Chapitre 5 : Optimisation non linéaire, méthodes de gradient

ENSIIE - Module de Recherche Opérationnelle

Dimitri Watel (dimitri.watel@ensiie.fr)

2024

Problèmes MIDA

Minimum Intersecting Disks Area

Soient $p_1, p_2, \ldots, p_n, q$ des points du plan, tracer, pour chaque point p_i un disque D_i centré en p_i tel que

$$\forall i < n, D_i \cap D_{i+1} \neq \emptyset$$

$$q \in D_n$$

et tel que la somme de l'aire des disques soit minimum.

Modélisation : MIDA

Soit d_i la distance entre p_i et p_{i+1} ; d_n la distance entre p_n et q.

$$
\begin{array}{lll}
\min & \displaystyle\sum_{i=1}^{n} r_i^2 & \\
s.c. & r_i + r_{i+1} \geq d_i & \forall\, 1 \leq i < n \\
 & r_n \geq d_n & \\
 & \cancel{(r_i \geq 0)} & \forall\, 1 \leq i \leq n
\end{array}
$$

Si r est solution avec $r_i < 0$ alors remplacer r_i par $-r_i$ est aussi solution, avec la même valeur objectif.

Conditions de Karush-Kuhn-Tucker

Soit $f, g_i, h_j : \mathbb{R}^n \to \mathbb{R}, \forall\ i \in [\![1; m]\!], \forall\ j \in [\![1; p]\!]$ de classe C^1.
On veut résoudre

$$\begin{array}{lcr} \min & f(x) & \\ s.c. & g_i(x) \leq 0 & \forall\ i \in [\![1; m]\!] \\ & h_j(x) = 0 & \forall\ j \in [\![1; p]\!] \end{array}$$

On note S l'ensemble des solutions réalisables.

Conditions de Karush-Kuhn-Tucker

$x \in S$ vérifie les conditions de Karush-Kuhn-Tucker (KKT) si :

$$\exists \lambda_i \geq 0, \mu_j \in \mathbb{R} \qquad \forall\ i \in [\![1; m]\!], j \in [\![1; p]\!]$$

$$\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) + \sum_{j=1}^{p} \mu_j \nabla h_j(x) = 0$$

$$\lambda_i \cdot g_i(x) = 0 \quad \forall\ i \in [\![1; m]\!]$$

Rappels : qualification

Soit $x \in S$. On note les contraintes saturées

$$I(x) = \{i \in [\![1; m]\!] | g_i(x) = 0\}$$

Qualification de l'indépendance linéaire

x est qualifié sous l'indépendance linéaire si la famille de vecteurs $(\nabla g_i(x), i \in I(x) \cup (\nabla h_j(x), j \in [\![1; p]\!])$ est linéairement indépendante.

Remarque : il existe une définition plus générale de la qualification, mais on ne s'en servira pas ici.

Rappels : Conditions nécessaires et suffisantes d'optimalité

Conditions nécessaires (KKT)

Si $x^* \in S$ est un minimum local de f et si x^* est qualifié alors x^* vérifie (KKT).

Conditions suffisante (KKT)

Si x^* satisfait (KKT), que les fonctions f et g_i sont convexes et que les fonctions h_j sont linéaires, alors x^* est un minimum global.

Qualification dans MIDA

MIDA : Tous les points sont qualifiés sous l'indépendance linéaire.

$\Rightarrow$ Les optimum globaux vérifient tous (KKT).

Algorithme du gradient projeté

On veut résoudre

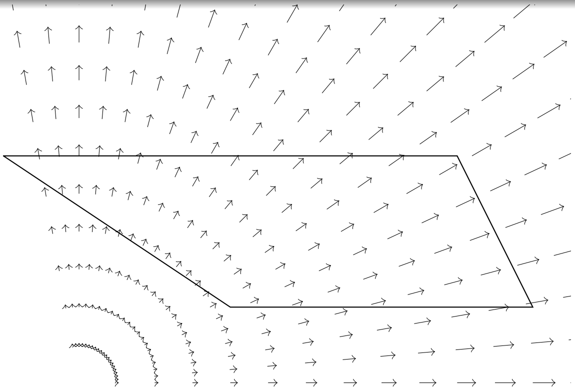
$$\begin{array}{lll} \min & f(x) & \\ s.c. & g_i(x) \leq 0 & \forall\ 1 \leq i \leq m \\ & h_j(x) = 0 & \forall\ 1 \leq j \leq p \end{array}$$

avec g_i et h_j **linéaires** et tous les points sont qualifiés sous l'indépendance linéaire.

Présentation succinte de l'algorithme

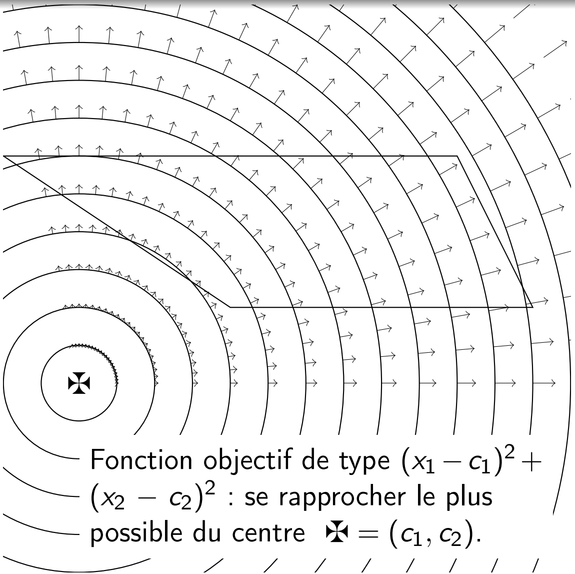
L'algorithme du gradient projeté est une descente de gradient qui ne sort pas de S. Pour cela, si on touche un bord de S, on ne suit pas le gradient mais le gradient projeté sur ce bord.
Il s'arrête sur un point vérifiant les conditions (KKT).

Gradient projeté : idée générale



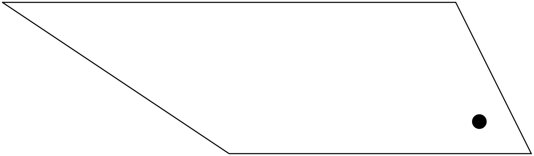
Les flèches représentent le gradient.
Le polygone représente S.

Gradient projeté : idée générale



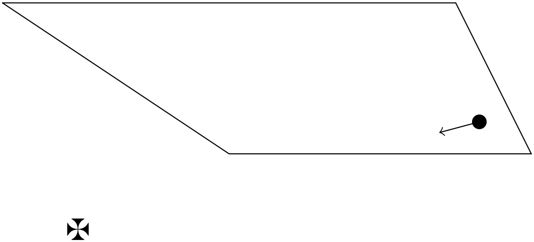
Fonction objectif de type $(x_1 - c_1)^2 + (x_2 - c_2)^2$: se rapprocher le plus possible du centre ✠ $= (c_1, c_2)$.

Gradient projeté : idée générale



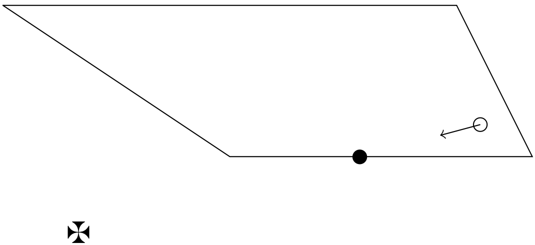
Etape 1 : trouver une solution réalisable x.

Gradient projeté : idée générale



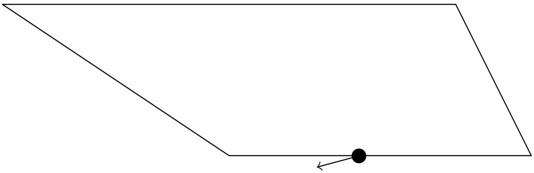
Etape 2 : regarder l'opposé du gradient $-\nabla f(x)$.

Gradient projeté : idée générale



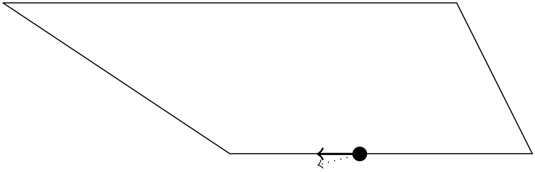
Etape 3 : On peut suivre cette direction, aller vers le minimum, sans sortir de S, et on recommence.

Gradient projeté : idée générale



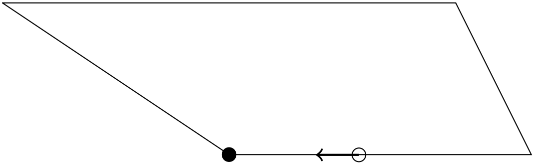
Etape 2 : regarder l'opposé du gradient $-\nabla f(x)$.

Gradient projeté : idée générale



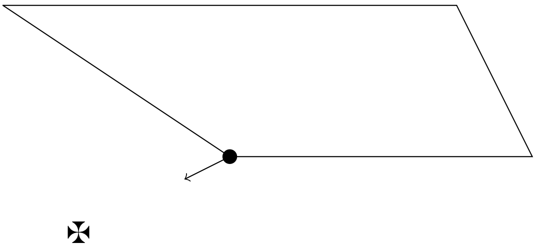
Etape 3 : On ne peut pas suivre cette direction sans sortir de S, on projette $-\nabla f(x)$ sur le bord de S.

Gradient projeté : idée générale



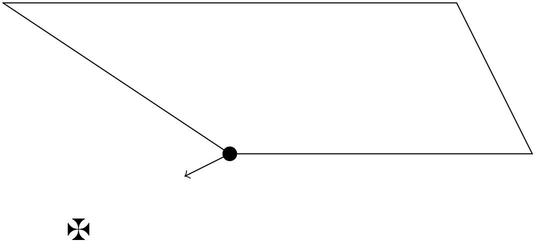
Etape 4 : On va vers le minimum dans cette direction, sans sortir de S et on recommence.

Gradient projeté : idée générale



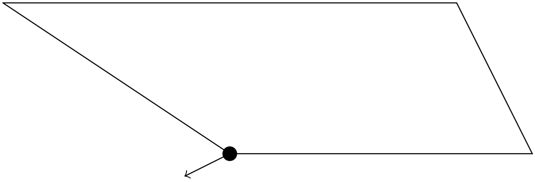
Etape 2 : regarder l'opposé du gradient $-\nabla f(x)$.

Gradient projeté : idée générale



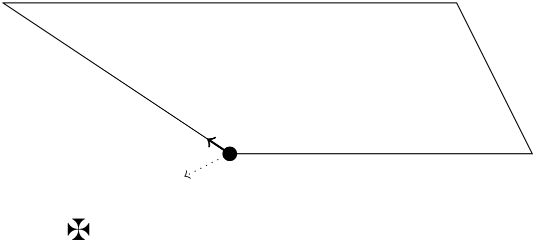
Etape 3 : On ne peut pas suivre cette direction sans sortir de S, on projette $-\nabla f(x)$ sur le bord de S.

Gradient projeté : idée générale



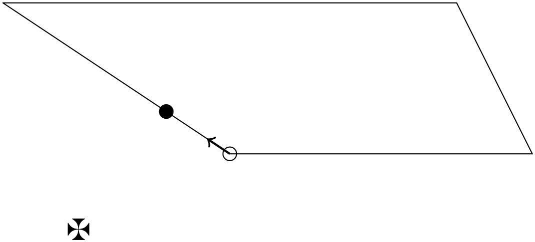
Etape 4 : Projection sur 2 bords non colinéaires en même temps ⇒ quel bord choisir ?

Gradient projeté : idée générale



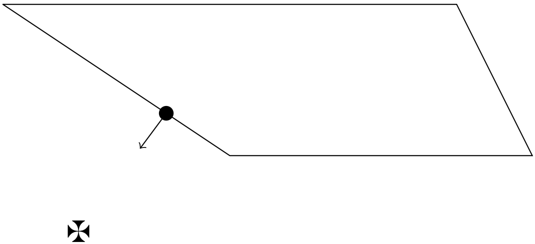
Etape 4 : Projeter sur le bord pour lequel la direction diminue l'objectif.

Gradient projeté : idée générale



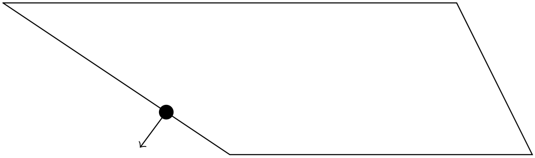
Etape 5 : On va vers le minimum dans cette direction, sans sortir de S et on recommence.

Gradient projeté : idée générale



Etape 2 : regarder l'opposé du gradient $-\nabla f(x)$.

Gradient projeté : idée générale



Etape 3 : Le gradient est orthogonal au bord, on s'arrête.

Dessin → formules

On rappelle que g_i et h_j sont linéaires :

$$g_i(x) = \sum_{k=1}^{n}(a_{ik}x_k) - b_i \leq 0 \ \forall i \in [\![1; m]\!]$$

$$h_j(x) = \sum_{k=1}^{n}(a'_{jk}x_k) - b'_j = 0 \ \forall j \in [\![1; p]\!]$$

On note

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \\ a'_{11} & a'_{12} & \cdots & a'_{1n} \\ a'_{21} & a'_{22} & \cdots & a'_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a'_{p1} & a'_{p2} & \cdots & a'_{pn} \end{pmatrix} \quad B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \\ b'_1 \\ b'_2 \\ \vdots \\ b'_p \end{pmatrix}$$

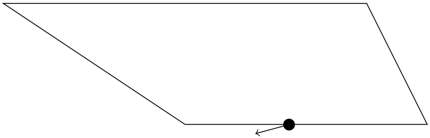
Dessin $\rightarrow$ formules

Lemme

Si L_i est la i^{e} ligne de A, alors

$$L_i = {}^t\nabla(g_i(x)) \qquad \forall i \in [\![1; m]\!]$$
$$L_{j+m} = {}^t\nabla(h_j(x)) \qquad \forall j \in [\![1; p]\!]$$

Détecter si on est sur la frontière

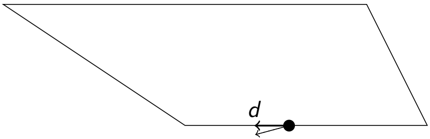


x touche les hyperplans d'équations $h_j(x) = 0$ et $g_i(x) = 0$.
On note

$$I(x) = \{i | g_i(x) = 0\}$$

$$J = \{j | h_j(x) = 0\} = [\![1; p]\!]$$

Projeter le gradient sur la frontière



On note $A_s = \{L_i, i \in I(x) \cup J\}$.

Lemme

Le gradient projeté d vaut

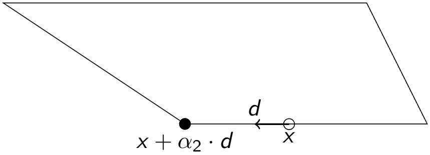
$$\begin{aligned} d &= \text{Projection de } (-\nabla f(x)) \text{ sur } \{y | L_i \cdot y = 0 \ \forall i \in I(x) \cup J\} \\ &= \left(I_n - {}^t A_s \cdot (A_s \cdot {}^t A_s)^{-1} \cdot A_s\right) \cdot (-\nabla f(x)) \end{aligned}$$

Remarques

$A_s \cdot {}^tA_s$ est toujours inversible du fait de la qualification par l'indépendance linéaire.

Si $I(x) = J = \emptyset$ alors, $A_s = \emptyset$. Dans ce cas, $d = (-\nabla f(x))$.

Déplacement dans la direction d



On note

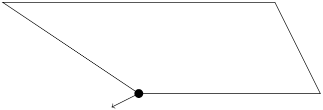
$$\alpha_1 = \max_{0 \leq \alpha}(x + \alpha \cdot d \in S)$$

$$\alpha_2 = \arg \min_{0 \leq \alpha \leq \alpha_1} (f(x + \alpha \cdot d))$$

Lemme

Si $d \neq \vec{0}$ alors $\alpha_1, \alpha_2 \neq 0$.

Cas où $d = \vec{0}$



Lemme

Si $d = \vec{0}$, alors il existe λ_i, pour $i \in I(x)$ et μ_j pour $j \in J$ tel que $-\nabla f(x) = \sum_{i \in I(x)} \lambda_i \cdot L_i + \sum_{j \in J} \mu_j \cdot L_{m+j}$.

Lemme

$$\begin{pmatrix} \lambda \\ \mu \end{pmatrix} = (A_s \cdot {}^t A_s)^{-1} A_s \cdot (-\nabla f(x))$$

Cas où $d = \vec{0}$

Lemme

Si, pour tout $i \in I(x)$, $\lambda_i \geq 0$ alors les conditions (KKT) sont vérifiées.

Lemme

S'il existe $\lambda_i < 0$, alors on peut supprimer i de $I(x)$. On recalcule la projection d' comme au slide 14.

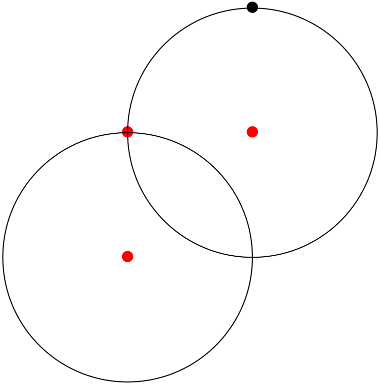
- $d' \neq \vec{0}$
- $\nabla g_i(x) \cdot d' < 0$

L'algorithme du gradient projeté

Trouver $x \in S$
$I(x) \leftarrow \{i \in [\![1; m]\!] | g_i(x) = 0\}$
Boucle
    $A_s \leftarrow \{L_i, i \in I(x) \cup J\}$
    $d \leftarrow (I_n - {}^tA_s \cdot (A_s \cdot {}^tA_s)^{-1} \cdot A_s) \cdot (-\nabla f(x))$
    Si $d = 0$ **Alors**
        ${}^t(\lambda, \mu) \leftarrow (A_s \cdot {}^tA_s)^{-1} A_s \cdot (-\nabla f(x))$
        Si $\exists i \in I(x), \lambda_i < 0$, Retirer i de $I(x)$
        Sinon Renvoyer x
    Sinon
        $\alpha_1 \leftarrow \max\limits_{0 \leq \alpha}(x + \alpha \cdot d \in S)$
        $\alpha_2 \leftarrow \arg \min\limits_{0 \leq \alpha \leq \alpha_1} (f(x + \alpha \cdot d))$
        $x \leftarrow x + \alpha_2 \cdot d$
        $I(x) \leftarrow \{i \in [\![1; m]\!] | g_i(x) = 0\}$

Exemple

Soit l'instance de MIDA suivante :
$P = \{p_1, p_2, p_3\} = \{[0,0], [0,5], [5,5]\}$ et $q = [5,10]$.
On veut partir de la solution $r = (r_1, r_2, r_3) = (5, 0, 5)$.



Exemple

Le problème s'écrit :

$$
\begin{array}{llr}
\min & r_1^2 + r_2^2 + r_3^2 & \\
s.c. & -r_1 - r_2 \leq -5 & (g_1) \\
 & -r_2 - r_3 \leq -5 & (g_2) \\
 & -r_3 \leq -5 & (g_3)
\end{array}
$$

$$
A = \begin{pmatrix} -1 & -1 & 0 \\ 0 & -1 & -1 \\ 0 & 0 & -1 \end{pmatrix} \; B = \begin{pmatrix} -5 \\ -5 \\ -5 \end{pmatrix}
$$

Exemple : Itération 1

Si $r = (5, 0, 5)$ alors $I(r) = \{1, 2, 3\}$.

$$A_s = \begin{array}{cc} \begin{pmatrix} -1 & -1 & 0 \\ 0 & -1 & -1 \\ 0 & 0 & -1 \end{pmatrix} & \begin{array}{c} L_1 \\ L_2 \\ L_3 \end{array} \end{array}$$

$$A_s \cdot {}^t A_s = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

Exemple : Itération 1

$$^t\nabla f(r) = (2r_1, 2r_2, 2r_3) = (10, 0, 10)$$

$$d = (I_n - {}^tA_s \cdot (A_s \cdot {}^tA_s)^{-1} \cdot A_s) \cdot (-\nabla f(r)) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot (-\nabla f(r)) = 0$$

Autre méthode : on calcule la projection de $-\nabla f(r)$ sur

$$\begin{aligned} L &= \{y | L_1 \cdot y = 0; L_2 \cdot y = 0; L_3 \cdot y = 0\} \\ &= \{y | -y_1 - y_2 = 0; -y_2 - y_3 = 0; -y_3 = 0\} \\ &= \{y = 0\} \end{aligned}$$

La projection vaut donc nécessairement $d = 0$.

Exemple : Itération 1

$$A_s = \begin{array}{c c} \begin{pmatrix} -1 & -1 & 0 \\ 0 & -1 & -1 \\ 0 & 0 & -1 \end{pmatrix} & \begin{array}{c} L_1 \\ L_2 \\ L_3 \end{array} \end{array}$$

$$\begin{pmatrix} \lambda \\ \mu \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = (A_s \cdot {}^t A_s)^{-1} \cdot A_s \cdot (-\nabla f(r)) = \begin{pmatrix} 10 \\ -10 \\ 20 \end{pmatrix}$$

Autre méthode :

$$-\nabla f(r) = \begin{pmatrix} -10 \\ 0 \\ -10 \end{pmatrix} = \lambda_1 L_1 + \lambda_2 L_2 + \lambda_3 L_3 = 10 \begin{pmatrix} -1 \\ -1 \\ 0 \end{pmatrix} - 10 \begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix} + 20 \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}$$

$$\lambda_2 < 0 \Rightarrow I(r) \leftarrow I(r) \backslash \{2\} = \{1, 3\}$$

Exemple : Itération 1.2

$I(r) = \{1, 3\}$.

$$A_s = \begin{array}{c} \begin{pmatrix} -1 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \end{array} \begin{array}{l} L_1 \\ L_3 \end{array}$$

$$A_s \cdot {}^t A_s = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

Exemple : Itération 1.2

$$^t\nabla f(r) = (2r_1, 2r_2, 2r_3) = (10, 0, 10)$$

$$d = (I_n - {}^tA_s \cdot (A_s \cdot {}^tA_s)^{-1} \cdot A_s) \cdot (-\nabla f(r)) = \begin{pmatrix} 0.5 & -0.5 & 0 \\ -0.5 & 0.5 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} -10 \\ 0 \\ -10 \end{pmatrix} = \begin{pmatrix} -5 \\ 5 \\ 0 \end{pmatrix}$$

Autre méthode : on calcule la projection de $-\nabla f(r)$ sur

$$\begin{aligned} L &= \{y | L_1 \cdot y = 0; L_3 \cdot y = 0\} \\ &= \{y | -y_1 - y_2 = 0; -y_3 = 0\} \\ &= \{y | y_1 = -y_2; y_3 = 0\} \end{aligned}$$

La projection est donc de la forme $d = \beta \cdot (1, -1, 0)$ sachant que

$$-\nabla f(r) \cdot d = d^2 \Rightarrow -10\beta = 2 \cdot \beta^2 \Rightarrow \beta = -5$$

Exemple : Itération 1, fin

$$\begin{aligned}
\alpha_1 &= \max_{0 \leq \alpha}(r + \alpha \cdot d \in S) \\
&= \max_{0 \leq \alpha}\left(\begin{pmatrix} 5 - 5 \cdot \alpha \\ 0 + 5 \cdot \alpha \\ 5 + 0 \cdot \alpha \end{pmatrix} \in S \right) \\
&= \max_{0 \leq \alpha} \begin{pmatrix} 5 - 5 \cdot \alpha + 0 + 5 \cdot \alpha \geq 5 \\ 0 + 5 \cdot \alpha + 5 + 0 \cdot \alpha \geq 5 \\ 5 + 0 \cdot \alpha \geq 5 \end{pmatrix} \\
&= \max_{0 \leq \alpha} \begin{pmatrix} 5 \geq 5 \\ 5 + 5 \cdot \alpha \geq 5 \\ 5 \geq 5 \end{pmatrix} \\
&= +\infty
\end{aligned}$$

Exemple : Itération 1, fin

$$\begin{aligned}
\alpha_2 &= \arg \min_{0 \leq \alpha \leq \alpha_1} (f(r + \alpha \cdot d)) \\
&= \arg \min_{0 \leq \alpha} (f(\begin{pmatrix} 5 - 5 \cdot \alpha \\ 0 + 5 \cdot \alpha \\ 5 + 0 \cdot \alpha \end{pmatrix})) \\
&= \arg \min_{0 \leq \alpha} ((5 - 5 \cdot \alpha)^2 + (0 + 5 \cdot \alpha)^2 + (5 + 0 \cdot \alpha)^2) \\
&= \arg \min_{0 \leq \alpha} (50 - 25\alpha + 50\alpha^2) \\
&= \frac{1}{2}
\end{aligned}$$

On se déplace donc vers $r + \frac{1}{2} d = \begin{pmatrix} 2.5 \\ 2.5 \\ 5 \end{pmatrix}$.

Exemple : Itération 2

Si $r = (2.5, 2.5, 5)$ alors $I(r) = \{1, 3\}$.

$$A_s = \begin{array}{cc} \begin{pmatrix} -1 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} & \begin{array}{c} L_1 \\ L_3 \end{array} \end{array}$$

$$A_s \cdot {}^tA_s = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

Exemple : Itération 2

$$^t\nabla f(r) = (2r_1, 2r_2, 2r_3) = (5, 5, 10)$$

$$d = (I_n - {}^tA_s \cdot (A_s \cdot {}^tA_s)^{-1} \cdot A_s) \cdot (-\nabla f(r)) = \begin{pmatrix} 0.5 & -0.5 & 0 \\ -0.5 & 0.5 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} -5 \\ -5 \\ -10 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Autre méthode : on calcule la projection de $-\nabla f(r)$ sur

$$\begin{aligned} L &= \{y | L_1 \cdot y = 0; L_3 \cdot y = 0\} \\ &= \{y | -y_1 - y_2 = 0; -y_3 = 0\} \\ &= \{y | y_1 = -y_2; y_3 = 0\} \end{aligned}$$

La projection est donc de la forme $d = \beta \cdot (1, -1, 0)$ sachant que

$$-\nabla f(r) \cdot d = d^2 \Rightarrow 0 = 2 \cdot \beta^2 \Rightarrow \beta = 0$$

Exemple : Itération 2

$$A_s = \begin{array}{c}\begin{pmatrix} -1 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}\end{array}\begin{array}{l} L_1 \\ L_2 \end{array}$$

$$\begin{pmatrix} \lambda \\ \mu \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \lambda_3 \end{pmatrix} = (A_s \cdot {}^t A_s)^{-1} \cdot A_s \cdot (-\nabla f(r)) = \begin{pmatrix} 5 \\ 10 \end{pmatrix}$$

Autre méthode :

$$-\nabla f(r) = \begin{pmatrix} -5 \\ -5 \\ -10 \end{pmatrix} = \lambda_1 L_1 + \lambda_3 L_3 = 5 \begin{pmatrix} -1 \\ -1 \\ 0 \end{pmatrix} + 10 \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}$$

$\lambda_1, \lambda_3 \geq 0 \Rightarrow$ (KKT) vérifiées. On renvoie $r^* = (2.5, 2.5, 5)$ avec $f(r^*) = 31.25$.

De plus, f est convexe, les g_i sont linéaires sont convexes, la condition suffisante de (KKT) montre donc que r^* est un **minimum global**.

Algorithme du gradient réduit

On veut résoudre

$$\begin{array}{lrl} \min & f(x) & \\ s.c. & h_j(x) = 0 & \forall\ 1 \leq j \leq p \\ & x_i \geq 0 & \forall\ 1 \leq i \leq n \end{array}$$

avec h_j **linéaires** et $p < n$.

Présentation succinte de l'algorithme

L'algorithme du gradient réduit est une descente de gradient qui ne sort pas de S. Du fait des égalités, une partie des variables, notée x_B, peut s'exprimer en fonction des autres, notées x_N. On va donc *réduire* $f(x)$ et $\nabla f(x)$ à $f(x_N)$ et $\nabla f(x_N)$.
Il s'arrête sur un point vérifiant les conditions (KKT).

Rappel

Programme standard

Soit le programme (P) suivant

$$\begin{array}{lrr} \min & f(x) & \\ s.c. & g_i(x) \leq 0 & \forall\ 1 \leq i \leq m \\ & h_j(x) = 0 & \forall\ 1 \leq j \leq p \end{array}$$

alors il existe un programme (P') équivalent à (P) de la forme

$$\begin{array}{lrr} \min & f(x) & \\ s.c. & h_j(x) = 0 & \forall\ 1 \leq j \leq p \\ & x_i \geq 0 & \forall\ 1 \leq i \leq n \end{array}$$

Rappel : astuces

Si $g_i(x) \leq 0$ alors rajouter une variable d'écart $s_i \geq 0$ et poser $h_i(x) = g_i(x) + s_i = 0$.

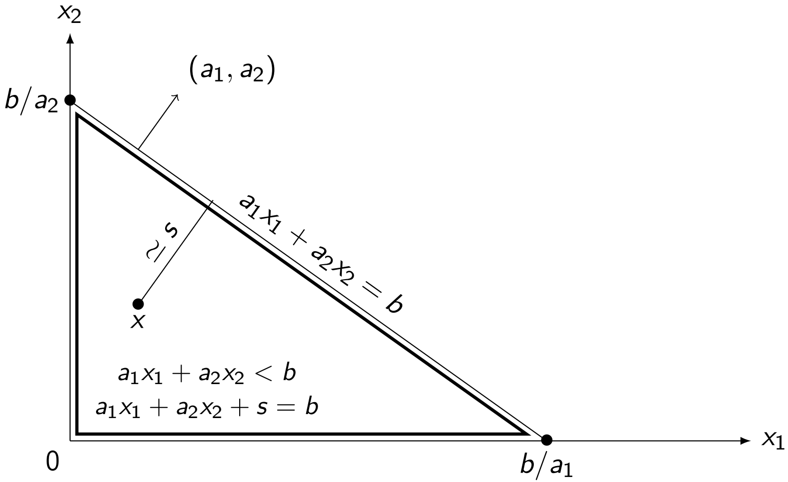
Si $x_i \in \mathbb{R}$ alors remplacer x_i par 2 variables $x_i^+ \geq 0$ et $x_i^- \geq 0$ en posant $x_i = x_i^+ - x_i^-$.

Si $g_i(x) = -x_i \leq 0$ alors **ne rien faire**.

Si $g_i(x) = x_i \leq 0$ alors remplacer x_i par $-x_i$.

Rappel : variables d'écart et dessin

On suppose $b > 0$



Pas de dessin

On rappelle que h_j sont linéaires :

$$h_j(x) = \sum_{k=1}^{n}(a_{jk}x_k) - b_j = 0 \ \forall j \in [\![1; p]\!]$$

On note

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pn} \end{pmatrix} \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix}$$

On suppose que $rg(A) = p$.

Base et non base

Soit $B \subset [\![1; n]\!]$ de taille p et $N = [\![1; n]\!] \backslash B$.
On note
$A_B =$ les colonnes de A dont l'indice est dans B.
$A_N =$ les colonnes de A dont l'indice est dans N.
$x_B =$ les variables de x dont l'indice est dans B.
$x_N =$ les variables de x dont l'indice est dans N.
$\nabla f_B =$ les coefficient de $\nabla f(x)$ dont l'indice est dans B.
$\nabla f_N =$ les coefficient de $\nabla f(x)$ dont l'indice est dans N.

On peut réécrire A, x et $\nabla f(x)$ ainsi :

$$A = \begin{pmatrix} A_B & A_N \end{pmatrix}, x = \begin{pmatrix} x_B \\ x_N \end{pmatrix}, \nabla f(x) = \begin{pmatrix} \nabla f_B \\ \nabla f_N \end{pmatrix}$$

Réduction, gradient réduit

Lemme

Si A_B est inversible, on peut réécrire $f(x)$ en fonction de x_N uniquement avec

$$x_B = A_B^{-1} \cdot b - A_B^{-1} \cdot A_N \cdot x_N$$

Soit $\bar{f} : \mathbb{R}^{n-p} \to \mathbb{R}$ tel que $\bar{f}(x_N) = f(x_B, x_N)$.

Lemme

Le *gradient réduit* est le gradient de $\bar{f}$.

$${}^t\nabla \bar{f}(x_N) = -\,{}^t\nabla f_B \cdot A_B^{-1} \cdot A_N + {}^t\nabla f_N$$

Direction à suivre

Définition

Connaissant $\nabla \bar{f}(x_N)$, on définit la direction d, coupée en d_B et d_N, ainsi

$$\forall j \in N, d_j = \begin{cases} 0 & \text{si } \nabla \bar{f}(x_N)_j > 0 \text{ et } x_j = 0 \\ -\nabla \bar{f}(x_N)_j & \text{sinon} \end{cases}$$

$$d_B = -A_B^{-1} \cdot A_N \cdot d_N$$

Lemme

Pour tout $\alpha \in \mathbb{R}$, $A \cdot (x + \alpha \cdot d) = b$.

Cas où $d = \vec{0}$

Lemme

Si $d = \vec{0}$ alors les conditions (KKT) sont vérifiées.

Déplacement dans la direction d si $d \neq \vec{0}$

Comme pour le gradient projeté, on note

$$\alpha_1 = \max_{0 \leq \alpha}(\alpha | x + \alpha \cdot d \geq 0)$$
$$\alpha_2 = \arg \min_{0 \leq \alpha \leq \alpha_1} (f(x + \alpha \cdot d))$$

On se déplace vers

$$y = x + \alpha_2 \cdot d$$

Remarque

Il est possible que $\alpha_1 = 0$.

Changement de base

Lemme

S'il existe $s \in B$ tel que $y_s = 0$; et si on conserve la même base B, alors, il est possible qu'à l'itération suivante $\alpha_1 = 0 \rightarrow$ surplace.

Hypothèse de non dégénerescence

L'hypothèse énonce que, quel que soit la base B, si A_B est inversible alors $A_B^{-1} \cdot b > 0$.

Lemme

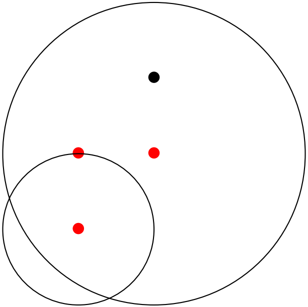
S'il existe $s \in B$ tel que $y_s = 0$, alors, sous hypothèse de non dégénérescence, il existe $r \in N$ tel que $y_r \neq 0$ et $A_{B \cup \{r\} \setminus \{s\}}$ est inversible.

L'algorithme du gradient réduit

Trouver $x \in S$ et B tel que A_B est inversible
Boucle
    ${}^t\nabla \bar{f}(x_N) \leftarrow -\,{}^t\nabla f_B \cdot A_B^{-1} \cdot A_N + {}^t\nabla f_N$
    Pour $j \in N$ **Faire**
        Si $\nabla \bar{f}(x_N)_j > 0$ et $x_j = 0$ **alors** $d_j \leftarrow 0$ **sinon** $d_j \leftarrow -\nabla \bar{f}(x_N)_j$
    $d_B \leftarrow -A_B^{-1} A_N d_N$
    Si $d = 0$ **alors Renvoyer** x
    $\alpha_1 \leftarrow \max_{0 \leq \alpha}(x + \alpha \cdot d \geq 0)$
    $\alpha_2 \leftarrow \arg\min_{0 \leq \alpha \leq \alpha_1} (f(x + \alpha \cdot d))$
    $x \leftarrow x + \alpha_2 \cdot d$
    Si $\exists s$ tel que $x_s = 0$ **Alors**
        Pour $r \in N$ par ordre décroissant de x_r **Faire**
            Si $A_{B \cup \{r\} \backslash \{s\}}$ est inversible **Alors**
                $B \leftarrow B \cup \{r\} \backslash \{s\}$
                Recommencer la boucle

Exemple

Soit l'instance de MIDA suivante :
$P = \{p_1, p_2, p_3\} = \{[0, 0], [0, 5], [5, 5]\}$ et $q = [5, 10]$.
On veut partir de la solution $r = (r_1, r_2, r_3) = (5, 0, 10)$.



Exemple

Le problème s'écrit :

$$\begin{array}{lcr}
\min & r_1^2 + r_2^2 + r_3^2 & \\
s.c. & -r_1 - r_2 + s_1 = -5 & (h_1) \\
 & -r_2 - r_3 + s_2 = -5 & (h_2) \\
 & -r_3 + s_3 = -5 & (h_3)
\end{array}$$

$$A = \begin{pmatrix} -1 & -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \end{pmatrix} \; B = \begin{pmatrix} -5 \\ -5 \\ -5 \end{pmatrix}$$

$$^t\nabla f(x) = (2r_1, 2r_2, 2r_3, 0, 0, 0)$$

$r = (5, 0, 10) \Rightarrow x = (5, 0, 10, 0, 5, 5)$
On part de la base $B = (1, 3, 6)$ et $N = (2, 4, 5)$.

Exemple : Itération 1

$$A_B = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 1 \end{pmatrix} \qquad A_N = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

$$x_B = (5, 10, 5) \qquad x_N = (0, 0, 5)$$

$$^t\nabla f_B = (10, 20, 0) \qquad ^t\nabla f_N = (0, 0, 0)$$

$$A_B^{-1} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$$

$$\begin{aligned} ^t\nabla \bar{f}(x_N) &= -\,^t\nabla f_B \cdot A_B^{-1} \cdot A_N + \,^t\nabla f_N \\ &= (-30, 10, 20) \end{aligned}$$

Exemple : Itération 1

Autre méthode : réécrire f en fonction de $x_N = (r_2, s_1, s_2) = (0, 0, 5)$

$$\begin{aligned} f(x) &= r_1^2 + r_2^2 + r_3^2 \\ \bar{f}(x_N) &= (5 + s_1 - r_2)^2 + r_2^2 + (5 + s_2 - r_2)^2 \\ &= 3r_2^2 + s_1^2 + s_2^2 - 2r_2 s_1 - 2r_2 s_2 - 20r_2 + 10s_1 + 10s_2 + 50 \end{aligned}$$

$$\nabla(\bar{f})(x_N) = \begin{pmatrix} 6r_2 - 2s_1 - 2s_2 - 20 \\ 2s_1 - 2r_2 + 10 \\ 2s_2 - 2r_2 + 10 \end{pmatrix} = \begin{array}{c} \begin{pmatrix} -30 \\ 10 \\ 20 \end{pmatrix} \end{array} \begin{array}{l} r_2 \\ s_1 \\ s_2 \end{array}$$

Exemple : Itération 1

$$d_N = \begin{pmatrix} 30 \\ 0 \\ -20 \end{pmatrix} \begin{array}{l} \leftarrow \nabla(\bar{f})(x_N)_{r_2} \leq 0, r_2 = 0 \\ \leftarrow \nabla(\bar{f})(x_N)_{s_1} > 0, s_1 = 0 \\ \leftarrow \nabla(\bar{f})(x_N)_{s_2} > 0, s_2 \neq 0 \end{array}$$

$$d_B \leftarrow -A_B^{-1} A_N d_N = \begin{pmatrix} -30 \\ -50 \\ -50 \end{pmatrix}$$

Exemple : Itération 1

$$
\begin{aligned}
\alpha_1 &= \max_{0 \leq \alpha}(x + \alpha \cdot d \geq 0) \\
&= \max_{0 \leq \alpha} \begin{pmatrix} 5 - 30 \cdot \alpha \geq 0 \\ 0 + 30 \cdot \alpha \geq 0 \\ 10 - 50 \cdot \alpha \geq 0 \\ 0 + 0 \cdot \alpha \geq 0 \\ 5 - 20 \cdot \alpha \geq 0 \\ 5 - 50 \cdot \alpha \geq 0 \end{pmatrix} \\
&= 0.1
\end{aligned}
$$

Exemple : Itération 1

$$\begin{aligned}
\alpha_2 &= \arg \min_{0 \leq \alpha \leq \alpha_1} (f(r + \alpha \cdot d)) \\
&= \arg \min_{0 \leq \alpha \leq 0.1} ((5 - 30 \cdot \alpha)^2 + (0 + 30 \cdot \alpha)^2 + (10 - 50 \cdot \alpha)^2) \\
&= \arg \min_{0 \leq \alpha \leq 0.1} (125 - 1300\alpha + 4300\alpha^2) \\
&= 0.1
\end{aligned}$$

On se déplace donc vers $x + 0.1d = \begin{pmatrix} 2 \\ 3 \\ 5 \\ 0 \\ 3 \\ 0 \end{pmatrix}$

Exemple : Itération 1, changement de base

$s_3 = 0$, or $s_3 \in B$ donc : changement de base.
$r_2 = 3, s_1 = 0, s_2 = 3$
Essayons $B' = B - \{s_3\} + \{r_2\} = (1, 2, 3)$

$$A_{B'} = \begin{pmatrix} -1 & -1 & 0 \\ 0 & -1 & -1 \\ 0 & 0 & -1 \end{pmatrix} \text{ inversible.}$$

Donc on remplace B par B'

Exemple : Itération 2

$$A_B = \begin{pmatrix} -1 & -1 & 0 \\ 0 & -1 & -1 \\ 0 & 0 & -1 \end{pmatrix} \qquad A_N = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$x_B = (2, 3, 5) \qquad x_N = (0, 3, 0)$$

$$^t\nabla f_B = (4, 6, 10) \qquad ^t\nabla f_N = (0, 0, 0)$$

$$A_B^{-1} = \begin{pmatrix} -1 & 1 & -1 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}$$

$$\begin{aligned} ^t\nabla \bar{f}(x_N) &= -\,^t\nabla f_B \cdot A_B^{-1} \cdot A_N + \,^t\nabla f_N \\ &= (4, 2, 8) \end{aligned}$$

Exemple : Itération 2

Autre méthode : réécrire f en fonction de $x_N = (s_1, s_2, s_3) = (0, 3, 0)$

$$
\begin{aligned}
f(x) &= r_1^2 + r_2^2 + r_3^2 \\
&= (5 + s_1 - r_2)^2 + (5 + s_2 - r_3)^2 + (5 + s_3)^2 \\
&= (5 + s_1 - (5 + s_2 - r_3))^2 + (5 + s_2 - r_3)^2 + (5 + s_3)^2 \\
&= (5 + s_1 - (5 + s_2 - (5 + s_3)))^2 + (5 + s_2 - (5 + s_3))^2 + (5 + s_3)^2 \\
\bar{f}(x_N) &= (5 + s_1 - s_2 + s_3)^2 + (s_2 - s_3)^2 + (5 + s_3)^2 \\
&= s_1^2 + 2s_2^2 + 3s_3^2 - 2s_1 s_2 + 2s_1 s_3 - 4s_2 s_3 + 10s_1 - 10s_2 + 20s_3 + 50
\end{aligned}
$$

$$
\nabla(\bar{f})(x_N) = \begin{pmatrix} 2s_1 + -2s_2 + 2s_3 + 10 \\ 4s_2 + -2s_1 - 4s_3 - 10 \\ 6s_3 + 2s_1 - 4s_2 + 20 \end{pmatrix} = \begin{array}{cc} \begin{pmatrix} 4 \\ 2 \\ 8 \end{pmatrix} & \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} \end{array}
$$

Exemple : Itération 2

$$d_N = \begin{pmatrix} 0 \\ -2 \\ 0 \end{pmatrix} \begin{array}{l} \leftarrow \nabla(\bar{f})(x_N)_{s_1} > 0, s_1 = 0 \\ \leftarrow \nabla(\bar{f})(x_N)_{s_2} > 0, s_2 \neq 0 \\ \leftarrow \nabla(\bar{f})(x_N)_{s_3} > 0, s_3 = 0 \end{array}$$

$$d_B \leftarrow -A_B^{-1} A_N d_N = \begin{pmatrix} 2 \\ -2 \\ 0 \end{pmatrix}$$

Exemple : Itération 2

$$\begin{aligned}
\alpha_1 &= \max_{0 \leq \alpha}(x + \alpha \cdot d \geq 0) \\
&= \max_{0 \leq \alpha} \begin{pmatrix} 2 + 2 \cdot \alpha \geq 0 \\ 3 - 2 \cdot \alpha \geq 0 \\ 5 + 0 \cdot \alpha \geq 0 \\ 0 + 0 \cdot \alpha \geq 0 \\ 3 - 2 \cdot \alpha \geq 0 \\ 0 + 0 \cdot \alpha \geq 0 \end{pmatrix} \\
&= 1.5
\end{aligned}$$

Exemple : Itération 2

$$\begin{aligned}
\alpha_2 &= \arg \min_{0 \leq \alpha \leq \alpha_1} (f(r + \alpha \cdot d)) \\
&= \arg \min_{0 \leq \alpha \leq 1.5} ((2 + 2 \cdot \alpha)^2 + (3 - 2 \cdot \alpha)^2 + (5 + 0 \cdot \alpha)^2) \\
&= \arg \min_{0 \leq \alpha \leq 1.5} (38 - 4\alpha + 8\alpha^2) \\
&= 0.25
\end{aligned}$$

On se déplace donc vers $x + 0.25d = \begin{pmatrix} 2.5 \\ 2.5 \\ 5 \\ 0 \\ 2.5 \\ 0 \end{pmatrix}$

Exemple : Itération 2, changement de base

Pas de $j \in B$ tel que $x_j = 0$, donc pas de changement de base.

Exemple : Itération 3

$$A_B = \begin{pmatrix} -1 & -1 & 0 \\ 0 & -1 & -1 \\ 0 & 0 & -1 \end{pmatrix}$$

$$x_B = (2.5, 2.5, 5)$$

$${}^t\nabla f_B = (5, 5, 10)$$

$$A_N = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$x_N = (0, 2.5, 0)$$

$${}^t\nabla f_N = (0, 0, 0)$$

$$A_B^{-1} = \begin{pmatrix} -1 & 1 & -1 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}$$

$$\begin{aligned} {}^t\nabla \bar{f}(x_N) &= -\,{}^t\nabla f_B \cdot A_B^{-1} \cdot A_N + {}^t\nabla f_N \\ &= (5, 0, 10) \end{aligned}$$

Exemple : Itération 3

Autre méthode : réécrire f en fonction de
$x_N = (s_1, s_2, s_3) = (0, 2.5, 0)$

$$\begin{aligned}
f(x) &= r_1^2 + r_2^2 + r_3^2 \\
&= (5 + s_1 - r_2)^2 + (5 + s_2 - r_3)^2 + (5 + s_3)^2 \\
&= (5 + s_1 - (5 + s_2 - r_3))^2 + (5 + s_2 - r_3)^2 + (5 + s_3)^2 \\
&= (5 + s_1 - (5 + s_2 - (5 + s_3)))^2 + (5 + s_2 - (5 + s_3))^2 + (5 + s_3)^2 \\
\bar{f}(x_N) &= (5 + s_1 - s_2 + s_3)^2 + (s_2 - s_3)^2 + (5 + s_3)^2 \\
&= s_1^2 + 2s_2^2 + 3s_3^2 - 2s_1s_2 + 2s_1s_3 - 4s_2s_3 + 10s_1 - 10s_2 + 20s_3 + 50
\end{aligned}$$

$$\nabla(\bar{f})(x_N) = \begin{pmatrix} 2s_1 + -2s_2 + 2s_3 + 10 \\ 4s_2 + -2s_1 - 4s_3 - 10 \\ 6s_3 + 2s_1 - 4s_2 + 20 \end{pmatrix} = \begin{array}{cc} \begin{pmatrix} 5 \\ 0 \\ 10 \end{pmatrix} & \begin{matrix} s_1 \\ s_2 \\ s_3 \end{matrix} \end{array}$$

Exemple : Itération 3

$$d_N = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \begin{array}{l} \leftarrow \nabla(\bar{f})(x_N)_{s_1} > 0, s_1 = 0 \\ \leftarrow \nabla(\bar{f})(x_N)_{s_2} = 0, s_2 \neq 0 \\ \leftarrow \nabla(\bar{f})(x_N)_{s_3} > 0, s_3 = 0 \end{array}$$

$$d_B \leftarrow -A_B^{-1} A_N d_N = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$d = 0 \Rightarrow$ (KKT) vérifiées. On renvoie $x^* = (2.5, 2.5, 5, 0, 2.5, 0)$ avec $f(x^*) = 31.25$.
De plus, f est convexe, les g_i (conditions de positivité des variables) sont linéaires sont convexes, les h_j sont linéaires, la condition suffisante de (KKT) montre donc que x^* est un **minimum global**.