Optimisation Non linèaire

Par

Professeur Abdellatif El Afia

Chapitre 6

Optimisation avec contraintes
Algorithmes primales

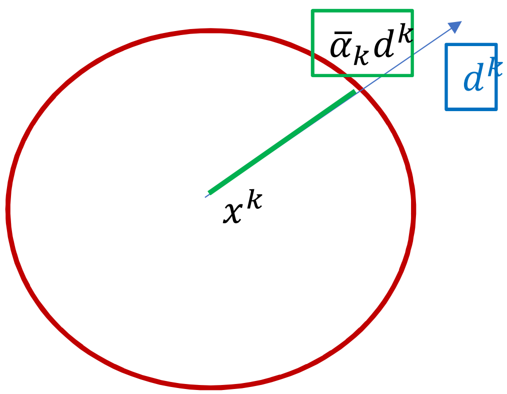
1. **Approche générale**
2. **Méthode direction réalisable**
3. **Frank and Wolfe dans le cas non linéaire convexe**
4. **Gradient réduit le cas ou il y a de la dégénérescence**
5. **Gradient projeté dans le cas non linéaire (variances)**

Algorithmes primales: Approche générale

Considérons le problème général de programmation mathématique suivant:

$$(P)\ \begin{cases} \text{Min} & f(x) \\ \text{sujet à} & f_i(x) \leq 0 \ \ i = 1, \dots, m \end{cases}$$

Soit le domaine réalisable $D_R = \{x \in \mathbb{R}^n : f_i(x) \leq 0, i = 1, \dots, m\}$



Approche de résolution: méthode itérative.

Itération initiale

Déterminer une solution initiale $x^0 \in D_R$.

Itération générale $(\boldsymbol{k+1})$: Etant donné x^k

1. Déterminer une direction d^k
2. Déterminer $\bar{\alpha}_k$ la plus grande valeur que peut prendre le pas α ≥ 0 de sorte que

$$x^{k+1} = x^k + \bar{\alpha}_k d^k \in D_R$$

1. Déterminer le pas optimal $\alpha_k$satisfait la relation :

$$\varphi(\alpha_k) = \min_{0 \leq \alpha \leq \bar{\alpha}_k} \varphi(\alpha) = \min_{0 \leq \alpha \leq \bar{\alpha}_k} \{f(x^k + \alpha_k d^k)\}$$

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

Répéter l'itération générale jusqu'à ce qu'un **critère d'arrêt** soit satisfait.

Algorithmes primales
Méthode des direction réalisables

1. **Calcule de la direction $\boldsymbol{d^k}$**
2. **Calcule de plus grand pas $\boldsymbol{\bar{\alpha}_k}$**
3. **Critère d'arrêt**
4. **Algorithme**
5. **Exemple**

Méthode des direction réalisables

Soit le problème (P)de la forme suivante :

$$(P)\ \begin{cases} \text{Min} & f(x) \\ \text{sujet à} & {a^i}^T x \leq b_i \quad i = 1, \dots, m \end{cases}$$

où $f \in C^1/D_R$ et où $D_R = \left\{x \in \mathbb{R}^n \middle| {a^i}^T x \leq b_i \quad i = 1, \dots, m\right\}$est défini à l'aide d'un ensemble de contraintes linéaires

Calcule de la direction $\mathbf{d}^\mathbf{k}$ ${a^i}^T \mathbf{d}^\mathbf{k} \leq \mathbf{0}\ i \in I_k$

Etant donné une solution réalisable x^k de (P) définissons l'ensemble des indices des contraintes actives par $\boldsymbol{I_k = \left\{i : {a^i}^T x^k = b_i\right\}}$

Puisque le pas $\alpha_k \geq 0$, alors pour que $x^k + \alpha_k d^k \in D_R$, il faut que : $\forall i \in I_k$

$${a^i}^T x^k + \alpha_k {a^i}^T d^k \leq b_i \implies \alpha_k {a^i}^T d^k \leq 0$$

Par conséquent la direction d^k est une solution optimale du problème :

- $$(P_k)\ \begin{cases} \text{Min} & \nabla f(x^k)^T d \\ \text{sujet à} & {a^i}^T d \leq 0 \qquad i \in I_k \\ & -1 \leq d_j \leq 1 \quad j = 1, \dots, n \end{cases}$$

Les contraintes additionnelles $-1 \leq d_j \leq 1$, permettent de comparer les quantités $\nabla f(x^k)^T d$ sur une base équitable pour toutes les directions.

Méthode des direction réalisables

Si la valeur optimale de (P_k) $\nabla f\left(x^k\right)^T d < 0$, alors d^k est une direction de descente de f au point x^k

Calcule de plus grand pas $\bar{\alpha}_k$

Etant donné x^k et d^k, dénotons par $\bar{\alpha}_k$ est la plus grande valeur que peut prendre le pas $\alpha_k \geq 0$ de sorte que $x^k + \bar{\alpha}_k d^k \in D_R$. Donc

$$\forall i = 1, \dots, m, \qquad \boldsymbol{a^{i^T}\left(x^k + \alpha_k d^k\right) \leq b_i \Longrightarrow a^{i^T} x^k + \alpha_k a^{i^T} d^k \leq b_i}$$

Or puisque d^k a été choisi de telle sorte $\boldsymbol{\forall i \in I_k}$: $\boldsymbol{a^{i^T} d^k \leq 0}$

$$\boldsymbol{\forall i \in I_k \left(a^{i^T} x^k = b_i\right)} : a^{i^T}\left(x^k + \alpha_k d^k\right) \leq b_i \Longrightarrow a^{i^T} x^k + \alpha_k a^{i^T} d^k \leq b_i$$

Pour $\boldsymbol{\forall i \notin I_k \left(a^{i^T} x^k < b_i\right)}$ alors il suffit de choisir $\bar{\alpha}_k$:

$$a^{i^T}\left(x^k + \alpha_k d^k\right) \leq b_i \Longrightarrow a^{i^T} x^k + \alpha_k \boldsymbol{a^{i^T} d^k} \leq b_i$$

il s'ensuit que :

- **Si $\boldsymbol{a^{i^T} d^k \leq 0}$, alors $\boldsymbol{a^{i^T} x^k + \alpha_k a^{i^T} d^k < b_i}$ pour toute valeur $\boldsymbol{\alpha_k \geq 0}$,**
- **Si $\boldsymbol{a^{i^T} d^k > 0}$** , alors $a^{i^T} x^k + \alpha_k a^{i^T} d^k \leq b_i \Leftrightarrow \alpha_k a^{i^T} d^k \leq b_i - a^{i^T} x^k \Leftrightarrow \alpha_k \leq \frac{b_i - a^{i^T} x^k}{a^{i^T} d^k}$

Par conséquent : $\bar{\alpha}_k = \min\limits_{i \notin I_k} \left\{ \frac{b_i - a^{i^T} x^k}{a^{i^T} d^k} : a^{i^T} d^k > 0 \right\}$

Méthode des direction réalisables

2.3 Critère d'arrêt

Théorème 1 : Supposons que f est convexe. Si la solution optimale d^k de (P_k) est telle que $\nabla f(x^k)^T d^k \geq 0$, alors x^k est une solution optimale de (P).

Preuve : Par contradiction, supposons que x^k n'est pas une solution optimale de (P).

Alors il existe une solution réalisable $\bar{x}$ de (P) telle que $f(\bar{x}) < f(x^k) \Leftrightarrow f(\bar{x}) - f(x^k) < 0$

Puisque f est convexe, l'inégalité de gradient s'applique est :

$$f(\bar{x}) \geq f(x^k) + \nabla f(x^k)(\bar{x} - x^k) \Longrightarrow 0 > f(\bar{x}) - f(x^k) \geq \nabla f(x^k)(\bar{x} - x^k)$$

D'autre part, **si** $\in I_k$, alors ${a^i}^T x^k = b_i$ et ${a^i}^T \bar{x} \leq b_i$ et ainsi : ${a^i}^T(\bar{x} - x^k) \leq b_i - b_i = 0$

Puisque $i \in I_k$, ${a^i}^T \left(\frac{\bar{x}-x^k}{\sum_{j=1}^n |\overline{x_j} - {x_j}^k|}\right) = \frac{1}{\sum_{j=1}^n |\overline{x_j} - {x_j}^k|} {a^i}^T(\bar{x} - x^k) \leq 0$

Alors $\bar{d} = \frac{\bar{x}-x^k}{\sum_{j=1}^n |\overline{x_j} - {x_j}^k|}$ est une solution réalisable de (P_k), Par conséquent, le fait que

$\nabla f(x^k)^T \frac{\bar{x}-x^k}{\sum_{j=1}^n |\overline{x_j} - {x_j}^k|} < 0$. Contredit l'hypothèse que d^k est une solution optimale de (P_k).

Méthode des direction réalisables: Algorithme

Initialiser une solution initiale $x^0 \in D_R$.

- $k = 0$

DO

- Déterminer la direction d^k est une solution optimale du problème(simplexe) :

$$(P_k) \begin{cases} \boldsymbol{Min} \quad \nabla f(x^k)^T d \\ \textbf{sujet à} \quad a^{i^T} d \leq 0 \qquad i \in I_k \\ \qquad -1 \leq d_j \leq 1 \quad j = 1, \dots, n \end{cases}$$

- Déterminer $\bar{\alpha}_k$ la plus grande valeur que peut prendre le pas $\alpha_k \geq 0$:

$$\bar{\alpha}_k = \underset{i \notin I_k}{Min} \left\{ \frac{b_i - a^{i^T} x^k}{a^{i^T} d^k} : a^{i^T} d^k > 0 \right\}$$

- *Trouver* α_k^* *la solution optimale du problème :*

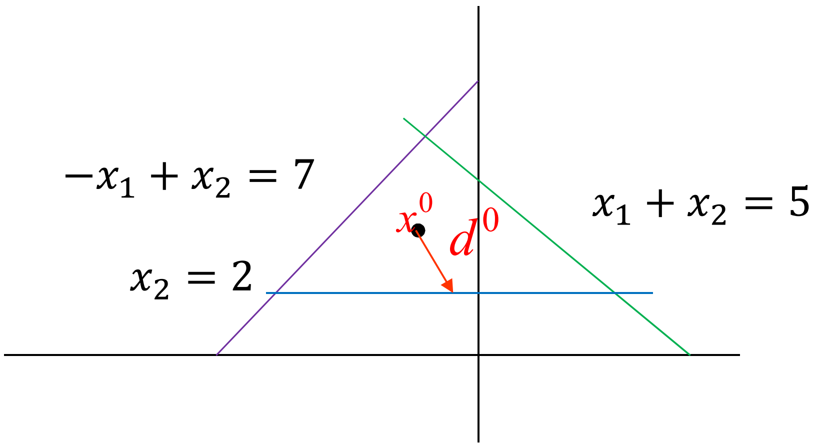
$$\alpha_k^* = \underset{0 \leq \alpha_k \leq \bar{\alpha}_k}{argMin}\, \varphi(\alpha_k) = \underset{0 \leq \alpha_k \leq \bar{\alpha}_k}{argMin}\, f(x^k + \alpha_k d^k)$$

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

WHILE $\nabla f(x^k)^T d \leq 0$

- $x^* = x^k$
- RETURN x^*

Méthode des direction réalisables: **Exemple**



$$(P)\begin{cases} \text{Min} \quad f(x) = \dfrac{1}{2}\,(x_1^2 + x_2^2) \\ \text{sujet à} \quad -x_1 + x_2 \leq 7 \\ \qquad\qquad x_1 + x_2 \leq 5 \\ \qquad\qquad -x_2 \leq -2 \end{cases} \Longrightarrow \begin{cases} {a^1}^T = (-1,1) \\ {a^2}^T = (1,1) \\ {a^3}^T = (0,-1) \end{cases}$$

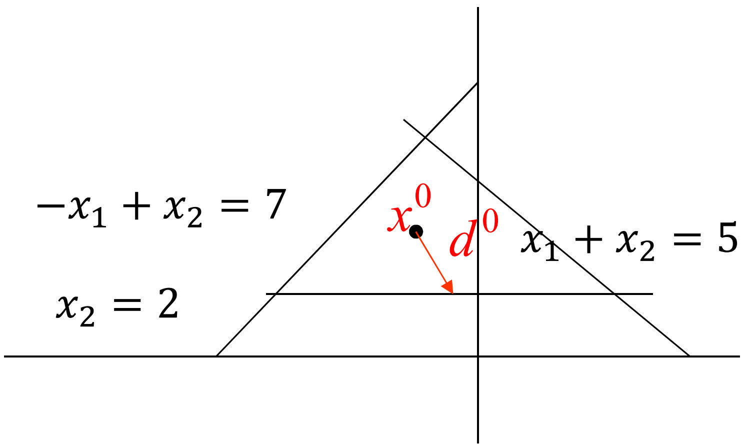
$$\nabla f(x) = [x_1, x_2]^T$$

Itération 1 : $x^0 = [-2,3]^T$. Alors $I_0 = \emptyset$ et $\nabla f(x^0) = [-2,3]^T$
La direction $d^0 = [1,-1]^T$ est solution du problème :

$$(P_0)\begin{cases} \text{Min } \nabla f(x^0)^T d = -2d_1 + 3d_2 \\ \qquad\qquad \text{Sujet à} \qquad\qquad -1 \leq d_j \leq 1 \quad j = 1,2 \end{cases}$$

$$x^1 = x^0 + \alpha_0 d^0 = [-2,3]^T + \alpha_0[1,-1]^T = [-2+\alpha_0, 3-\alpha_0]^T.$$

Méthode des direction réalisables: Exemple



$$(P)\begin{cases} \text{Min} \quad f(x) = \frac{1}{2}\,(x_1^2 + x_2^2) \\ \text{sujet à} \quad -x_1 + x_2 \le \ \ 7 \\ \qquad\qquad\ \ x_1 + x_2 \le \ \ 5 \\ \qquad\qquad\quad\ \ -x_2 \le -2 \end{cases} \Longrightarrow \begin{cases} {a^1}^T = (-1,1) \\ {a^2}^T = (1,1) \\ {a^3}^T x^0 = (0,-1) \end{cases}$$

$$\nabla f(x) = [x_1, x_2]^T$$
$${a^1}^T x^0 = 5 < 7$$
$${a^2}^T x^0 = 1 < 5$$
$${a^3}^T x^0 = -3 < -2$$

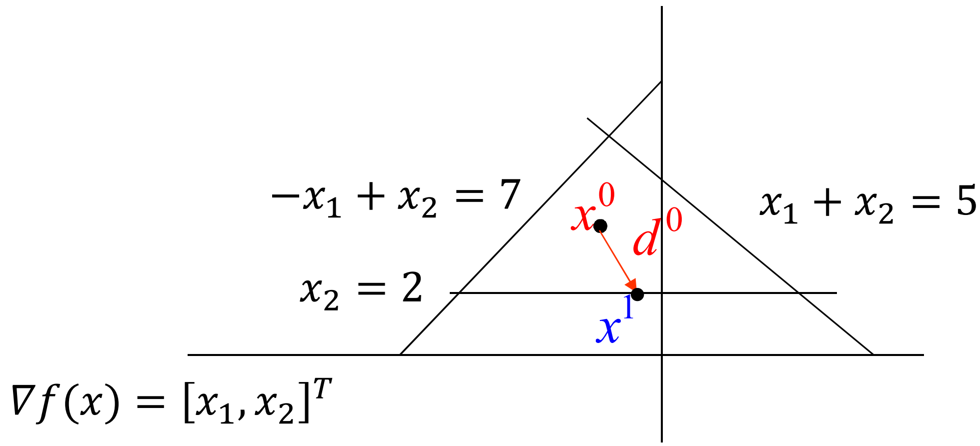
Itération 1 : $x^0 = [-2{,}3]^T.\ d^0 = [1,-1]^T,\ x^1 = [-2+\alpha_0, 3-\alpha_0]^T$.

La plus grande valeur $\bar{\alpha}_0 = 1{:}\ x^1 = x^0 + \alpha_0 d^0 \in D_R$

$$\left.\begin{aligned} {a^1}^T d^0 &= -1\times 1 - 1\times 1 = -2 < 0 \\ {a^2}^T d^0 &= 1\times 1 - 1\times 1 = 0 \le 0 \\ {a^3}^T d^0 &= -1\times -1 = 1 > 0 \end{aligned}\right\} \Longrightarrow \ \bar{\alpha}_0 = \frac{b_3 - \ {a^3}^T x^0}{{a^3}^T d^0} = \frac{-2+3}{1} = 1$$

Méthode des direction réalisables: Exemple

$$(P)\begin{cases} \text{Min} \quad f(x) = \frac{1}{2}\,(x_1^2 + x_2^2) \\ \text{sujet à} \quad -x_1 + x_2 \leq \;\; 7 \\ \qquad\qquad\quad x_1 + x_2 \leq \;\; 5 \\ \qquad\qquad\qquad\; -x_2 \leq -2 \end{cases}$$



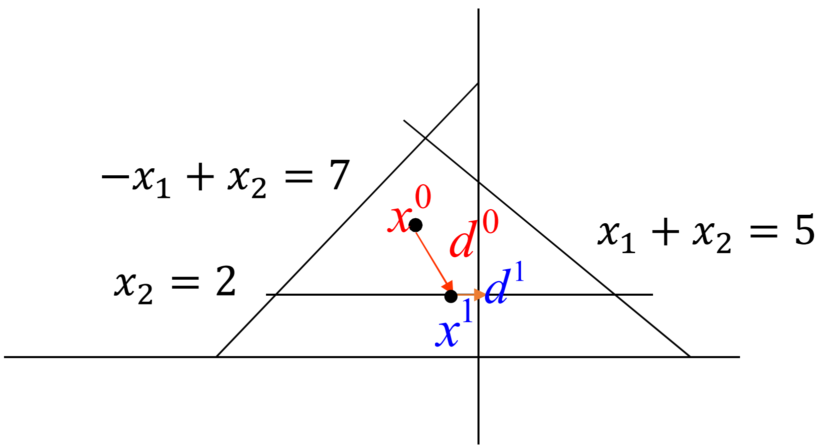
- $\nabla f(x) = [x_1, x_2]^T$
- $\varphi(\alpha) = f\left(x^k + \alpha d^k\right)$
- $k = 0$

$\varphi(\alpha) = f(-2 + \alpha, 3 - \alpha)$

Itération 1 : $x^0 + \alpha_0 d^0 = [-2{,}3]^T + \alpha_0[1,-1]^T = [-2 + \alpha_0, 3 - \alpha_0]^T.$

- $\varphi(\alpha) = \frac{1}{2}(13 - 10\alpha + 2\alpha^2) \Longrightarrow \frac{d\varphi}{d\alpha} = -5 + 2\alpha = 0 \Longrightarrow \alpha = \frac{5}{2} > 1$
- $\min\limits_{0\leq\alpha\leq 1} \varphi(\alpha) = \varphi(\alpha_0) = \frac{5}{2} \Longrightarrow \alpha_0 = 1$
- $\Longrightarrow x^1 = x^0 + \alpha_0 d^0 = [-2{,}3]^T + 1[1,-1]^T = [-2 + 1{,}3 - 1]^T = [-1{,}2]^T$

Méthode des direction réalisables: **Exemple**



$$(P)\begin{cases} \text{Min} \quad f(x) = \frac{1}{2}\left(x_1^2 + x_2^2\right) \\ \text{sujet à} \quad -x_1 + x_2 \le 7 \\ \qquad x_1 + x_2 \le 5 \\ \qquad -x_2 \le -2 \end{cases} \Longrightarrow \begin{cases} {a^1}^T = (-1,1) \\ {a^2}^T = (1,1) \\ {a^3}^T x^0 = (0,-1) \end{cases}$$

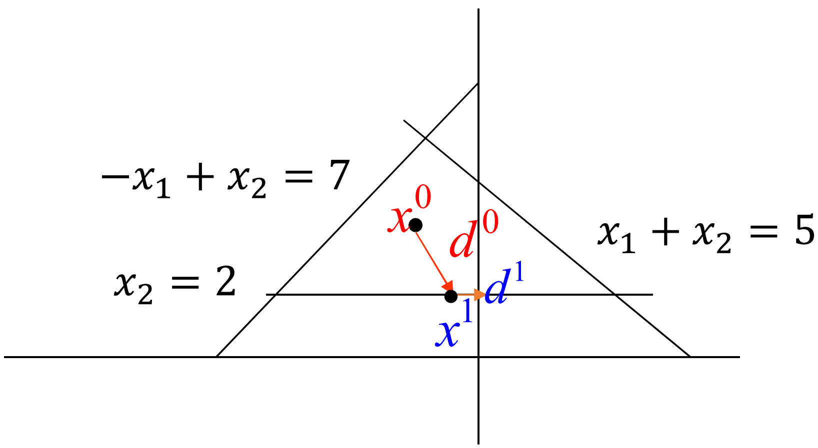
- $\nabla f(x) = [x_1, x_2]^T$
- ${a^1}^T x^1 = 5 < 7$
- ${a^2}^T x^2 = 1 < 5$
- ${a^3}^T x^3 = -2 = -2$

Itération 2 : $x^1 = [-1{,}2]^T$. Alors $I_1 = \{3\}$.
La direction $d^1 = [1{,}0]^T$ est la solution du problème :

$$(P_1)\begin{cases} Min\ \nabla f(x^1)^T d = -d_1 + 2d_2 \\ \qquad Sujet\ à \qquad -d_2 \le 0 \\ \qquad -1 \le d_j \le 1 \quad j = 1{,}2 \end{cases}$$

$$x^2 = x^1 + \alpha_1 d^1 = [-1{,}2]^T + \alpha_1 [1{,}0]^T = [-1 + \alpha_1, 2]^T.$$

Méthode des direction réalisables: **Exemple**



$$(P)\begin{cases} \text{Min} \quad f(x) = \frac{1}{2}\,(x_1^2 + x_2^2) \\ \text{sujet à} \quad -x_1 + x_2 \leq \quad 7 \\ \qquad\qquad x_1 + x_2 \leq \quad 5 \\ \qquad\qquad\quad -x_2 \leq -2 \end{cases} \Longrightarrow \begin{cases} {a^1}^T = (-1{,}1) \\ {a^2}^T = (1{,}1) \\ {a^3}^T x^0 = (0,-1) \end{cases}$$

- $\nabla f(x) = [x_1, x_2]^T$
- ${a^1}^T x^1 = 5 < 7$
- ${a^2}^T x^2 = 1 < 5$
- ${a^3}^T x^3 = -2 = -2$

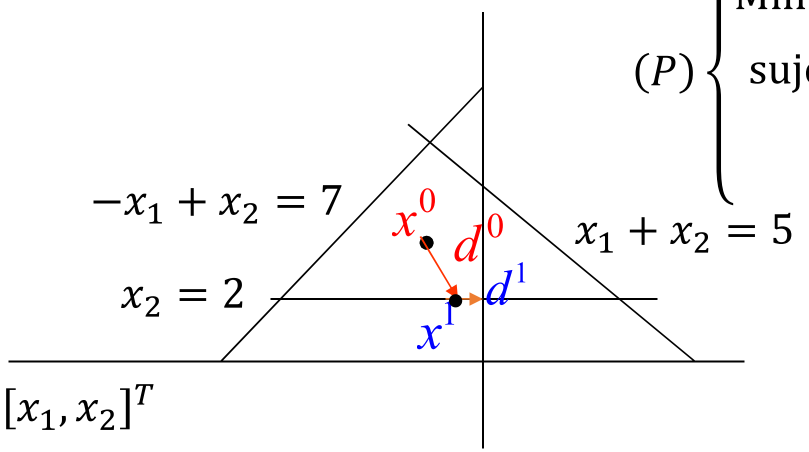
Itération 2 : $x^1 = [-1{,}2]^T.\, d^1 = [1{,}0]^T, x^2 = [-1 + \alpha_1, 2]^T$

La plus grande valeur $\bar{\alpha}_1 : x^2 = x^1 + \alpha_1 d^1 \in D_R$

$$\left.\begin{matrix} {a^1}^T d^1 = -1 \times 1 + 0 = -1 < 0 \\ {a^2}^T d^1 = 1 \times 1 + 0 = 1 > 0 \end{matrix}\right\} \Longrightarrow \quad \bar{\alpha}_1 = \frac{b_2 - {a^2}^T x^1}{{a^2}^T d^1} = \frac{5-1}{1} = 4$$

Méthode des direction réalisables: **Exemple**

$$(P)\begin{cases}\text{Min} \quad f(x) = \frac{1}{2}\left(x_1^2 + x_2^2\right) \\ \text{sujet à} \quad -x_1 + x_2 \leq 7 \\ \quad x_1 + x_2 \leq 5 \\ \quad -x_2 \leq -2\end{cases}$$



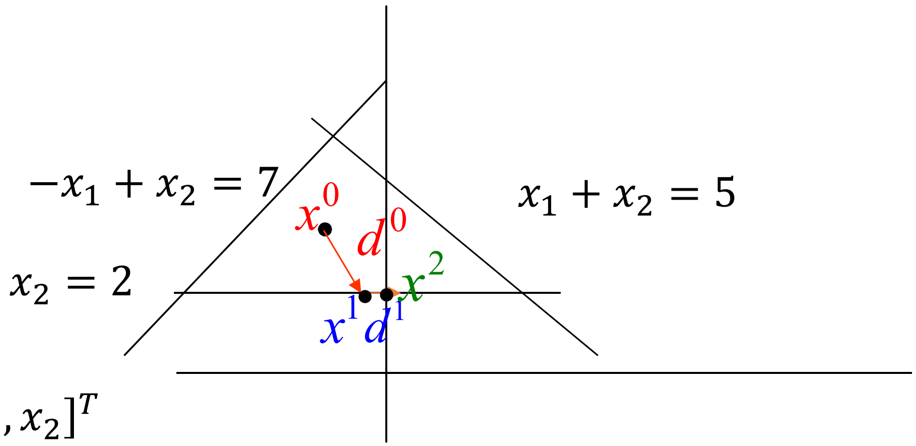
$\nabla f(x) = [x_1, x_2]^T$

Itération 2 : $x^1 = [-1,2]^T. d^1 = [1,0]^T, x^2 = [-1 + \alpha_1, 2]^T$

La plus grande valeur $\bar{\alpha}_1 : x^2 = x^1 + \alpha_1 d^1 \in D_R$

$$\left.\begin{array}{r} {a^1}^T d^1 = -1 \times 1 + 0 = -1 < 0 \\ {a^2}^T d^1 = 1 \times 1 + 0 = 1 > 0 \end{array}\right\} \Longrightarrow \quad \bar{\alpha}_1 = \frac{b_2 - {a^2}^T x^1}{{a^2}^T d^1} = \frac{5-1}{1} = 4$$

Méthode des direction réalisables: Exemple



$$(P)\begin{cases} \text{Min} \quad f(x) = \frac{1}{2}\,(x_1^2 + x_2^2) \\ \text{sujet à} \quad -x_1 + x_2 \le 7 \\ \qquad\qquad x_1 + x_2 \le 5 \\ \qquad\qquad -x_2 \le -2 \end{cases}$$

- $\nabla f(x) = [x_1, x_2]^T$
- $\varphi(\alpha) = f\left(x^k + \alpha d^k\right)$
- $k = 0$
- $\varphi(\alpha) = f(-1 + \alpha, 2)$

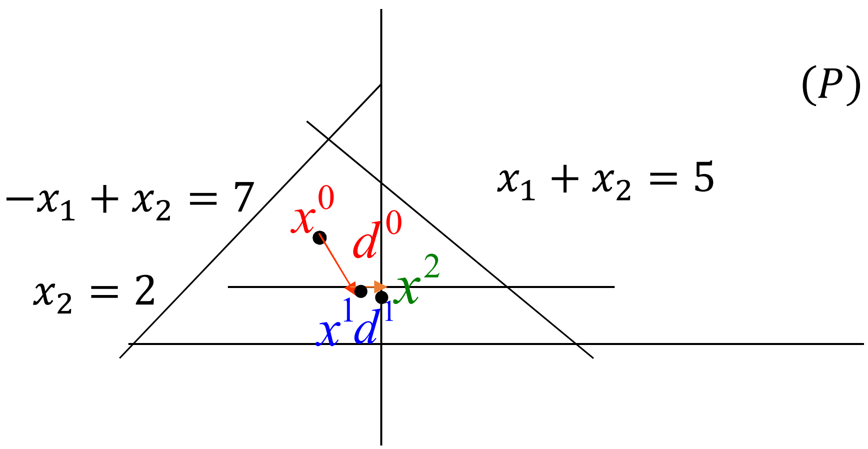
$\nabla f(x) = [x_1, x_2]^T$

Itération 2 : $x^1 = [-1,2]^T . d^1 = [1,0]^T, x^2 = [-1 + \alpha, 2]^T$

- $\varphi(\alpha) = f(-1 + \alpha_1, 2)$
- $\varphi(\alpha) = \frac{1}{2}(-1 + \alpha)^2 + (2)^2 \Longrightarrow \frac{d\varphi}{d\alpha} =- -2 + 4\alpha = 0 \Longrightarrow \alpha = 1$
- $\min\limits_{0 \le \alpha \le 4} \varphi(\alpha) \Longrightarrow \alpha_2 = 1.$

Donc $x^2 = [-1 + \alpha_1, 2]^T = [0,2]^T$

Méthode des direction réalisables: Exemple



$$(P)\begin{cases} \text{Min} \quad f(x) = \frac{1}{2}\,(x_1^2 + x_2^2) \\ \text{sujet à} \quad -x_1 + x_2 \le \ \ 7 \\ \qquad\qquad\quad x_1 + x_2 \le \ \ 5 \\ \qquad\qquad\qquad\ -x_2 \le -2 \end{cases}$$

$$\nabla f(x) = [x_1, x_2]^T$$
$$\nabla f(0{,}2) = [0{,}2\,]^T$$
$$\nabla f(0{,}2)^T d = 2d_2 = 0 \ge 0$$
$$-1 \le d_1 \le 1$$

Itération 3 : $x^2 = [0{,}2]^T$. Alors $I_1 = \{3\}$.

La direction $d^2 = [0{,}0]^T$ est la solution du problème :

$$(P_2)\begin{cases} Min\ \nabla f(x^2)^T d = 2d_2 \\ \qquad Sujet\ \text{à} \qquad\qquad -d_2 \le 0 \\ \qquad\qquad\qquad\qquad -1 \le d_j \le 1 \quad j = 1{,}2 \end{cases}$$

Donc $x^2 = [0{,}2]^T$ est une solution optimale de (P).

Algorithmes primales
Méthode de Frank-Wolfe

1. **Calcule de la direction**
2. **Calcule de pas optimal**
3. **Critère d'arrêt**
4. **Algorithme**
5. **Exemple:**

Méthode de Frank-Wolfe:

Calcule de la direction d^k

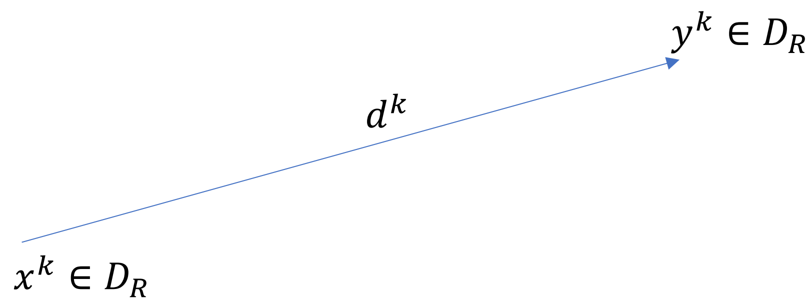
La méthode de Frank-Wolfe résout de façon itérative le problème $\min_{x \in D_R} f(x)$

Où D_R est un ensemble convexe de $\mathbb{R}^n$.

À une itération générale, étant donnée une solution réalisable x^k, la prochaine solution réalisable x^{k+1} est déterminée comme suit :

- Définissons : $y^k = d^k + x^k$
- $x^{k+1} = \lambda_k^* y^k + (1 - \lambda_k^*) x^k, \;\; \lambda_k^* \in [0,1]$
- $x^{k+1} = x^k + \lambda_k^* \left(y^k - x^k\right) = x^k + \lambda_k^* d^k$
- d^k est une direction de descente si y^k est une solution optimale du problème auxiliaire :

$$(PFW_k) \qquad \begin{cases} Min\ \nabla f\left(x^k\right)^T y \\ \quad Sujet\ \text{à} \qquad y \in D_R \end{cases}$$



- $x^{k+1} \in [x^k, y^k]$
- $\exists \lambda \in [0,1]$: $x^{k+1} = \lambda y^k + (1 - \lambda) x^k$
- $x^{k+1} = \lambda_k^* (d^k + x^k) + (1 - \lambda_k^*) x^k$
- $\nabla f\left(x^k\right)^T d = \nabla f\left(x^k\right)^T \left(y^k - x^k\right)$
- $\nabla f\left(x^k\right)^T d^k = \nabla f\left(x^k\right)^T y^k - \nabla f\left(x^k\right)^T x^k \leq 0$
- $\nabla f\left(x^k\right)^T y^k \leq \nabla f\left(x^k\right)^T x^k$
- Si $y^k = x^k \Leftrightarrow \nabla f\left(x^k\right)^T y^k = \nabla f\left(x^k\right)^T x^k$ stop
- $\Leftrightarrow \nabla f\left(x^k\right)^T d = \nabla f\left(x^k\right)^T y - \nabla f\left(x^k\right)^T x^k$

  $\min_d \nabla f\left(x^k\right)^T d \Leftrightarrow \min_{y \in D_R} \nabla f\left(x^k\right)^T y$

Méthode de Frank-Wolfe

Calcule de pas optimal λ_k^*

- $\varphi(\lambda_k) = f\left(\lambda_k y^k + (1-\lambda_k)x^k\right) = f\left(x^k + \lambda_k^* d^k\right)$
- $\lambda_k^* = \underset{0\leq\lambda_k\leq 1}{\operatorname{argmin}}\, \varphi(\lambda_k) \Leftrightarrow \varphi(\lambda_k^*) = \min_{0\leq\lambda_k\leq 1} \varphi(\lambda_k)$

Critère d'arrêt :

La méthode s'arrête à l'itération k lorsque x^k est une solution optimale de (PFW_k).

Sous l'hypothèse que f est convexe sur D_R, alors x^k est une solution optimale du problème lorsque le critère d'arrêt est satisfait. $\nabla f\left(x^k\right)^T x^k \leq \nabla f\left(x^k\right)^T y \Leftrightarrow \nabla f\left(x^k\right)^T\left(y - x^k\right) \geq 0 \quad \forall y \in D_R$

Et se référant à l'inégalité de gradient : $f(y) \geq f\left(x^k\right) + \nabla f\left(x^k\right)^T\left(y - x^k\right) \geq f\left(x^k\right) \qquad \forall y \in D_R$

Lorsque D_R est défini avec des contraintes linéaires, nous pouvons comparer cette méthode avec celle des directions réalisables. Nous notons alors les différences suivantes :

- $\lambda_k \in [0,1]$.
- Toutes les contraintes sont considérées dans le problème auxiliaire (PFW_k) pour déterminer y^k , alors que seules les contraintes actives dans I_k étaient utilisées dans le problème auxiliaire (P_k) de MDR.
- Si la méthode du simplexe est utilisé pour résoudre(PFW_k) , alors y^k est un point extrême appartenant à la frontière de D_R.

Méthode de Frank-Wolfe: Algorithme

Hypothèse: f est convexe sur D_R

Initialiser une solution initiale $x^0 \in D_R$.

METTRE $k = 0$

DO

- *Trouver une solution optimal* y^k *du problème*

$$(PFW_k) \qquad \begin{cases} Min\ \nabla f\left(x^k\right)^T y \\ \quad Sujet\ \text{à} \qquad\qquad y \in D_R \end{cases}$$

- $d^k = y^k - x^k$
- *Trouver* λ_k^* *la solution optimale du problème :*

$$\varphi(\lambda_k^*) = \min_{0 \le \lambda_k \le 1} \varphi(\lambda_k) = f\left(x^k + \lambda_k d^k\right)$$

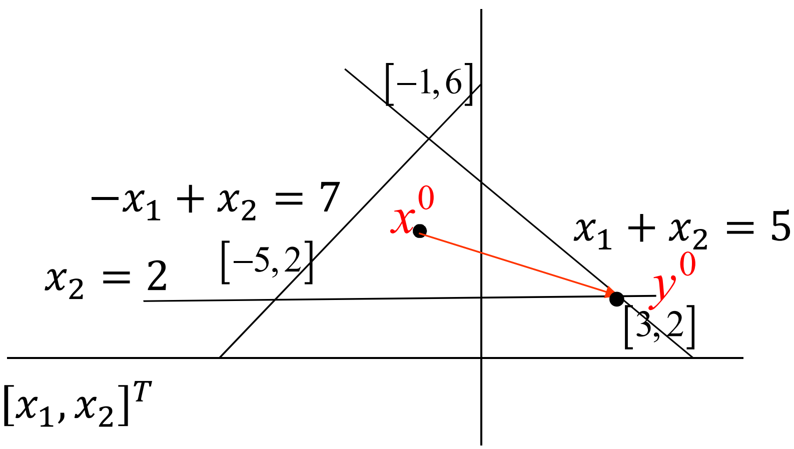
- $x^{k+1} = x^k + \lambda_k^* d^k$
- $k = k + 1$

WHILE x^k *est une solution optimale de* (PFW_k)*:* $\lambda_k^* < \tau_1$ ou $\left\|d^k\right\| < \tau_2$

- $x^* = x^k$
- RETURN x^*

Méthode de Frank-Wolfe: Exemple

$$(P)\begin{cases} \text{Min} \quad f(x) = \frac{1}{2}\left(x_1^2 + x_2^2\right) \\ \text{sujet à} \quad -x_1 + x_2 \leq 7 \\ \qquad\qquad x_1 + x_2 \leq 5 \\ \qquad\qquad -x_2 \leq -2 \end{cases}$$



$\nabla f(x) = [x_1, x_2]^T$

Itération 1 :$x^0 = [-2{,}3]^T, x^1 = \lambda_0 y^0 + (1 - \lambda_0)x^0,$

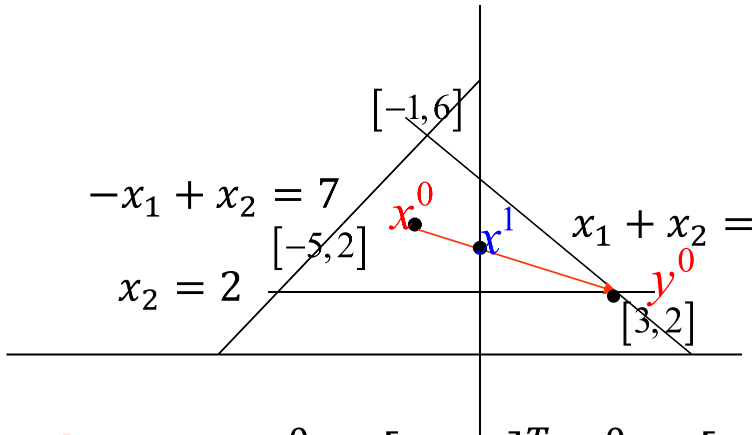
- $\lambda_0 = \underset{0 \leq \lambda \leq 1}{arg\min}\, \varphi(\lambda)$, $y^0 = [3{,}2]^T$ est une solution du problème :

$$(PFW_0)\begin{cases} Min\ \nabla f(x^0)^T y = -2y_1 + 3y_2 \\ \qquad Sujet\ à \qquad -y_1 + y_2 \leq 7 \\ \qquad\qquad\qquad y_1 + y_2 \leq 5 \\ \qquad\qquad\qquad -y_2 \leq -2 \end{cases}$$

Méthode de Frank-Wolfe: Exemple

$$(P)\begin{cases} \text{Min} \quad f(x) = \frac{1}{2}\,(x_1^2 + x_2^2) \\ \text{sujet à} \quad -x_1 + x_2 \leq 7 \\ \qquad\qquad x_1 + x_2 \leq 5 \\ \qquad\qquad -x_2 \leq -2 \end{cases}$$

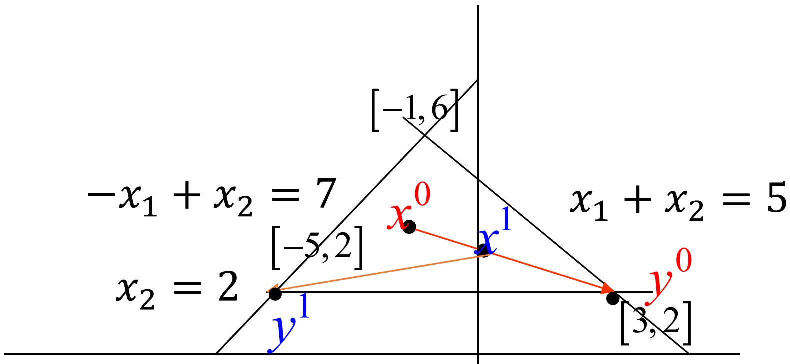
- $\nabla f(x) = [x_1, x_2]^T$



Itération1: $x^0 = [-2{,}3]^T, y^0 = [3{,}2]^T, x^1 = \lambda_0 y^0 + (1-\lambda_0)x^0, \qquad \lambda_0 = \underset{0\leq\lambda\leq 1}{argmin}\,\varphi(\lambda)$

- $\varphi(\lambda) = f(\lambda y^0 + (1-\lambda)x^0) = \frac{1}{2}\left((3\lambda - 2(1-\lambda))^2 + (2\lambda + 3(1-\lambda))^2\right)$
- $\varphi(\lambda) = \frac{1}{2}\left((5\lambda - 2)^2 + (-\lambda + 3)^2\right)$
- $\frac{d\varphi(\lambda)}{d\lambda} = 5(5\lambda - 2) - (-\lambda + 3) = 25\lambda - 10 + \lambda - 3 = 26\lambda - 13 = 0 \Longrightarrow \lambda_0 = \frac{1}{2}$
- $x^1 = \lambda_0 y^0 + (1-\lambda_0)x^0 = \left(-2\frac{1}{2} + 3\frac{1}{2}, 3\frac{1}{2} + 2\frac{1}{2}\right)^{\boldsymbol{T}} = \left(\frac{1}{2}, \frac{5}{2}\right)^T$

Méthode de Frank-Wolfe: Exemple



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

- $\nabla f(x) = [x_1, x_2]^T$

Itération 2 : $x^1 = \left[\frac{1}{2}, \frac{5}{2}\right]^T, x^2 = \lambda_1 y^1 + (1-\lambda_1)x^1,$

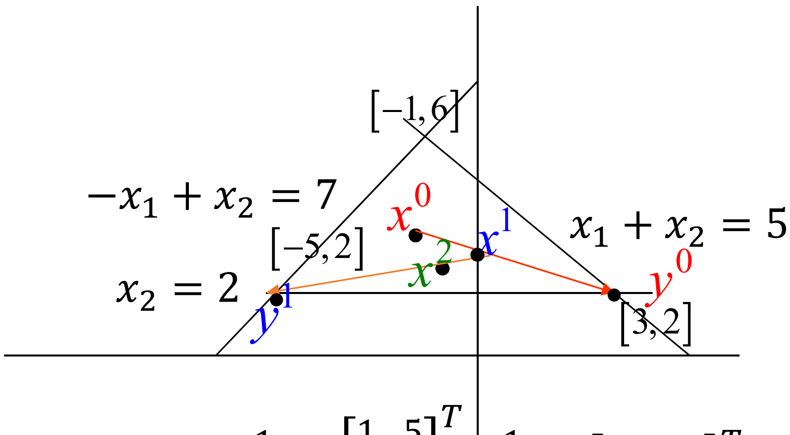
- $\lambda_1 = arg\min\limits_{0\leq\lambda\leq 1} \varphi(\lambda), y^1 = [-5{,}2]^T$ est la solution du problème :

$$(PFW_1)\begin{cases} Min\, \nabla f(x^1)^T y = \frac{1}{2}y_1 + \frac{5}{2}y_2 \\ Sujet\ \text{à} \qquad -y_1 + y_2 \leq 7 \\ \qquad\qquad y_1 + y_2 \leq 5 \\ \qquad\qquad -y_2 \leq -2 \end{cases}$$

Méthode de Frank-Wolfe: Exemple

$$(P)\begin{cases}\text{Min} & f(x)=\frac{1}{2}\left(x_1^2+x_2^2\right)\\ \text{sujet à} & -x_1+x_2\le 7\\ & x_1+x_2\le 5\\ & -x_2\le -2\end{cases}$$

- $\nabla f(x)=[x_1,x_2]^T$



Itération 2 : $x^1=\left[\frac{1}{2},\frac{5}{2}\right]^T y^1=[-5{,}2]^T, x^2=\lambda_1 y^1+(1-\lambda_1)x^1,\ \ \lambda_1=\underset{0\le\lambda\le1}{argmin}\,\varphi(\lambda)$

- $\varphi(\lambda)=f(\lambda y^1+(1-\lambda)x^1)=\frac{1}{2}\left(\left(-5\lambda+\frac{1}{2}(1-\lambda)\right)^2+\left(2\lambda+\frac{5}{2}(1-\lambda)\right)^2\right)$
- $\varphi(\lambda)=\frac{1}{2}\left(\left(-\frac{11}{2}\lambda+\frac{1}{2}\right)^2+\left(-\frac{1}{2}\lambda+\frac{5}{2}\right)^2\right)$
- $\frac{d\varphi(\lambda)}{d\lambda}=\frac{1}{4}\left(-(-121\lambda+11)-1(-\lambda+5)\right)=\frac{1}{4}(122\lambda-16)\Longrightarrow\lambda_1=\frac{16}{122}=0.13$
- $x^2=\left(-0.13\times5+0.87\frac{1}{2},0.13\times2+0.87\frac{5}{2}\right)^T=(-0.215{,}2.435)^T$

Méthode du gradient réduit

1. **Calcule de la direction d^k**
2. **Calcule plus grande valeur $\overline{\alpha}_k$:**
3. **Critère d'arrêt :**
4. **Algorithme**
5. **Exemple:**

Méthode du gradient réduit

Cette méthode a beaucoup de similarités avec l'algorithme du simplexe

Considérons le problème général de programmation mathématique linéaire suivant :

$$(PGR)\ \begin{cases} \text{Min} & f(x) \\ \text{sujet à} & Ax = b \\ & x \geq 0 \end{cases}$$

Où A est une matrice $m \times n$, $b \in \mathbb{R}^m$, $x \in \mathbb{R}^n$ et $f \in C^1/D_R$.

Comme dans la méthode des directions réalisables, à l'itération $(k+1)$ la direction d^k est déterminée en résolvant un problème auxiliaire (PGR_k). Pour définir le problème auxiliaire, associons à la solution x^k une base B_k de A (B_k $m \times m$ est non singulière). Supposons sans perte de généralité que B_k est constituée des m premières colonnes de A et R_k est constituée des $(n-m)$ dernières colonnes de A (R_k $m \times (n-m)$). Donc $A = [B_k \ \ R_k]$ et $x^k = \begin{pmatrix} x_{B_k} \\ x_{R_k} \end{pmatrix}$

- x_{B_k} est une variables de base et de x_{R_k} est une variables de hors base.
- Posons une hypothèse de non dégénérescence spécifiant que $x_{B_k}^k > 0$.

Méthode du gradient réduit

Étant donnée la partition précédente, ce problème s'écrit également :

$$(PGR)\begin{cases} \text{Min} \quad f\left(x_{B_k}, x_{R_k}\right) \\ \text{sujet à} \quad Ax = [B_k \ \ R_k]\begin{bmatrix} x_{B_k} \\ x_{R_k} \end{bmatrix} = b \\ \qquad x_{B_k} \geq 0 \ , x_{R_k} \geq 0 \end{cases}$$

B_k est une base de A et $[B_k \ \ R_k]\begin{bmatrix} x_{B_k} \\ x_{R_k} \end{bmatrix} = b$ alors :

$$B_k^{-1}[B_k \ \ R_k]\begin{bmatrix} x_{B_k} \\ x_{R_k} \end{bmatrix} = B_k^{-1}\left(B_k x_{B_k} + R_k x_{R_k}\right) = I x_{B_k} + B_k^{-1} R_k x_{R_k} = B_k^{-1} b \Longrightarrow x_{B_k} = B_k^{-1} b - B_k^{-1} R_k x_{R_k}$$

par conséquent en remplaçant x_{B_k}, (PGR) devient :

$$(PGR_k)\begin{cases} \text{Min} \quad F\left(x_{R_k}\right) = f\left(B_k^{-1} b - B_k^{-1} R_k x_{R_k}, x_{R_k}\right) \\ \text{sujet à} \quad x_{B_k} \geq 0 \ , x_{R_k} \geq 0 \end{cases}$$

Calcule de la direction $\mathbf{d^k}$

La direction d^k à l'itération $(k+1)$ est obtenue en considérant le problème (PGR_k)

La direction d^k est partitionnée de la même façon que x sous la forme $\boldsymbol{d^k} = \begin{bmatrix} \boldsymbol{d^k_{B_k}} \\ \boldsymbol{d^k_{R_k}} \end{bmatrix}$

Méthode du gradient réduit

- Nous allons d'abord déterminer $d_{R_k}^k$ en considérant le problème (PGR_k),
- puis nous allons déterminer $d_{B_k}^k$ de telle sorte qu'en prenant un pas $\alpha_k > 0$ dans la direction d^k,
- la nouvelle solution $x^{k+1} = x^k + \alpha_k\, d^k$ soit réalisable pour (PGR).

Nous voulons déterminer une valeur pour $d_{R_k}^k$.

- Qui soit une direction de descente de $F\left(x_{R_k}\right)$ au point $x_{R_k}^k$;
- Qui permette de prendre un pas $\alpha_k^{R_k} > 0$ de telle sorte que $x_{R_k}^{k+1} = x_{R_k}^k + \alpha_k^{R_k}\, d_{R_k}^k \geq 0$

Déterminons donc $d_{R_k}^k = \left[d_{m+1}^k, \dots, d_n^k\right]^T$ comme suit.

Pour $j = m+1, \dots, n$

$$d_j^k = \begin{cases} -\dfrac{dF\left(x_{R_k}^k\right)}{dx_j} & Si \quad \dfrac{dF\left(x_{R_k}^k\right)}{dx_j} < 0 \quad ou \ \ si \ x_j^k > 0 \\ 0 & autrement \end{cases}$$

Méthode du gradient réduit

- **$d_{R_k}^k$ est une direction de descente de $F(x_{R_k})$ au point $x_{R_k}^k$:**

Dénotons $J_k = \left\{ j: \frac{dF(x_{R_k}^k)}{dx_j} < 0 , x_j^k > 0, j = m+1, \dots, n \right\}$ Alors $\nabla_{R_k} F(x_{R_k}^k)^T d_{R_k}^k = \sum_{j \in J_k} \frac{dF(x_{R_k}^k)}{dx_j} d_j^k < 0$

A moins que $J_k = \emptyset$ dans lequel cas $d_{R_k}^k = 0$, ce qui est le critère d'arrêt de la méthode.

- **Calcule de $\alpha_k^{R_k}$ tel que $x_{R_k}^{k+1} = x_{R_k}^k + \alpha_k^{R_k} d_{R_k}^k \geq 0$**

Considérons $x_{R_k}(\alpha_k) = x_{R_k}^k + \alpha_k d_{R_k}^k$, pour $j = m+1, \dots, n$ on a :

$$x_j(\alpha_k) = \begin{cases} x_j^k - \alpha_k \dfrac{dF(x_{R_k}^k)}{dx_j} & Si \ \dfrac{dF(x_{R_k}^k)}{dx_j} < 0 \quad ou \ \ si \ x_j^k > 0 \\ x_j^k & autrement \end{cases}$$

- $x_{R_k(\alpha_k)} = x_{R_k}^k + \alpha_k d_{R_k}^k \geq 0 \Leftrightarrow x_j^k - \alpha_k \frac{dF(x_{R_k}^k)}{dx_j} \geq 0 \colon \forall j \in J_k$
- $\Leftrightarrow \alpha_k \leq \frac{x_j^k}{\frac{dF(x_{R_k}^k)}{dx_j}} \quad \forall j : \frac{dF(x_{R_k}^k)}{dx_j} > 0 \ et \ \ x_j^k > 0 \Leftrightarrow \alpha_k \leq \alpha_k^{R_k}$

Méthode du gradient réduit

Alors la valeur maximale de pas sur la direction $d_{R_k}^k$

$$\alpha_k^{R_k} = \min_{j=m+1,\dots,n}\left\{\frac{x_j^k}{\frac{dF\left(x_{R_k}^k\right)}{dx_j}} : \frac{dF\left(x_{R_k}^k\right)}{dx_j} > 0 \;\; et \;\; x_j^k > 0\right\}$$

Nous utilisons la dérivation en chaîne pour déterminer le gradient $\nabla_{R_k} F\left(x_{R_k}^k\right)$

- $F\left(x_{R_k}\right) = f\left(B_k^{-1}b - B_k^{-1}R_k x_{R_k}, x_{R_k}\right) \Longrightarrow \nabla_{R_k} F\left(x_{R_k}^k\right) = -\left(B_k^{-1}R_k\right)^T \; \nabla_{B_k} f\left(x^k\right) + \; \nabla_{R_k} f\left(x^k\right)$
- $\nabla_{B_k} f\left(x^k\right) = \left(\frac{df\left(x^k\right)}{dx_1}, \dots, \frac{df\left(x^k\right)}{dx_m}\right)^T, \quad \nabla_{R_k} f\left(x^k\right) = \left(\frac{df\left(x^k\right)}{dx_{m+1}}, \dots, \frac{df\left(x^k\right)}{dx_n}\right)^T$

Déterminons maintenant le vecteur des directions $\boldsymbol{d_{B_k}^k} = \left[\boldsymbol{d_1^k}, \dots, \boldsymbol{d_m^k}\right]$ associé aux variables de base.

Considérons un certain pas α_k dans la direction $d_{B_k}^k$ pour se déplacer à partir de $x_{B_k}^k$, et dénotons :

$$x_{B_k}(\alpha_k) = x_{B_k}^k + \alpha_k d_{B_k}^k \qquad (4)$$

Or $Ax(\alpha_k) = [B_k \quad R_k]\begin{bmatrix} x_{B_k}(\alpha_k) \\ x_{R_k}(\alpha_k)\end{bmatrix} = B_k\, x_{B_k}(\alpha_k) + R_k\, x_{R_k}(\alpha_k) = b$

- $x_{R_k}(\alpha_k) = x_{R_k}^k + \alpha_k d_{R_k}^k$
- $x_{B_k}(\alpha_k) = x_{B_k}^k + \alpha_k d_{B_k}^k \qquad (4)$
- $x_{B_k}(\alpha_k) = B_k^{-1} b - B_k^{-1} R_k\, x_{R_k}(\alpha_k)$

- $x_{B_{k}(\alpha_k)} = B_k^{-1} b - B_k^{-1} R_k\left(x_{R_k}^k + \alpha_k d_{R_k}^k \right)$
- $x_{B_{k}(\alpha_k)} = B_k^{-1} b - B_k^{-1} R_k\ x_{R_k}^k - \alpha_k B_k^{-1} R_k d_{R_k}^k$
- $x_{B_{k}(\alpha_k)} = x_{B_k}^k - \alpha_k B_k^{-1} R_k d_{R_k}^k$
- $d_{B_k}^k = -\alpha_k B_k^{-1} R_k d_{R_k}^k$

Méthode du gradient réduit

Et puisque B_k est une base alors:

- $B_k^{-1}\left(B_k\, x_{B_k}(\alpha_k) + R_k\, x_{R_k}(\alpha_k)\right) = I\, x_{B_k}(\alpha_k) + B_k^{-1} R_k\, x_{R_k}(\alpha_k) = B_k^{-1} b$
- $\Longrightarrow x_{B_k}(\alpha_k) = B_k^{-1} b - B_k^{-1} R_k\, x_{R_k}(\alpha_k) \qquad (5)$

Puisque $x_{R_k}(\alpha_k) = x_{R_k}^k + \alpha_k d_{R_k}^k$ alors en substituant dans (5):

$$x_{B_k(\alpha_k)} = B_k^{-1} b - B_k^{-1} R_k \left(x_{R_k}^k + \alpha_k d_{R_k}^k\right)$$

$$= B_k^{-1} b - B_k^{-1} R_k\; x_{R_k}^k - \alpha_k B_k^{-1} R_k d_{R_k}^k$$

$$= x_{B_k}^k - \alpha_k B_k^{-1} R_k d_{R_k}^k \quad \text{puisque } (x_{B_k} = B_k^{-1} b - B_k^{-1} R_k\, x_{R_k})$$

En comparant avec l’expression (4) et (5), il s’ensuit que

$$d_{B_k}^k = -B_k^{-1} R_k d_{R_k}^k$$

Méthode du gradient réduit

En somme pour déterminer la direction $d^k = \left(d_{B_k}^k, d_{R_k}^k\right)^T$,

a) Nous déterminons d'abord la composante $d_{R_k}^k$ à l'aide du problème (PGR_k) en s'assurant qu'elle soit une direction de descente de $F\left(x_{R_k}\right)$ au point $x_{R_k}^k$;

- $d_j^k = \begin{cases} -\dfrac{dF\left(x_{R_k}^k\right)}{dx_j} & Si \;\; \dfrac{dF\left(x_{R_k}^k\right)}{dx_j} < 0 \;\; ou \;\; si \;\; x_j^k > 0 \\ 0 & autrement \end{cases}$
- $\alpha_k^{R_k} = \min\limits_{j=m+1,\ldots,n} \left\{ \dfrac{x_j^k}{\frac{dF\left(x_{R_k}^k\right)}{dx_j}} : \dfrac{dF\left(x_{R_k}^k\right)}{dx_j} > 0 \;\; et \;\; x_j^k > 0 \right\}$

b) Nous spécifions ensuite $d_{B_k}^k$ en termes de $d_{R_k}^k$.

- $d_{B_k}^k = -B_k^{-1} R_k d_{R_k}^k$

C'est de là que vient le nom de la méthode.

Méthode du gradient réduit

Calcule plus grande valeur $\bar{\alpha}_k$:

Déterminons la plus grande valeur $\bar{\alpha}_k$ que peut prendre le pas pour que $x^{k+1} = x^k + \bar{\alpha}_k d^k$ demeure réalisable pour le problème (PGR).

Puisque les contraintes $Ax = b$ sont prises en considération en définissant la composante $d_{B_k}^k$ en termes de $d_{R_k}^k$, alors il suffit de s'assurer que $x(\alpha_k) = \left(x_{R_k}(\alpha_k), x_{B_k}(\alpha_k) \right)^T = x^k + \alpha_k d^k \geq 0$

Nous avons déjà analysé $x_{R_k}(\alpha_k)$

Pour $x_{B_k}(\alpha_k)$ rappelons que nous avons posé une hypothèse de non dégénérescence où $x_{B_k}^k > 0$.

Par conséquent : pour toute $j \in \{1, .., m\}$ tel que $d_j^k < 0$

- $x_{B_k}(\alpha_k) = x_{B_k}^k + \alpha_k d_{B_k}^k \geq 0 \Leftrightarrow x_j^k + \alpha_k d_j^k \geq 0$
- $\Leftrightarrow \alpha_k \leq \dfrac{x_j^k}{-d_j^k} \Leftrightarrow \alpha_k \leq \alpha_k^{B_k} = \min\limits_{j=1,\ldots,m} \left\{ \dfrac{x_j^k}{-d_j^k} : d_j^k < 0 \right\}$

Ainsi : $\bar{\alpha}_k = \min\{\alpha_k^{B_k}, \alpha_k^{R_k}\}$

Méthode du gradient réduit

Nous déterminons le pas optimal comme étant la valeur α_k^* telle que :

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

Analysons les deux cas où $\alpha_k^* = \alpha_k^{B_k}$ et ou $\alpha_k^* < \alpha_k^{B_k}$:

- Si $\alpha_k^* = \alpha_k^{B_k}$, alors une variable de base prend une valeur 0 au point x^{k+1}. Dans ce cas, la base pour la prochaine itération $(k+2)$ est obtenue à partir de la base actuelle en éliminant la colonne associée à la variable de base qui devient 0 pour la remplacer par la colonne associée à une variable actuellement hors base qui prend une valeur positive au point x^{k+1}(ce qui est toujours possible de faire sous l'hypothèse de non dégénérescence). Suivant la notation en programmation linéaire, la mise à jour du tableau des contraintes se fait à l'aide d'un pivot.

- Si $\alpha_k^* < \alpha_k^{B_k}$, alors nous utilisons la même base pour la prochaine itération $(k+2)$.

Méthode du gradient réduit

Critère d'arrêt :

La méthode s'arrête lorsque le vecteur de direction par rapport aux variables hors base $d_{R_k}^k = 0$. La justification de ce critère d'arrêt est présentée dans le prochain résultat.

Théorème 2 : Si $d_{R_k}^k = 0$, alors les conditions d'optimalité K-K-T pour le problème (PGR) sont satisfaites au point x^k.

Preuve : Rappelons que $d_{R_k}^k = \left[d_{m+1}^k, \dots, d_n^k\right]^T$ est défini comme suit :

$$\forall j = m+1, \dots, n \quad d_j^k = \begin{cases} -\dfrac{dF\left(x_{R_k}^k\right)}{dx_j} & \text{si } \dfrac{dF\left(x_{R_k}^k\right)}{dx_j} < 0 \quad \text{ou si } x_j^k > 0 \\ 0 & \text{autrement} \end{cases}$$

Par conséquent, si $d_j^k = 0$, alors $\left(\dfrac{dF\left(x_{R_k}^k\right)}{dx_j} = 0 \wedge x_j^k > 0\right) \vee \left(\dfrac{dF\left(x_{R_k}^k\right)}{dx_j} \geq 0 \wedge x_j^k = 0\right)$

Il s'ensuit que : $\nabla_{R_k} F\left(x_{R_k}^k\right) \geq 0 \quad et \quad \nabla_{R_k} F\left(x_{R_k}^k\right)^T x_{R_k}^k = 0$

Remplaçons $\nabla_{R_k} F\left(x_{R_k}^k\right)$ par son expression en termes du $\nabla f(x^k)$:

$$\nabla_{R_k} F\left(x_{R_k}^k\right)^T x_{R_k}^k = \left(\nabla_{R_k} f\left(x^k\right) - \left(B_k^{-1} R_k\right)^T \nabla_{B_k} f\left(x^k\right)\right)^T x_{R_k}^k = 0$$

Méthode du gradient réduit

Critère d'arrêt :

Introduisons également les deux relations trivialement satisfaites :

$$\nabla_{B_k} f(x^k) - I\ \nabla_{B_k} f(x^k) \geq 0 \quad et \quad \left(\nabla_{B_k} f(x^k) - I\ \nabla_{B_k} f(x^k)\right)^T x_{B_k}^k = 0$$

Ceci s'écrit également :

- $\begin{bmatrix} \nabla_{B_k} f(x^k) \\ \nabla_{R_k} f(x^k) \end{bmatrix} - \begin{bmatrix} I \\ (B_k^{-1} R_k)^T \end{bmatrix} \nabla_{B_k} f(x^k) \geq 0$
- $\Rightarrow \left(\begin{bmatrix} \nabla_{B_k} f(x^k) \\ \nabla_{R_k} f(x^k) \end{bmatrix} - \begin{bmatrix} I \\ (B_k^{-1} R_k)^T \end{bmatrix} \nabla_{B_k} f(x^k) \right)^T \begin{pmatrix} x_{B_k}^k \\ x_{R_k}^k \end{pmatrix} = 0$

En ajoutant les expressions $I x_{B_k}^k = B_k^{-1} b - B_k^{-1} R_k x_{R_k}^k \quad x_{B_k}^k \geq 0$ et $x_{R_k}^k \geq 0$

Nous retrouvons les conditions de K-K-T pour le problème :

$$\begin{cases} \text{Min} \quad f(x_{B_k}, x_{R_k}) \\ \text{sujet à} \quad x_{B_k} + B_k^{-1} R_k x_{R_k} = B_k^{-1} b \\ \qquad\qquad x_{B_k} \geq 0 \ , x_{R_k} \geq 0 \end{cases}$$

(Equivalant au problème (PGR)) où le vecteur de multiplicateurs est égal à $-\nabla_{B_k} f(x^k)$.

Méthode du gradient réduit

Critère d'arrêt :

Transformons les conditions comme suit et Regroupons les termes $\left(B_k^{-1}\right)^T$ et $\nabla_{B_k} f\left(x^k\right)$:

$$\left\{\begin{array}{lcl} \left(\begin{pmatrix} \nabla_{B_k} f\left(x^k\right) \\ \nabla_{R_k} f\left(x^k\right) \end{pmatrix} - \begin{pmatrix} B_k^{-1}B_k \\ B_k^{-1}R_k \end{pmatrix}^T \nabla_{B_k} f\left(x^k\right)\right) & \geq & 0 \\ \left(\begin{pmatrix} \nabla_{B_k} f\left(x^k\right) \\ \nabla_{R_k} f\left(x^k\right) \end{pmatrix} - \begin{pmatrix} B_k^{-1}B_k \\ B_k^{-1}R_k \end{pmatrix}^T \nabla_{B_k} f\left(x^k\right)\right)^T \begin{pmatrix} x_{B_k}^k \\ x_{R_k}^k \end{pmatrix} & = & 0 \\ B_k\left(I x_{B_k}^k + B_k^{-1}R_k x_{R_k}^k\right) & = & B_k B_k^{-1} b \\ \begin{pmatrix} x_{B_k}^k \\ x_{R_k}^k \end{pmatrix} & \geq & 0 \end{array}\right. \Rightarrow \left\{\begin{array}{lcl} \left(\begin{pmatrix} \nabla_{B_k} f\left(x^k\right) \\ \nabla_{R_k} f\left(x^k\right) \end{pmatrix} - \begin{pmatrix} B_k \\ R_k \end{pmatrix}^T \left(B_k^{-1}\right)^T \nabla_{B_k} f\left(x^k\right)\right) & \geq & 0 \\ \left(\begin{pmatrix} \nabla_{B_k} f\left(x^k\right) \\ \nabla_{R_k} f\left(x^k\right) \end{pmatrix} - \begin{pmatrix} B_k \\ R_k \end{pmatrix}^T \left(B_k^{-1}\right)^T \nabla_{B_k} f\left(x^k\right)\right)^T \begin{pmatrix} x_{B_k}^k \\ x_{R_k}^k \end{pmatrix} & = & 0 \\ B_k x_{B_k}^k + R_k x_{R_k}^k & = & b \\ \begin{pmatrix} x_{B_k}^k \\ x_{R_k}^k \end{pmatrix} & \geq & 0 \end{array}\right.$$

Ces relations représentent les conditions d'optimalité K-K-T pour le problème (PGR) au point x^k où le vecteur de multiplicateurs est égal à $-\left(B_k^{-1}\right)^T \nabla_{B_k} f\left(x^k\right)$.

Méthode du gradient réduit: Algorithme(hypothèse de non dégénérescence)

Initialiser une solution initiale $x^0 \in D_R$.

METTRE $k = 0$

DO

- $\nabla_{R_k} F(x_{R_k}^k) = -\left(B_k^{-1} R_k\right)^T \nabla_{B_k} f(x^k) + \nabla_{R_k} f(x^k)$
- $d_{R_k}^k = \left(-\nabla_{R_k} F(x_{R_k}^k)\right)^+ \Longrightarrow \alpha_k^{R_k} = \min\limits_{j=m+1,\dots,n} \left\{ x_j^k \Big/ \frac{dF(x_{R_k}^k)}{dx_j} : \frac{dF(x_{R_k}^k)}{dx_j} > 0 \ \ et \ \ x_j^k > 0 \right\}$
- $d_{B_k}^k = -\alpha_k B_k^{-1} R_k d_{R_k}^k,\ \forall \alpha_k \in [0, \alpha^{R_k}] \Longrightarrow \alpha_k^{B_k} = \min\limits_{j=1,\dots,m} \left\{ x_j^k \Big/ -d_j^k : d_j^k < 0 \right\}$
- $\bar{\alpha}_k = \min\{\alpha_k^{B_k}, \alpha_k^{R_k}\}$
- *Trouver* α_k^* *la solution optimale du problème :* $f(x^k + \alpha_k^* d^k) = \min\limits_{0 \le \alpha_k \le \bar{\alpha}_k} \{f(x^k + \alpha_k d^k\}$
- $x^{k+1} = x^k + \alpha_k^* d^k$
- IF $\alpha_k^* = \alpha_k^{B_k}$ THEN Mise a jour de base ELSE $\alpha_k^* < \alpha_k^{B_k}$ utilisons la même base ENDIF
- $k = k + 1$

WHILE $\left\|d_{R_k}^k\right\| > \tau$

- $x^* = x^k$
- RETURN x^*

Méthode du gradient réduit: Exemple

$$\begin{cases} Min \quad f(x_1,x_2) = (x_1)^2+3x_1x_2+4(x_2)^2 \\ s.t \qquad\qquad x_1+x_2=1 \\ \qquad\qquad\quad x_1 \geq 0, x_2 \geq 0 \end{cases}$$

- $A=(1,1)$, $\mathrm{B}=1$, $\mathrm{R}=1 \Longrightarrow B^{-1}1 = B^{-1}\mathrm{R} = 1$
- $\mathrm{A}x = \mathrm{B}x_2 + \mathrm{R}x_1 = 1 \Longrightarrow x_2 = B^{-1}1 - B^{-1}\mathrm{R}x_1$

Itération 1 : $x^0 = [0,1]^T$, x_2 est une variable de base. $x_2 = 1 - B^{-1}Rx_1 \Longrightarrow F(x_1) = f(x_1, 1 - x_1)$

Par conséquent $\frac{dF(x_1)}{dx_1} = \frac{df(x_1,x_2)}{dx_1} - 1\,\frac{df(x_1,x_2)}{dx_2} = (2x_1+3x_2) - (3x_1+8x_2) = -x_1 - 5x_2$

Donc $\frac{dF(x_1^0)}{dx_1} = -5 \quad \Longrightarrow \quad \boldsymbol{d_{R_0}^0} = -\frac{dF(x_1^0)}{dx_1} = 5$

Calculons $\boldsymbol{d_{B_0}^0}$: $\quad d_{B_0}^0 = -B_0^{-1}R_0d_1^0 = -1\times 1\times 5 = -5$

Déterminons $\boldsymbol{\alpha_1^{B_0}}$ et $\boldsymbol{\alpha_1^{R_0}}$:

- $x_1(\alpha_1) = x_1^0 + \alpha_0 d_1^0 = 0 + 5\alpha_0 \geq 0 \;\forall\, \alpha_0 \;\Longrightarrow\; \alpha_0^{R_0} = \infty$
- $x_2(\alpha_1) = x_2^0 + \alpha_0 d_2^0 = 1 - 5\alpha_0 \geq 0 \Leftrightarrow \alpha_0 \leq \frac{1}{5} = \alpha_0^{B_0} \Longrightarrow \boldsymbol{\bar{\alpha}_0 = min\{\alpha_0^{B_0}, \alpha_0^{R_0}\} = \alpha_0^{B_0} = \frac{1}{5}}$

On a $\varphi(\alpha_0) = f(x_1^0+\alpha_0 d_1^0, x_2^0+\alpha_0 d_2^0) = (5\alpha_0)^2 + 3\times 5\alpha_0(1-5\alpha_0) + 4(1-5\alpha_0)^2 = 50\alpha_0^2 - 25\,\alpha_0 + 4$

- $\varphi'(\alpha_0) = 100\alpha_0 - 25 = 0 \;\Longrightarrow\; \alpha_0 = \frac{1}{4} > \frac{1}{5}$

Méthode du gradient réduit: Exemple

Et Alors $\min\limits_{0\le\alpha_1\le\frac{1}{5}} \varphi(\alpha_0) = \varphi\left(\frac{1}{5}\right) = 1,$ donc le pas optimal $\alpha_0^* = \bar{\alpha}_0 = \alpha_0^{B_0} = \frac{1}{5}$ et

- $x_1^1 = x_1^0 + \alpha_0^* d_1^0 = 0 + \frac{1}{5}5 = 1$
- $x_2^1 = x_2^0 + \alpha_0^* d_2^0 = 1 - \frac{1}{5}5 = 0$

Ainsi $x^1 = [1,0]^T$. De plus puisque $\alpha_0^* = \alpha_0^{B_0}$, il faut modifier la base en remplaçant x_2 dans la base par x_1.

Itération 2 : $x^1 = [1,0]^T$, x_1 est la variable de base. $x_1 = 1 - x_2 \Longrightarrow F(x_2) = f(1 - x_2, x_2)$

Par conséquent :

- $\frac{dF(x_2)}{dx_2} = \frac{df(x_1,x_2)}{dx_2} - 1\,\frac{df(x_1,x_2)}{dx_1} = (3x_1 + 8x_2) - (2x_1 + 3x_2) = x_1 + 5x_2$

Donc $\frac{dF(x^1)}{dx_2} = 3 - 2 = 1 > 0 \Longrightarrow\ d_2^1 = 0$

Le critère d'arrêt est satisfait, et par le théorème (2) il s'ensuit que $x^1 = [1,0]^T$ satisfait les conditions d'optimalité K-K-T pour le problème.

De plus, puisque f est convexe, alors $x^1 = [1,0]^T$ est une solution optimale du problème.

Méthode du gradient projeté

1. Calcule de la direction d^k
2. Algorithme
3. Exemple:

Méthode du gradient projeté:

Considérons le problème (P) où $f \in C^1/D_R$ et où $D_R = \left\{x \in \mathbb{R}^n \middle| {a^i}^T x \leq b_i, i = 1, \dots, m\right\}$:

$$(P) \begin{cases} \text{Min} & f(x) \\ \text{sujet à} & x \in D_R \end{cases}$$

Etant donné une solution réalisable x^kde (P) définissons l'ensemble des indices des contraintes actives :

$$I_k = \left\{i: {a^i}^T x^k = b_i\right\}$$

Calcule de la direction d^k

La direction de descente d^k est obtenue en projetant le vecteur $-\nabla f(x^k)$ sur le sous espace

$$S_k = \left\{x \in \mathbb{R}^n:\ {a^i}^T x^k = 0, i \in I_k\right\}$$

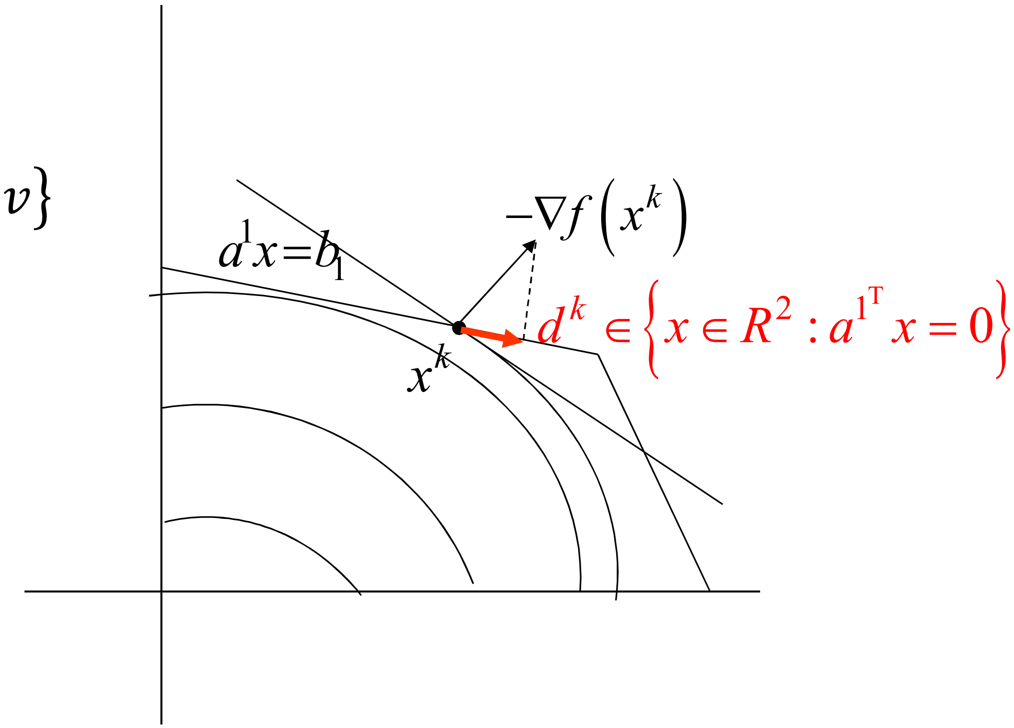
Pour dériver une formulation explicite de d^k, dénotons A_k la matrice $|I_k| \times n$ dont les lignes sont les vecteurs$({a^i}^T, i \in I_k)$. De plus supposons que les vecteurs $({a^i}^T, i \in I_k)$ sont linéairement indépendants de sorte que la matrice A_k soit de rang $|I_k| < n$.

Ainsi, le sous espace $S_k$est orthogonal au sous espace $S_k^\perp = \left\{x \in \mathbb{R}^n:\ \exists v \in \mathbb{R}^{|I_k|} \text{ tel que } x = A_k^T v\right\}$

et de plus : $\mathbb{R}^n = S_k + S_k^\perp$

Méthode du gradient projeté:

- $S_k = \left\{x \in \mathbb{R}^n\colon\ {a^i}^T x^k = 0, i \in I_k\right\}$
- $S_k^\perp = \ \left\{x \in \mathbb{R}^n\colon\ \ \exists v \in \mathbb{R}^{|I_k|} \text{ tel que } \ x = A_k^T\ v\right\}$
- $\mathbb{R}^n = S_k + S_k^\perp$



En effet, si $y \in S_k$ et $t \in S_k^\perp$, alors
il existe un $v(t)$ tel que $t = A_k^T\ v(t)$, et ainsi :

$$y^T t = y^T A_k^T v(t) = 0^T v(t) = 0.$$

Puisque d^k est la projection de $-\nabla f(x^k)$ sur S_k,
alors $d^k \in S_k$ et il existe $v^k \in \mathbb{R}^{|I_k|}$ tel que

$$-\nabla f\left(x^k\right) = d^k + A_k^T v^k$$

Méthode du gradient projeté:

- $S_k = \left\{x \in \mathbb{R}^n\colon\ {a^i}^T x^k = 0, i \in I_k\right\}$
- $S_k^\perp = \left\{x \in \mathbb{R}^n\colon\ \ \exists v \in \mathbb{R}^{|I_k|} \text{ tel que } x = A_k^T\, v\right\}$
- $\mathbb{R}^n = S_k + S_k^\perp$

En effet, si $y \in S_k$ et $t \in S_k^\perp$, alors
il existe un $v(t)$ tel que $t = A_k^T\, v(t)$, et ainsi :

$$y^T t = y^T A_k^T v(t) = 0^T v(t) = 0.$$

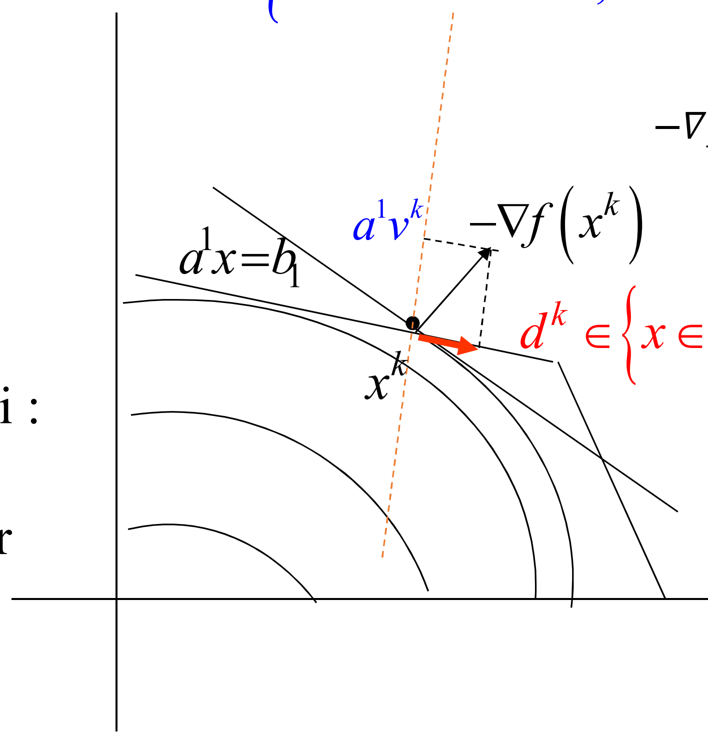
Puisque d^k est la projection de $-\nabla f(x^k)$ sur S_k,
alors $d^k \in S_k$ et il existe $v^k \in \mathbb{R}^{|I_k|}$ tel que

$$-\nabla f\left(x^k\right) = d^k + A_k^T v^k$$

$$-\nabla f\left(x^k\right) \in \mathbb{R}^n = d^k + A_k^T v^k$$

$\left\{x \in R^2 : x = a^1 v,\ \ v \in R^1\right\}$

$-\nabla f\left(x^k\right) \in \mathbb{R}^n = d^k + A_k^T v^k$



Méthode du gradient projeté:

Ainsi : $-\nabla f(x^k) = d^k + A_k^T v^k \Longrightarrow -A_k \nabla f(x^k) = A_k d^k + A_k A_k^T v^k$

Puisque $A_k d^k = 0$ alors $-A_k \nabla f(x^k) = A_k A_k^T v^k$

Ceci implique que: $\boldsymbol{v^k = -(A_k A_k^T)^{-1} A_k \nabla f(x^k)}$

Et par conséquent : $\boldsymbol{d^k = -\nabla f(x^k) - A_k^T v^k = -[I - A_k^T (A_k A_k^T)^{-1} A_k] \nabla f(x^k)}$

(Notons que la matrice $A_k A_k^T$ de dimension $|I_k| \times |I_k|$ est non singulière puisque la matrice $A_k$de dimension $|I_k| \times n$ est de plein rang par hypothèse).

Nous devons analyser les deux cas où $d^k = 0$ et $d^k \neq 0$.

Cas 1 : Si $d^k \neq 0$, alors d^k est une direction de descente de f au point x^k.

En effet $\nabla f(x^k)^T d^k = (\nabla f(x^k) + d^k - d^k)^T d^k = (\nabla f(x^k) + d^k)^T d^k - {d^k}^T d^k$

Or puisque $A_k^T v^k = -\nabla f(x^k) - d^k$ est orthogonal à d^k, alors $(\nabla f(x^k) + d^k)^T d^k = 0$,

Et il s'ensuit $\nabla f(x^k)^T d^k = (\nabla f(x^k) + d^k)^T d^k - {d^k}^T d^k = -{d^k}^T d^k < 0 \quad si \;\; d^k \neq 0$.

Méthode du gradient projeté:

Cas 2 : Si $d^k = 0$, alors il y a deux possibilités.

i. x^k satisfait les conditions d'optimalité K-K-T pour le problème (P).

ii. une nouvelle direction de descente $\tilde{d}^k$ peut être définie.

i) Supposons que $v^k \geq 0$.

Dénotons $\bar{A}_k$:$|n - I_k| \times n$ dont les lignes sont les vecteurs ${a^i}^T$ dont les indices $i \notin I_k$. Alors la matrice de

contrainte A du problème (P), $A = \begin{bmatrix} A_k \\ \bar{A}_k \end{bmatrix}$ (en supposant que les $|I_k|$ contraintes actives sont les $|I_k|$ premières).

Considérons v^k comme étant le vecteur des multiplicateurs associés aux contraintes actives, et supposons que les multiplicateurs associés aux autres contraintes sont tous égaux à 0. Alors les K-K-T sont vérifiées à x^k

- le gradient du lagrangien $\nabla f(x^k) + \begin{bmatrix} A_k^T & \bar{A}_k^T \end{bmatrix} \begin{bmatrix} v^k \\ 0 \end{bmatrix} = \nabla f(x^k) + A_k^T v^k = -d^k = 0$
- $Ax^k \leq 0$
- $v_k^T A_k x^k = 0$ $\left(\text{puisque } A_k x^k = 0\right)$
- $0^T \bar{A}_k x^k = 0$
- $v^k \geq 0$

Méthode du gradient projeté:

Cas 2 : Si $d^k = 0$, alors il y a deux possibilités.
ii) supposons que $v^k \ngeq 0$. Alors il existe un indice $\tilde{\imath} \in I_k$ tel que $v_{\tilde{\imath}}^k < 0$.
Dénotons :

- $\tilde{I}_k = I_k - \tilde{\imath}$
- $\tilde{A}_k$ la matrice obtenue en éliminant la ligne $\tilde{\imath}$ de A_k.
- $\tilde{S}_k = \left\{x \in \mathbb{R}^n \colon {a^i}^T x = 0, i \in \tilde{I}_k\right\}$

Et considérons le théorème suivant.

Théorème 3 :

Si $\tilde{d}^k$ est la projection de $-\nabla f\left(x^k\right)$ sur le sous espace $\tilde{S}_k$ alors il possède les propriétés suivantes :

- $\tilde{d}^k \neq 0$
- $\nabla f(x^k)^T \tilde{d}^k < 0$ (direction de descente)
- $a_{\tilde{\imath}}^T \tilde{d}^k < 0$ (direction réalisable)

Méthode du gradient projeté:

Preuve :

Puisque $\tilde{d}^k$ est la projection de $-\nabla f\left(x^k\right)$ sur le sous espace $\tilde{S}_k$, alors $\exists \tilde{v}^k \in \mathbb{R}^{|\tilde{I}_k|}$ tel que $\tilde{d}^k = -\nabla f\left(x^k\right) - \tilde{A}_k^T \tilde{v}^k$. De plus, puisque $d^k = 0$ alors $-\nabla f\left(x^k\right) = A_k^T v^k$. Ce qui implique

$$-\nabla f\left(x^k\right) = A_k^T v^k \ (6)$$

$$-\nabla f\left(x^k\right) = \tilde{d}^k + \tilde{A}_k^T \tilde{v}^k \quad (7)$$

a) (Preuve par contradiction) $-\nabla f\left(x^k\right) = A_k^T v^k = \tilde{A}_k^T \tilde{v}^k \quad (8)$

Dénotons par $\hat{v}^k$ le vecteur obtenu de v^k en éliminant la composante $v_{\tilde{i}}^k$.

Alors $A_k^T v^k = a_i v_{\tilde{i}}^k + \tilde{A}_k^T \hat{v}^k$,

Et combinant avec (8), $a_i v_{\tilde{i}}^k + \tilde{A}_k^T \hat{v}^k = \tilde{A}_k^T \tilde{v}^k \Longrightarrow a_i v_{\tilde{i}}^k = \tilde{A}_k^T (\tilde{v}^k - \hat{v}^k)$

Et il s'ensuit que a_i s'exprime comme une combinaison linéaire des colonnes de $\tilde{A}_k^T$, contredisant que $Rang(A_k) = |I_k|$.

Méthode du gradient projeté:

Preuve :

b) Du fait que $\tilde{d}^k \neq 0$, le même argument qu'au Cas 1 s'applique.

$$\nabla f(x^k)^T \tilde{d}^k = \left(\nabla f(x^k) + \tilde{d}^k - \tilde{d}^k\right)^T \tilde{d}^k = \left(\nabla f(x^k) + \tilde{d}^k\right)^T \tilde{d}^k - \tilde{d}^{k^T} \tilde{d}^k.$$

Or puisque $\tilde{A}_k^T \tilde{v}^k = -\nabla f(x^k) - \tilde{d}^k$ est orthogonal à $\tilde{d}^k$, alors $\left(\nabla f(x^k) + \tilde{d}^k\right)^T \tilde{d}^k = 0$,

Et il s'ensuit que si $\tilde{d}^k \neq 0$, alors $\nabla f(x^k)^T \tilde{d}^k = \left(\nabla f(x^k) + \tilde{d}^k\right)^T \tilde{d}^k - \tilde{d}^{k^T} \tilde{d}^k = -\tilde{d}^{k^T} \tilde{d}^k < 0$

c) Il se découle de (6) et de la partie ***b***) que : $\nabla f(x^k)^T \tilde{d}^k = -v^{k^T} A_k \tilde{d}^k < 0. \quad (9)$

Considérons le vecteur $\hat{v}^k$ obtenu de v^k en éliminant la composante $v_{\tilde{i}}^k$,

$$v^{k^T} A_k \tilde{d}^k = \hat{v}^{k^T} \tilde{A}_k^T \tilde{d}^k + v_{\tilde{i}}^k a_{\tilde{i}}^T \tilde{d}^k \qquad (10)$$

Puisque $\tilde{A}_k \tilde{d}^k = 0$ étant donné que $\tilde{d}^k$ est la projection de $-\nabla f(x^k)$ sur le sous espace $\tilde{S}_k$.

$$v_{\tilde{i}}^k a_{\tilde{i}}^T \tilde{d}^k > 0,$$

Et puisque par hypothèse $v_{\tilde{i}}^k < 0$, il s'ensuit que $a_{\tilde{i}}^T \tilde{d}^k < 0$.

5- Méthode du gradient projeté:

Une fois que la direction est déterminée, il suffit de poursuivre comme dans la méthode des directions réalisables pour déterminer :

a) le plus grand pas dans cette direction

b) le pas optimale dans cette direction.

Méthode du gradient projeté: Algorithme

Initialiser une solution initiale $x^0 \in D_R$. $k = 0$

DO

-METTRE le vecteur des multiplicateurs associés aux contraintes actives

$$v^k = -(A_k A_k^T)^{-1} A_k \nabla f\left(x^k\right)$$

-METTRE direction $d^k = -\nabla f\left(x^k\right) - A_k^T v^k = -\left[I - A_k^T (A_k A_k^T)^{-1} A_k\right] \nabla f\left(x^k\right)$

-Déterminer $\bar{\alpha}_k$ la plus grande valeur que peut prendre le pas $\alpha_k \geq 0$

$$\bar{\alpha}_k = \min_{i \notin I_k} \left\{ \frac{b_i - {a^i}^T x^k}{{a^i}^T d^k} : {a^i}^T d^k > 0 \right\}$$

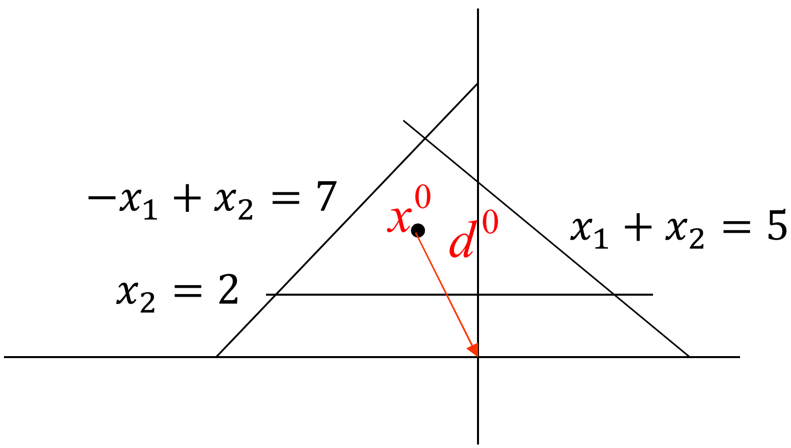
-*Trouver* α_k^* *la solution optimale du problème* : $f\left(x^k + \alpha_k^* d^k\right) = \min_{0 \leq \alpha_k \leq \bar{\alpha}_k} \{f(x^k + \alpha_k d^k\}$

- $k = k + 1$

WHILE $\left\|d^k\right\| > \tau$

- $x^* = x^k$
- RETURN x^*

Méthode du gradient projeté: Exemple



$$(P)\begin{cases} \text{Min} \quad f(x) = \frac{1}{2}(x_1^2 + x_2^2) \\ \text{sujet à} \quad -x_1 + x_2 \leq 7 \\ \qquad x_1 + x_2 \leq 5 \\ \qquad -x_2 \leq -2 \end{cases}$$

$$\nabla f(x) = [x_1, x_2]^T$$

Itération 1 : $x^0 = [-2,3]^T$. Alors $I_0 = \emptyset$.

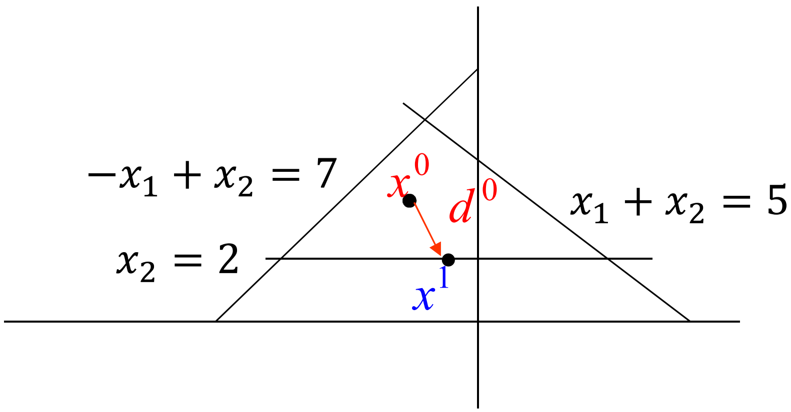
La direction $d^0 = -\nabla f(x^0) = [2,-3]^T$

$x^0 + \alpha_0 d^0 = [-2,3]^T + \alpha_0[2,-3]^T = [-2 + 2\alpha_0, 3 - 3\alpha_0]^T$

La plus grande valeur $\bar{\alpha}_0$:

$$\left.\begin{array}{l} {a^1}^T d^0 = -1 \times 2 - 1 \times 3 = -5 < 0 \\ {a^2}^T d^0 = \quad 1 \times 2 - 1 \times 3 = -1 < 0 \\ {a^3}^T d^0 = -1 \times -3 = 3 > 0 \end{array}\right\} \Rightarrow \bar{\alpha}_0 = \frac{b_3 - {a^3}^T x^0}{{a^3}^T d^0} = \frac{-2+3}{3} = \frac{1}{3}$$

Méthode du gradient projeté: Exemple



$$(P)\begin{cases} \text{Min} \quad f(x) = \frac{1}{2}\,(x_1^2 + x_2^2) \\ \text{sujet à} \quad -x_1 + x_2 \leq \ \ 7 \\ \qquad\qquad\quad x_1 + x_2 \leq \ \ 5 \\ \qquad\qquad\qquad -x_2 \leq -2 \end{cases}$$

$$\nabla f(x) = [x_1, x_2]^T$$

Itération 1 : $x^0 + \alpha_0 d^0 = [-2,3]^T + \alpha_0[2,-3]^T = [-2+2\alpha_0, 3-3\alpha_0]^T$.

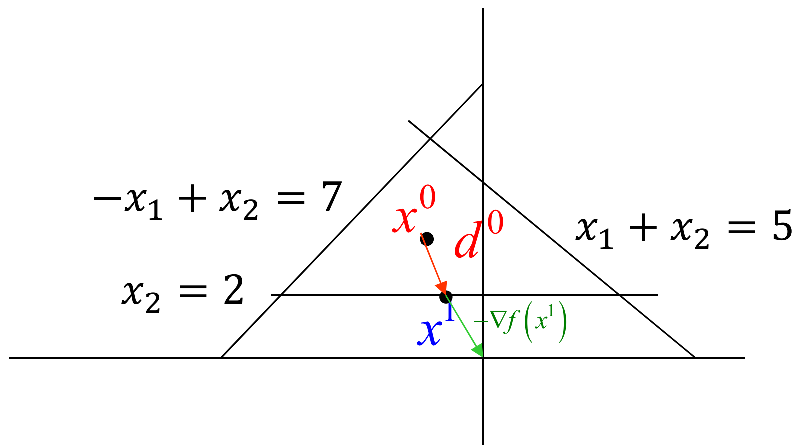
La plus grande valeur $\bar{\alpha}_0$ que peut prendre α_0 pour que $x^0 + \alpha_0 d^0 \in D_R$: lorsque la troisième contrainte devient active $-3 + 3\alpha_0 \leq -2 \Longrightarrow \bar{\alpha}_0 = \frac{1}{3}$, On a

$$\varphi(\alpha_0) = f(-2+2\alpha_0, 3-3\alpha_0) = \frac{1}{2}\left((-2+2\alpha_0)^2 + (3-3\alpha_0)^2\right) = \frac{1}{2}(13 - 26\alpha_0 + 13\alpha_0^2)$$

Ce qui implique $\varphi'(\alpha_0) = 13\alpha_0 - 13 = 0 \Longrightarrow \bar{\alpha}_0 = 1$

Et alors : $\min_{0\leq\alpha_0\leq\frac{1}{3}} \varphi(\alpha_0) = \varphi\left(\frac{1}{3}\right)$, Donc $\alpha_0^* = \frac{1}{3}$ $\quad x^1 = \left[-\frac{4}{3}, 2\right]^T$

Méthode du gradient projeté: Exemple



$$(P)\begin{cases} \text{Min} & f(x) = \frac{1}{2}(x_1^2 + x_2^2) \\ \text{sujet à} & -x_1 + x_2 \le 7 \\ & x_1 + x_2 \le 5 \\ & -x_2 \le -2 \end{cases}$$

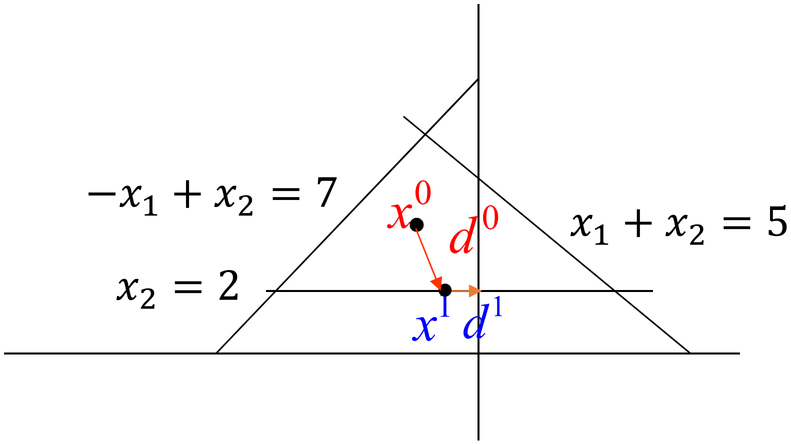
$$\nabla f(x) = [x_1, x_2]^T$$

Itération 2 : $x^1 = \left[-\frac{4}{3}, 2\right]^T$. Alors $I_1 = \{3\}$.La direction d^1 est obtenue en faisant la projection de $\nabla f(x^1) = -\left[-\frac{4}{3}, 2\right]$ sur $\left\{x \in R^n : {a^3}^T x = -x_2 = 0\right\}$

$$d^1 = -\left(I - \binom{0}{-1}\left((0,-1)\binom{0}{-1}\right)^{-1}(0,-1)\right)\begin{pmatrix} -\frac{4}{3} \\ 2 \end{pmatrix} = -\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\right)\begin{pmatrix} -\frac{4}{3} \\ 2 \end{pmatrix}$$

$$= \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} -\frac{4}{3} \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{4}{3} \\ 0 \end{pmatrix}$$

Méthode du gradient projeté: Exemple



$$(P)\begin{cases} \text{Min} \quad f(x) = \frac{1}{2}\left(x_1^2 + x_2^2\right) \\ \text{sujet à} \quad -x_1 + x_2 \leq 7 \\ \qquad\qquad x_1 + x_2 \leq 5 \\ \qquad\qquad -x_2 \leq -2 \end{cases}$$

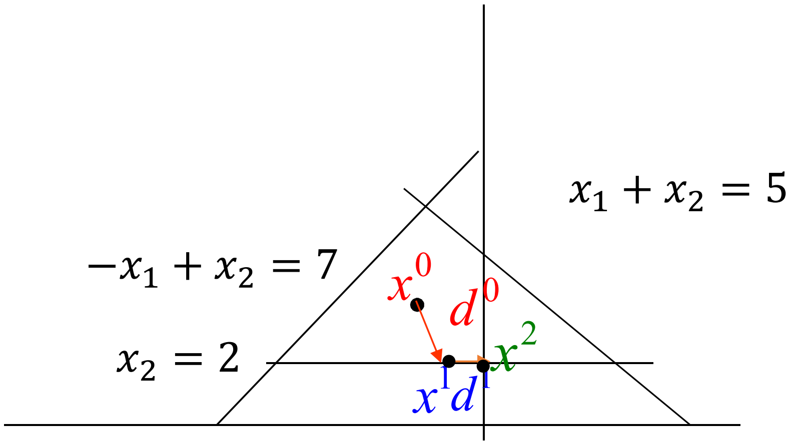
$$\nabla f(x) = [x_1, x_2]^T$$

Itération 2 : $x^1 = \left[-\frac{4}{3}, 2\right]^T . d^1 = \left[\frac{4}{3}, 0\right]^T$

- $x^1 + \alpha_1 d^1 = \left(-\frac{4}{3}, 2\right)^T + \alpha_1\left(\frac{4}{3}, 0\right)^T = \left(-\frac{4}{3} + \frac{4}{3}\alpha_1, 2\right)^T$

- $$\left.\begin{array}{l} {a^1}^T d^1 = -1 \times \frac{4}{3} + 0 = -\frac{4}{3} < 0 \\ {a^2}^T d^1 = 1 \times \frac{4}{3} + 0 = \frac{4}{3} > 0 \end{array}\right\} \Longrightarrow \quad \bar{\alpha}_1 = \frac{b_2 - {a^2}^T x^1}{{a^2}^T d^1} = \frac{5 - \left(-\frac{4}{3} + 2\right)}{\frac{4}{3}} = \frac{13}{4}$$

Méthode du gradient projeté: Exemple



$$(P)\begin{cases} \text{Min} \quad f(x) = \frac{1}{2}\,(x_1^2 + x_2^2) \\ \text{sujet à} \quad -x_1 + x_2 \le \ \ 7 \\ \qquad\qquad x_1 + x_2 \le \ \ 5 \\ \qquad\qquad\quad -x_2 \le -2 \end{cases}$$

$$\nabla f(x) = [x_1, x_2]^T$$

Itération 2 : $x^1 + \alpha_1 d^1 = \left(-\frac{4}{3}, 2\right)^T + \alpha_1\left(\frac{4}{3}, 0\right)^T = \left(-\frac{4}{3} + \frac{4}{3}\alpha_1, 2\right)^T$

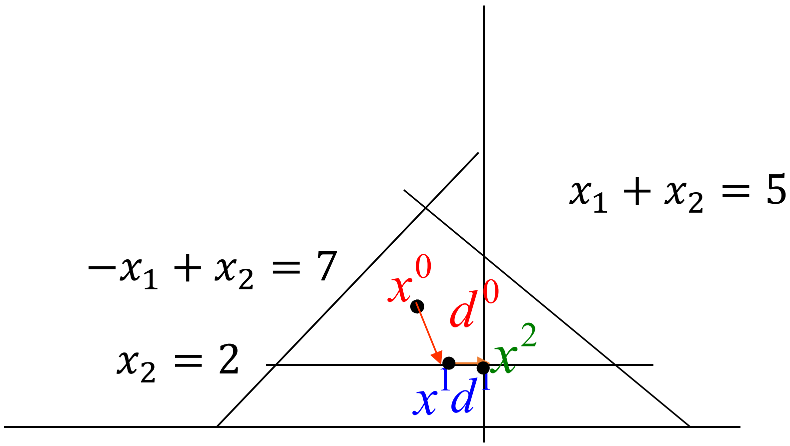
La plus grande valeur $\bar{\alpha}_1$ que peut prendre α_1 pour que $x^1 + \alpha_1 d^1 \in D_R$ lorsque la deuxième contrainte devient active $\quad -\frac{4}{3} + \frac{4}{3}\alpha_1 + 2 \le 5 \Rightarrow \bar{\alpha}_1 = \frac{13}{4}$

On a $\varphi(\alpha_1) = f\left(-\frac{4}{3} + \frac{4}{3}\alpha_1, 2\right) = \frac{1}{2}\left(\left(-\frac{4}{3} + \frac{4}{3}\alpha_1\right)^2 + 4\right) = \frac{8}{9}(-1 + \alpha_1)^2 + 2$

Ce qui implique $\varphi'(\alpha_1) = \frac{16}{9}\alpha_1 - \frac{16}{9} = 0 \implies \bar{\alpha}_1 = 1 \implies \min\limits_{0 \le \alpha_1 \le \frac{13}{4}} \varphi(\alpha_1) = \varphi(1)$

Donc le pas optimal $\alpha_1^* = 1 \implies x^2 = [0,2]^T$

Méthode du gradient projeté: Exemple



$$(P)\begin{cases} \text{Min} \quad f(x) = \frac{1}{2}\,(x_1^2 + x_2^2) \\ \text{sujet à} \quad -x_1 + x_2 \leq 7 \\ \qquad\qquad x_1 + x_2 \leq 5 \\ \qquad\qquad -x_2 \leq -2 \end{cases}$$

$$\nabla f(x) = [x_1, x_2]^T$$

Itération 3 : $x^2 = [0,2]^T$. Alors $I_1 = \{3\}$, La direction d^2 est obtenue en faisant la projection de $-\nabla f(x^2) = -[0,2]$ sur $\left\{x \in \mathbb{R}^n : {a^3}^T x = -x_2 = 0\right\}$.

$$d^2 = -\left(I - \begin{pmatrix} 0 \\ -1 \end{pmatrix}\left((0,-1)\begin{pmatrix} 0 \\ -1 \end{pmatrix}\right)^{-1}(0,-1)\right)\begin{pmatrix} 0 \\ 2 \end{pmatrix} = -\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\right)\begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$v^2 = -\left([0,-1]\begin{bmatrix} 0 \\ -1 \end{bmatrix}\right)^{-1}[0,-1]\begin{bmatrix} 0 \\ 2 \end{bmatrix} = [0,1]\begin{bmatrix} 0 \\ 2 \end{bmatrix} = 2$$

Le point $[0,2]^T$ satisfait les conditions K-K-T pour le problème puisque : $d^2 = [0,0]^T$ $v^1 = 2 > 0$

Méthode du gradient projeté

Remarques :

i. Dans son livre, Luenberger mentionne que même s'il n'existe pas de preuve de convergence pour cette méthode, elle est utilisée avec succès et la suite des points générés converge toujours vers un point satisfaisant les conditions K-K-T pour le problème.

ii. La méthode peut être généralisée au cas où les contraintes sont non linéaires. Dans ce cas, la projection de $-\nabla f\left(x^k\right)$ s'effectue sur l'hyperplan tangent aux contraintes actives. Par contre, le pas doit être ajusté pour que x^{k+1} appartienne au domaine réalisable.