suit : $d = \begin{bmatrix} d_B \\ d_N \end{bmatrix}$. Puisque d est une direction admissible donc $Ad = 0$. [1]

$$\begin{aligned} \begin{bmatrix} B & N \end{bmatrix} \begin{bmatrix} d_B \\ d_N \end{bmatrix} &= 0 \\ Bd_B + Nd_N &= 0 \end{aligned} \tag{3.7}$$

est puisque B inversible donc :

$$\begin{aligned} B^{-1}Bd_B + B^{-1}Nd_N &= 0 \\ d_B &= -B^{-1}Nd_N \end{aligned} \tag{3.8}$$

* Introduisons maintenant le gradient (dit "réduit") associé au "PGR".
* Il a pour expression :

$$\begin{aligned} \nabla f(x)d &= \nabla f(x_B, x_N) \begin{bmatrix} d_B \\ d_N \end{bmatrix}, \text{avec}, d_B = -B^{-1}Nd_N \\ &= \nabla_B f(x)d_B + \nabla_N f(x)d_N \\ &= -\nabla_B f(x)B^{-1}Nd_N + \nabla_N f(x)d_N \\ &= [-\nabla_B f(x)B^{-1}N + \nabla_N f(x)]d_N \end{aligned}$$

On suppose $R_N = \nabla_N f(x) - \nabla_B f(x)B^{-1}N$ donc $\nabla f(x)d = R_N d_N$ [1]. Et par définition R_N est appelé "**Gradient réduit**" relativement a la base B.
A l'étape courante de l'algorithme, soit $x = \begin{bmatrix} x_B & x_N \end{bmatrix}$ la solution réalisable obtenue.

Pour améliorer cette solution , en se déplace dans la direction $d = \begin{bmatrix} d_B & d_N \end{bmatrix}$, où d_N est définie par :

$$d_{Nj} = \begin{cases} -R_j & \text{Si} \quad R_j < 0 \\ 0 & \text{Si} \quad R_j \geq 0, x_j = 0 \end{cases}$$

- Si $d_N = 0$ ce qui est le critère d'arrêt (critère d'optimisation).
- Si $d_N \neq 0$ on peut améliorer x, en effectuant une recherche linéaire convenable suivant d.

Recherche linéaire

- Si $d_N \neq 0$ alors on peut améliorer x en remplaçant par : $x^* + \alpha^* d$.

  où :

$$\alpha^* = min \nabla f(x + \alpha^* d), \qquad 0 \leq \alpha^* \leq \alpha_{max}$$

$$\alpha_{max} = min_{d_j} \left\{ \frac{-x_j}{d_j}, \qquad d_j < 0, j = 1, \cdots, n \right\}$$

  d'après la détermination de α^* deux situation peuvent alors se présenter :

1• $\alpha^* < \alpha_{max}$:

  Dans ce cas, aucune variable ne s'annule. La procédure ce poursuit à partir du nouveau point $x^* = x + \alpha^* d$.

  On recalculera R_N en utilisant la même base B.

2• $\alpha^* = \alpha_{max}$:

  Dans ce cas, une variable s'annule, soit x_s, on distingue aussi deux cas :

  - Si x_s est une variable hors base alors on poursuit la procédure sans changer de base.
  - Si x_s est une variable de base on effectue un changement de base : n'importe quelle variable hors base $x_t > 0$ peut venir remplacer x_s.

La procédure ce poursuit alors avec cette nouvelle base.

Pour terminer. vérifions que la condition $d_N = 0$, implique les condition de "Kuhn-Tuker" (puisque les contraintes sont linéaires, l'hypothèse de qualification est toujours vérifiée). En associant aux contraintes $Ax = b$, des multiplicateurs $\mu \in \mathbb{R}^n$ (de signe quelconque) et aux contraintes $x \geq 0$, des multiplicateurs $\lambda \geq 0$ ($\lambda \in \mathbb{R}^n_+$), les conditions de "Kuhn-Tucker". s'écrivent :

$$\begin{cases} -\nabla_x f + \mu A + \lambda I &= 0 \\ \lambda x &= 0 \end{cases} \tag{3.9}$$

Où I est la matrice unité $(n * n)$.
En posant :

$$\begin{array}{llll} \mu = \nabla f B^{-1}, & \lambda = [\lambda_B, \lambda_N] & \text{avec} & \\ \lambda_B = 0 & \text{et} & \lambda_N & = \quad (\nabla_N f(x) - \nabla_B f(x) B^{-1} N) = R_N \end{array}$$

La relation (6) est vérifiée.
Remarquons alors que $d_N = 0$, implique $R_N \geq 0$ d'où $\lambda_N \geq 0$ et par suite $\lambda \geq 0$.
D'autre part, comme pour tout j hors base : $d_j = 0$ on a nécessairement : $\mu_j > 0 \Rightarrow x_j = 0$, et on deduit que les conditions de complémentarité (7) sont satisfaites.
Les conditions de '"**Kuhn-Tucker**" sont donc bien vérifiées lorsque $d_N = 0$.
'La méthode du gradient réduit, sous sa forme primitive exposée ci-dessus, ne converge pas globalement.
Il est facile de voir, en effet, que la direction de déplacement d peut subir des discontinuités puisque par exemple lorsqu'une contrainte cesse active, néanmoins, il est possible d'obtenir la convergence globale en modifient la méthode et en définissant la direction de déplacement d_N par :

$$d_{Nj} = \begin{cases} -R_j & \text{Si} \quad R_j < 0 \\ -x_j R_j & \text{Si} \quad R_j > 0 \end{cases}$$

Pour tout les indices j correspondant à des variables hors base [3].

3.4.1 Exemple

$$PGR\begin{cases} minf(x_1,x_2) = x_1^2 + 3x_1x_2 + 4x_2^2 \\ x_1 + x_2 = 1 \\ x_1, x_2 \geq 0 \end{cases} \tag{3.10}$$

minimiser le problème avec la méthode de gradient réduit avec $x^0 = (0,1)^t$.

Itération 1 : $x^0 = (0,1)^t$

On pose x_2 la variable de base donc d'après (9) $x_2 = 1 - x_1$ donc

$$\begin{aligned} f(x_1, 1-x_1) = f(x_1) &= x_1^2 + 3x_1(1-x_1) + 4(1-x_1)^2 \\ &= x_1^2 + 3x_1 - 3x_1^2 + 4(1 + x_1^2 - 2x_1) \\ &= 2x_1^2 - 5x_1 + 4 \end{aligned}$$

Par conséquent :

$$\begin{aligned} R_N &= \nabla_N f(x) - \nabla_B f(x) B^{-1} \\ \nabla_N f(x) &= 2x_1 + 3x_2; \text{tel que } : B = 1, N = 1 \\ \nabla_B f(x) &= 3x_1 + 8x_2 \end{aligned}$$

Donc :

$$\begin{aligned} R_N &= 2x_1 + 3x_2 - (3x_1 + 8x_2)(1)(1) \\ R_N &= 2x_1 + 3x_2 - 3x_1 - 8x_2 \\ R_N &= -5 < 0 \end{aligned}$$

La direction de descente :

$$\begin{aligned} d_N^{(0)} &= -R_N &= 5 \neq 0 \\ d_B^{(0)} &= -B^{-1}Nd_N^{(0)} &= -5 \end{aligned} \quad \text{donc : } d^{(0)} = \begin{pmatrix} 5 \\ -5 \end{pmatrix}$$

Calculons α_{max} :

$$\alpha_{max} = min\left\{\frac{-x_j}{d_j} \qquad , d_j < 0, j = 1, \cdots, n\right\}$$

Alors :

$$\alpha_{max} = min\left\{\frac{-x_B^{(0)}}{d_B} \qquad , d_B < 0, j = 1, \cdots, n\right\}$$

$\alpha_{max} = \frac{1}{5}$, donc : $\alpha \in [0, \alpha_{max}] = \left[0, \frac{1}{5}\right]$.

Calculons α^* :

$$\alpha^* = min_{\alpha \in [0,\alpha_{max}]} \nabla f(x + \alpha^* d) \qquad , 0 \leq \alpha^* \leq \alpha_{max}$$

$$f(x^{(0)} + \alpha d^{(0)}) = = f\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \alpha \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) = f\begin{pmatrix} 5\alpha \\ 1 - 5\alpha \end{pmatrix}$$

$$\begin{aligned} f\begin{pmatrix} 5\alpha \\ 1 - 5\alpha \end{pmatrix} &= (5\alpha)^2 + 3(5\alpha)(1 - 5\alpha) + 4(1 - 5\alpha)^2 \\ &= 25\alpha^2 + 15\alpha(1 - 5\alpha) + 4(1 + 25\alpha^2 - 10\alpha) \\ &= 25\alpha^2 + 15\alpha - 75\alpha^2 + 4 + 100\alpha^2 - 40\alpha \\ &= 50\alpha^2 - 25\alpha + 4 \end{aligned}$$

$f'(x^{(0)} + \alpha d^{(0)}) = 100\alpha - 25$

$f'(x^{(0)} + \alpha d^{(0)}) = 0 \Rightarrow 100\alpha - 25 = 0 \Rightarrow \alpha^* = \frac{25}{100} = \frac{1}{4} \notin [0, \frac{1}{5}]$

donc :

$$\alpha^* = \alpha_{max}$$

Iteration 2 :

$x_N^{(1)} = x_N^{(0)} + \alpha d_N^{(0)} = 0 + \frac{1}{5} * 5 = 1$

$x_B^{(1)} = x_B^{(0)} + \alpha d_B^{(0)} = 1 - \frac{1}{5} * 5 = 0$

Donc : $x^{(1)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

De plus puisque $\alpha^* = \alpha_{max}$ il faut modifier la base en remplaçant x_2 dans la base par x_1.

On pose : x_1 la variable de base.

x_2 la variable hors base.

Donc : $x_1 = 1 - x_2$

$$\begin{aligned} f(x_2) &= (1 - x_2)^2 + 3(1 - x_2)x_2 + 4x_2^2 \\ &= 2x_2^2 + x_2 + 1 \end{aligned}$$

$$\begin{aligned} R_N &= \nabla_N f(x)] - \nabla_B f(x) B^{-1} \\ \nabla_N f(x) &= 8x_2 + 3x_1; tq : B = 1, N = 1 \\ \nabla_B f(x) &= 2x_1 + 3x_2 \end{aligned}$$

$R_N = 8x_1 + 3x_1 - 2x_1 - 3x_2 = x_1 + 5x_2$.
$R_N(x^{(1)}) = R_N \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.
Donc : $x^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ est la solution optimale du problème.

3.5 Méthode de Frank-Wolfe

La méthode de "**Frank-Wolfe**" résout de façon itérative le problème :

$$(P) \begin{cases} min f(x) \\ x \in S \end{cases} \tag{3.11}$$

où S est un ensemble convexe de $\mathbb{R}^n$.
f est convexe.

3.5.1 Construction de la méthode

Soit $x^0 \in S$ un point arbitraire, pour construire $x^{k+1} \in S$ on a :

$$x^{k+1} = x^k + \alpha_k^* d^k$$

3.5.2 Direction de descente

Le choix de direction de descente consiste a linéariser en x^k la fonction f.
D'après la développement de Taylor d'ordre 1 en x^k on trouve

$$f(x, y) = f(x^k) + \nabla f(x^k)(y - x^k)$$

On pose $d^k = y^k - x^k$ qui satisfait la condition de descente $\nabla f(x^k)d^k < 0$:

$$\begin{aligned} \nabla f(x^k)d^k &= \nabla f(x^k)(y^k - x^k) \\ &= \nabla f(x^k)y^k - \nabla f(x^k)x^k < 0 \\ &\Leftrightarrow \nabla f(x^k)y^k < \nabla f(x^k)x^k \end{aligned}$$

y^k la solution optimale du problème :

$$(PFW)\begin{cases} min\nabla f(x^k)y \\ y \in S \end{cases} \tag{3.12}$$

3.5.3 Déterminer la valeur de α_k^*

Étant donnée une solution réalisable x^k, prochaine solution réalisable x^{k+1} est déterminer comme suit :

$$\begin{aligned} x^{k+1} &= x^k + \alpha_k^* d^k \\ &= x^k + \alpha_k^*(y^k - x^k) \\ &= (x^k - \alpha_k^* x^k) + \alpha_k^* y^k \\ x^{k+1} &= (1 - \alpha_k^*)x^k + \alpha_k^* y^k \end{aligned}$$

Le pas α_k^* est la solution optimale du problème :

$$\begin{aligned} f(x^{k+1}) &= f(x^k + \alpha_k^* d^k) \\ &= f(x^k + \alpha_k^*(y^k - x^k)) \\ &= f(x^k - \alpha_k^* x^k) + \alpha_k^* y^k) \\ x^{k+1} &= f((1 - \alpha_k^*)x^k + \alpha_k^* y^k) \\ &= min_{\alpha_k^* \in [0,1]} f((1 - \alpha_k^*)x^k + \alpha_k^* y^k) \end{aligned}$$

3.5.4 Le critère d'arrêt

$\nabla f(x_k)(y^k - x^k) \geq 0$ [11], [17].

3.5.5 Exemple

$min f(x) = 5x_1 - x_1^2 - 8x_2 + 2x_2^2$

$$\begin{cases} 3x_1 + 2x_2 \leq 6 \\ -x_1 \leq 0 \\ -x_2 \leq 0 \end{cases}$$

Le point initiale est : $x^0 = (0, 0)$

minimiser le problème par la méthode de Frank-Wolfe.

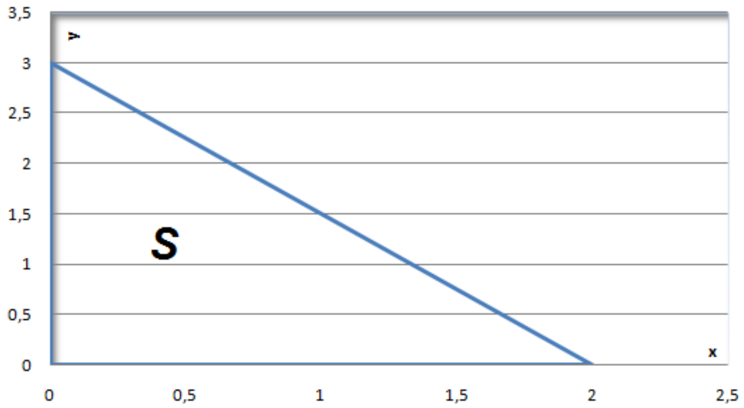

FIGURE 3.2 – Ensemble réalisable

<u>Itération 1 :</u>

1- Comme seul l'objectif et non linéaire on va remplacer $f(x)$ par son approximation de Taylor limitée a d'ordre 1 :
$f(x) = f(0,0) + \dfrac{\partial f}{\partial x_1}(0,0)(y_1 - 0) + \dfrac{\partial f}{\partial x_2}(0,0)(y_2 - 0)$
Évaluons le gradient en $x^0 = (0, 0)$.
$\dfrac{\partial f}{\partial x_1}(0,0) = 5 - 2(0) = 5; \dfrac{\partial f}{\partial x_2}(0,0) = -8 + 4(0) = -8$
D'où l'approximation linéaire de la fonction en $(0, 0)$:

$$g(y) = 5y_1 - 8y_2$$

2- on peut donc résoudre le problème linéaire :

$$ming(y) = 5y_1 - 8y_2$$

$$\begin{cases} 3y_1 + 2y_2 \leq 6 \\ -y_1 \leq 0 \\ -y_2 \leq 0 \end{cases}$$

donc la solution optimale du problème est : $y^{(0)} = (0, 3)$

$d^{(0)} = y^{(0)} - x^{(0)} = \begin{pmatrix} 0 \\ 3 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \end{pmatrix}$

Test d'arrêt

$\nabla g(x^{(0)})d^{(0)} = [5, -8] \begin{pmatrix} 0 \\ 3 \end{pmatrix} = -24 < 0$

- **Calculon α_0^*** :

$f((1-\alpha_0^*)x^{(0)} + \alpha_0^* y^{(0)}) = f \begin{pmatrix} 0 \\ 3\alpha_0^* \end{pmatrix}, f \begin{pmatrix} 0 \\ 3\alpha_0^* \end{pmatrix} = -24\alpha_0^* + 18\alpha_0^{*2}.$

$\nabla f \begin{pmatrix} 0 \\ 3\alpha_0^* \end{pmatrix} = -24 + 36\alpha_0^* = 0 \Rightarrow \alpha_0^* = \frac{2}{3} \in [0, 1].$

$x^{(1)} = x^{(0)} + \alpha_0^* d^{(0)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \frac{2}{3} \begin{pmatrix} 0 \\ 3 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$

Itération 2 :

1- Évaluation du gradient en $x^{(1)} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$:

$\frac{\partial f}{\partial x_1}(0, 0) = 5 - 2(0) = 5; \frac{\partial f}{\partial x_2}(0, 0) = -8 + 4(2) = 0$

2- Résoudre le problème linéaire

$$minz(y) = 5y_1$$

$$\begin{cases} 3y_1 + 2y_2 \leq 6 \\ -y_1 \leq 0 \\ -y_2 \leq 0 \end{cases}$$

donc la solution optimale du problème est : $y^{(1)} = (2,0)$

$d^{(1)} = y^{(1)} - x^{(1)} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \end{pmatrix}$

Test d'arrêt

$\nabla z(x^{(1)})d^{(1)} = (5,0)\begin{pmatrix} 2 \\ -2 \end{pmatrix} = 10 > 0$

Donc la solution optimale est : $x^{(1)} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$

Conclusion

En conclusion, on peut dire que les méthodes primales sont des méthodes itératives pour résoudre des problèmes optimisations non linéaires avec contraintes.

L'objectif des méthodes de direction réalisable et méthode de Frank-Wolfe est de reformuler le problème non linéaire à un problème linéaire, et la méthode de gradient réduit a un principe de transférer le problème avec contrainte à un problème son contrainte.

CONCLUSION GÉNÉRALE

Dans ce travail, nous avons étudié une classe d'optimisation non linéaire avec contraintes. Pour une bonne présentation de ce mémoire, il est indispensable dans le premier chapitre le bagage mathématique nécessaire pour le traitement du problème considérés. Ensuite, dans le deuxième chapitre, nous avons traité les conditions d'optimalité (nécessaires et suffisantes) dans les cas des contraintes de type égalité et inégalité (non linéaire) et aussi dans le cas des contraintes mixtes. Enfin, nous avons étudiés dans le dernier chapitre les méthodes numériques pour résoudre des problèmes d'optimisation non linéaire avec contraintes qui sont la méthode de direction réalisable, la méthode de gradient réduit et la méthode de Frank-Wolfe avec des exemples illustratifs pour chaque méthode.

BIBLIOGRAPHIE

[1] M.Bierlaire,2006, Introduction à l'optimisation différentiable, $1^{ème}$ édition, 2-88074-669-8, PP[23-33-34-35-44-69-494-497].

[2] J.C.Culioli, 2012, Introduction à l'optimisation, ellipses, $2^{ème}$ édition, 978-2-7298-76241, PP[142].

[3] M.Bouzerara, N.Bouchama, M.Boukaroura, 2011, La méthode du gradient réduit pour un problème d'optimisation avec contrainte, mémoire de fin d'étude Présenté pour l'obtention du diplôme de Licence Académique mathématique fondamentales, CENTRE UNIVERSITRAIRE DE MILA, PP[17-19].

[4] S.Le Digabel, 2019, Optimisation non linéaire : Théorie, Polytechnique Montréal, pp[25]

[5] A.BELLOUFI,H.Mebrouk, 2008, Optimisation du processus d'usinage à l'aide de la programmation non linéaire (P.N.L), Mémoire de Magistere en Construction Mécanique,Université Mohammed Khider BISKRA, PP[7].

[6] S.Mottelet, 2003, Optimisation non linéaire, Université de Technologie de Comiégne, PP[132-133-138-139].

[7] S.KETTAB, M.ACHACHE, 2015, Généralisation d'une méthode de trajectoire centrale de points intéreurs pour la programmation semi-définie, thése de doctorat en sciences, Université Ferhat Abbas Setif 1, PP[18-19-20].

[8] F.Delbos, 2004, Problèmes d'Optimisation Non Linéaire avec contraintes en Tomographie de Réflexion 3D, thèse de Doctorat en sciences Mathématiques de paris centre, L'Université Pierre et Marie Curie (Paris VI), PP[91].

[9] I.CIUPERCA, COURS OPTIMISATION, Cours à l'ISFA,en M1SAF,PP[16-17].

[10] A.Crosnier, 2017, M2 optimisation(concepts de base et méthode), PP[33]

[11] G.Gauthier, 7 Octobre 2016, Extensions de l'algorithme de Frank-Wolfe pour la recherche de points selles, ÉCOLE NORMALE SUPÉRIEURE, DÉPARTEMENT DE MATHÉMATIQUES ET APPLICATIONS, PP[6-7]

[12] A.Rondepierre, 4ème année, 2017-2018, Méthodes numériques pour l'optimisation non linéaire déterministe,Département Génie Mathématique et Modélisation, PP[30-41].

[13] I.Charon, O.Hudry, Optimisation non linéaire, École nationale supérieure des télécommunications, PP[15]

[14] E.RADJEF,S.DOUAR, 2019, Optimisation , PP[62-63].

[15] O.Boudjaatat, N.Nesrouche, R.Bouderbane, S.Dekkiche, M.Azi, 2014,OPTIMISATION NON LINÉAIRE AVEC, CONTRAINTE, Mémoire préparé En vue de l'obtention du diplôme de licence Mathématiques fondamentales,Centre Universitaire de Mila,PP[5-6].

[16] R.AREZKI, M.CHEBAH, 08 .07.2013 La programmation mathématique Avec les méthodes des points intérieurs, MÉMOIRE DE FIN D'ÉTUDES EN VUE DE L'OBTENTION DU DIPLÔME DE MASTER EN MATHÉMATIQUES, RECHERCHE OPERATIONNELLE, UNIVERSITÉ MOULOUD MAMMERI DE TIZI-OUZOU, PP[3-4-5-14]

[17] Dankerque, Recherche opérationnelle Daniel DE WOLF, pp[149]