# Chapter 5 : Non linear optimization, gradient methods

## ENSIIE - Operations Research Module

Dimitri Watel (dimitri.watel@ensiie.fr)

2024

# Problems MIDA

## Minimum Intersecting Disks Area

Let $p_1, p_2, \ldots, p_n, q$ be points in 2D space, build for each point $p_i$ a disk $D_i$ centered at $p_i$ such that

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

$$q \in D_n$$

and such that the sum of the areas of the disks is minimum.

# Modelization : MIDA

Let $d_{ij}$ the distance between $p_i$ and $p_j$ ; $d_n$ the distance between $p_n$ and $q$.

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

If $r$ is a solution with $r_i < 0$ we can safely replace $r_i$ by $-r_i$ and get a solution with the same objective value.

# Karush-Kuhn-Tucker conditions

Let $f, g_i, h_j : \mathbb{R}^n \to \mathbb{R}, \forall\ i \in [\![1; m]\!], \forall\ j \in [\![1; p]\!]$ of class $C^1$.
We want to solve

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

We write $S$ the set of feasible solutions.

## Karush-Kuhn-Tucker conditions

$x \in S$ satisfies the Karush-Kuhn-Tucker conditions (KKT) if:

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

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

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

# Reminder : qualification

Let $x \in S$. We write the saturated constraints

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

**Linear independance qualification**

$x$ satisfies the linear independance qualification if the vectors $(\nabla g_i(x), i \in I(x^*) \cup (\nabla h_j(x), j \in [\![1; p]\!])$ are linearly independant.

**Remark**: there exists a more general definition of qualification, but we will not use it.

# Reminder : necessary and sufficient optimality conditions

**(KKT) necessary conditions**

If $x^* \in S$ is a local minimum of $f$ and if $x^*$ is qualified then $x^*$ satisfies (KKT).

**(KKT) sufficient conditions**

If $x^*$ satisfies (KKT), if the functions $f$ et $g_i$ are convex and if the functions $h_j$ are linear, then $x^*$ is a global minimum.

# Qualification for MIDA

MIDA : Every point satisfies the linear independance qualification.

$\Rightarrow$ Every global optimum satisfies (KKT).

# Projected gradient algorithm

We want to solve

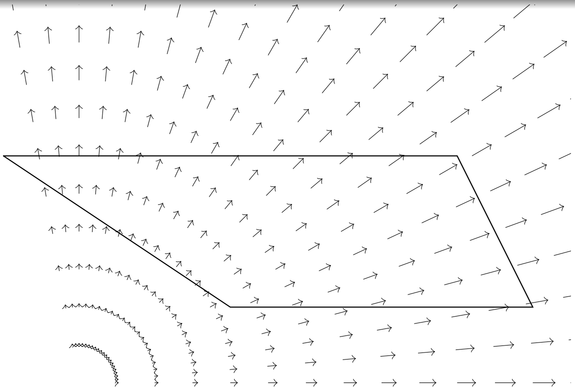
$$\begin{array}{lll} \min & f(x) & \\ s.c. & g_i(x) \leq 0 & \forall\ 1 \leq i \leq m \\ & h_j(x) = 0 & \forall\ 1 \leq j \leq p \end{array}$$

where $g_i$ and $h_j$ are **linear** and every point satisfies the linear independance qualification.

## Informal presentation of the algorithm

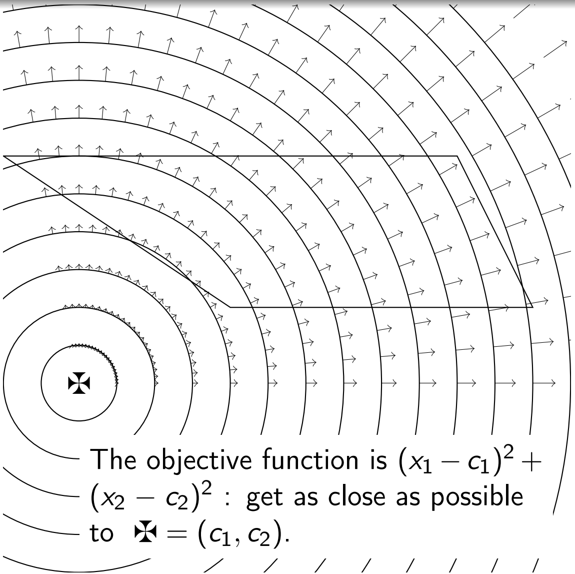
The projected gradient algorithm is a gradient descent that does not go out of $S$. To do so, when we touch the boundary of $S$, we do not follow the gradient but a projection of the gradient on the boundary.
It stops when the (KKT) conditions are satisfied.

# Projected gradient : overall idea



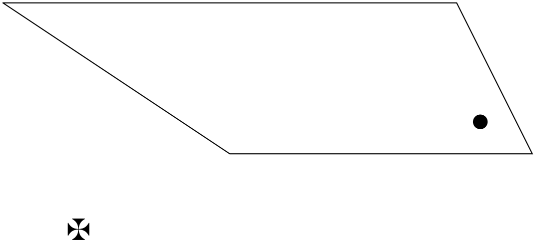
Arrows are the gradient. The polygon is $S$.

# Projected gradient : overall idea



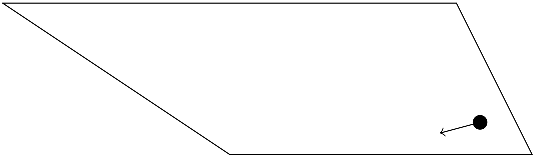
The objective function is $(x_1 - c_1)^2 + (x_2 - c_2)^2$ : get as close as possible to ✠ $= (c_1, c_2)$.

# Projected gradient : overall idea



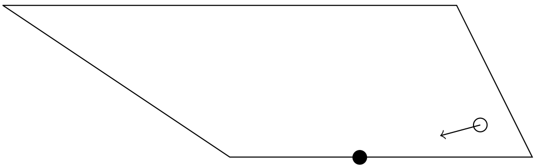
Step 1 : find a feasible solution $x$.

# Projected gradient : overall idea



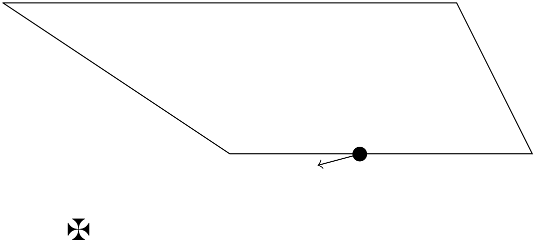
Step 2 : watch the opposite of the gradient $-\nabla f(x)$.

# Projected gradient : overall idea



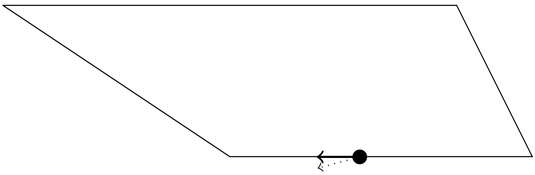
Step 3 : we can follow that direction, go to the minimum but do not go out of $S$. Then start again.

# Projected gradient : overall idea



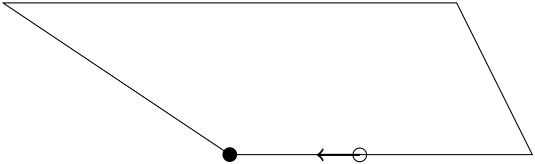
Step 2 : watch the opposite of the gradient $-\nabla f(x)$.

# Projected gradient : overall idea



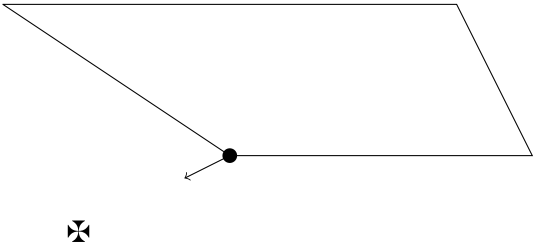
Step 3 : it is not possible to follow that direction without going out of $S$. We project $-\nabla f(x)$ on the boundary of $S$.

# Projected gradient : overall idea



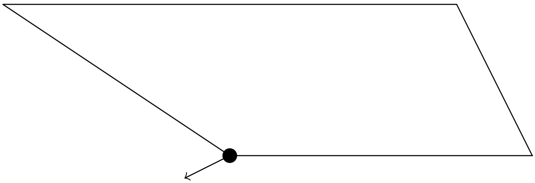
Step 4 : go to the minimum in that direction, but do not go out of $S$. And start again.

# Projected gradient : overall idea



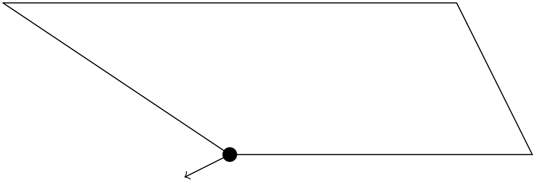
Step 2 : watch the opposite of the gradient $-\nabla f(x)$.

# Projected gradient : overall idea



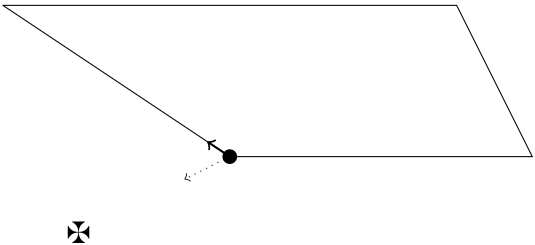
Step 3 : it is not possible to follow that direction without going out of $S$. We project $-\nabla f(x)$ on the boundary of $S$.

# Projected gradient : overall idea



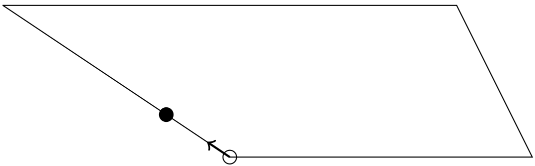
Step 4 : We need to project on 2 non-colinear borders ⇒ which border do we choose ?

# Projected gradient : overall idea



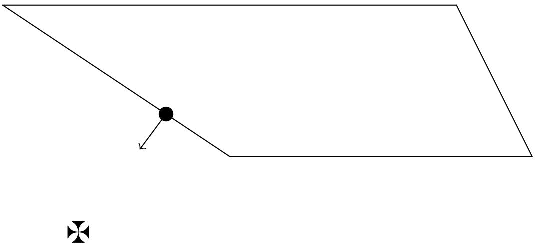
Step 4 : Project on the border for which the projection goes toward a lower value of the objective.

# Projected gradient : overall idea



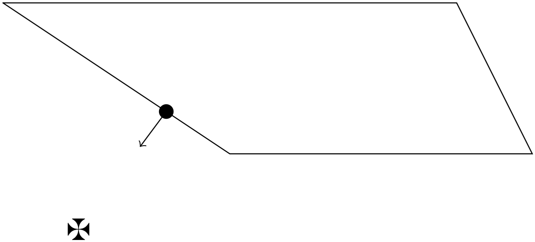
Step 5 : go to the minimum in that direction, but do not go out of $S$.
And start again.

# Projected gradient : overall idea



Step 2 : watch the opposite of the gradient $-\nabla f(x)$.

# Projected gradient : overall idea



Step 3 : The gradient is orthogonal to the boundary, we stop.

# Drawing $\rightarrow$ formulas.

We remind that $g_i$ and $h_j$ are linear:

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

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

We write

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

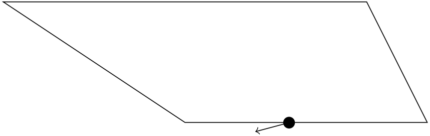
# Drawing $\rightarrow$ formulas.

**Lemma**

If $L_i$ is the $i$-th line of $A$, then

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

# Detect if we are on the boundary

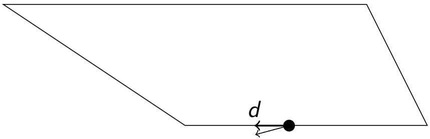


$x$ touches the hyperplanes with equations $h_j(x) = 0$ and $g_i(x) = 0$.
We write

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

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

# Project the gradient on the boundary




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

**Lemma**

The projected gradient $d$ is

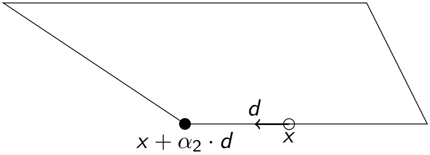
$$\begin{aligned} d &= \text{Projection of } (-\nabla f(x)) \text{ on } \{y | L_i \cdot y = 0 \ \forall i \in I(x) \cup J\} \\ &= \left(I_n - {}^tA_s \cdot (A_s \cdot {}^tA_s)^{-1} \cdot A_s\right) \cdot (-\nabla f(x)) \end{aligned}$$

# Remarks

$A_s \cdot {}^t A_s$ is always invertible due to the linear independance qualification.

If $I(x) = J = \emptyset$ then, $A_s = \emptyset$. In that case, $d = (-\nabla f(x))$.

# Translation in the direction $d$




We write

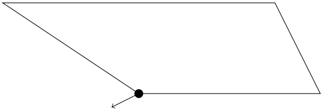
$$\alpha_1 = \max_{0 \leq \alpha}(x + \alpha \cdot d \in S)$$

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

## Lemma

If $d \neq \vec{0}$ then $\alpha_1, \alpha_2 \neq 0$.

# Case $d = \vec{0}$



## Lemma

If $d = \vec{0}$, then, there exists $\lambda_i$, for $i \in I(x)$ and $\mu_j$ for $j \in J$ such that $-\nabla f(x) = \sum_{i \in I(x)} \lambda_i \cdot L_i + \sum_{j \in J} \mu_j \cdot L_{m+j}$.

## Lemma

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

# Case $d = \vec{0}$

**Lemma**

If, for every $i \in I(x)$, $\lambda_i \geq 0$ then, the (KKT) conditions are satisfied.

**Lemma**

If there exists $\lambda_i < 0$, then we can remove $i$ from $I(x)$. We compute again the projection $d'$ like we did at slide 14.

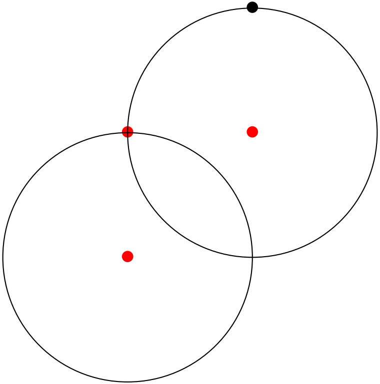
- $d' \neq \vec{0}$
- $\nabla g_i(x) \cdot d' < 0$

# The projected gradient algorithm

Find $x \in S$
$I(x) \leftarrow \{i \in [\![1; m]\!] | g_i(x) = 0\}$
**loop**
  $A_s \leftarrow \{L_i, i \in I(x) \cup J\}$
  $d \leftarrow (I_n - {}^tA_s \cdot (A_s \cdot {}^tA_s)^{-1} \cdot A_s) \cdot (-\nabla f(x))$
  **if** $d = 0$ **then**
    ${}^t(\lambda, \mu) \leftarrow (A_s \cdot {}^tA_s)^{-1} A_s \cdot (-\nabla f(x))$
    **If** $\exists i \in I(x), \lambda_i < 0$, Remove $i$ from $I(x)$
    Else **return** $x$
  **else**
    $\alpha_1 \leftarrow \max\limits_{0 \leq \alpha}(x + \alpha \cdot d \in S)$
    $\alpha_2 \leftarrow \arg \min\limits_{0 \leq \alpha \leq \alpha_1} (f(x + \alpha \cdot d))$
    $x \leftarrow x + \alpha_2 \cdot d$
    $I(x) \leftarrow \{i \in [\![1; m]\!] | g_i(x) = 0\}$

# Example

We consider the following MIDA instance
$P = \{p_1, p_2, p_3\} = \{[0, 0], [0, 5], [5, 5]\}$ and $q = [5, 10]$.
We start from the solution $r = (r_1, r_2, r_3) = (5, 0, 5)$.



# Example

The formulation of the problem is

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

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

# Example : Iteration 1

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

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

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

# Example : Iteration 1

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

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

Other method : we compute the projection of $-\nabla f(r)$ on

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

The projection is then necessarily $d = 0$.

# Example : Iteration 1

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

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

Other method :

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

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

# Example : Iteration 1.2

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

$$A_s = \begin{pmatrix} -1 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{matrix} L_1 \\ L_3 \end{matrix}$$

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

# Example : Iteration 1.2

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

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

Other method : we compute the projection of $-\nabla f(r)$ on

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

The projection is then necessarily $d = \beta \cdot (1, -1, 0)$ knowing that

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

# Example : Iteration 1, end

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

# Example : Iteration 1, end

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

We move to $r + \frac{1}{2}d = \begin{pmatrix} 2.5 \\ 2.5 \\ 5 \end{pmatrix}$.

# Example : Iteration 2

If $r = (2.5, 2.5, 5)$ then $I(r) = \{1, 3\}$.

$$A_s = \begin{pmatrix} -1 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{matrix} L_1 \\ L_3 \end{matrix}$$

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

# Example : Iteration 2

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

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

Other method : we compute the projection of $-\nabla f(r)$ on

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

The projection is then necessarily $d = \beta \cdot (1, -1, 0)$ knowing that

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

# Example : Iteration 2

$$A_s = \begin{pmatrix} -1 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{matrix} L_1 \\ L_2 \end{matrix}$$

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

Other method :

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

$\lambda_1, \lambda_3 \geq 0 \Rightarrow$ (KKT) satisfied. We return $r^* = (2.5, 2.5, 5)$ with $f(r^*) = 31.25$.
Moreover, $f$ is convex and the $g_i$ are convex thus linear, consequently, the sufficient conditions of (KKT) show that $r^*$ is a **global minimum**.

# Reduced gradient algorithm

We want to solve

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

where $h_j$ are **linear** and $p < n$.

## Informal presentation of the algorithm

The reduced gradient algorithm is a gradient descent that does not go out of $S$. Because of the equalities, we can rewrite some of the variables $x_B$ with the rest of the variables $x_N$. We then *reduce* $f(x)$ and $\nabla f(x)$ to $f(x_N)$ and $\nabla f(x_N)$
It stops when the (KKT) conditions are satisfied.

# Reminder

## Augmented form

Let $(P)$ be the following program

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

then, there exists a program $(P')$ equivalent to $(P)$ with the following form

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

# Reminder : tricks

If $g_i(x) \leq 0$ then add a slack variable $s_i \geq 0$ and set $h_i(x) = g_i(x) + s_i = 0$.

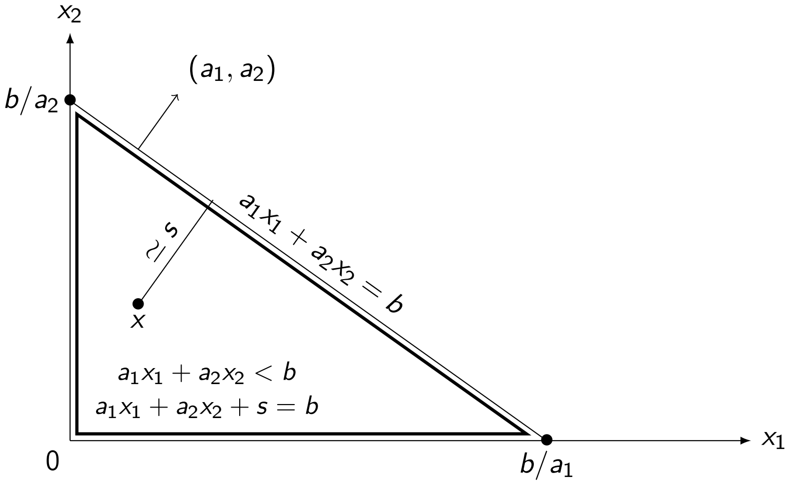
If $x_i \in \mathbb{R}$ then remplace $x_i$ by two variables $x_i^+ \geq 0$ and $x_i^- \geq 0$ and set $x_i = x_i^+ + x_i^-$.

If $g_i(x) = -x_i \leq 0$ then **do nothing**.

If $g_i(x) = x_i \leq 0$ then replace $x_i$ by $-x_i$.

# Reminder : slack variable and drawing

We assume $b > 0$



# No drawing

We remind that $h_j$ are linear:

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

We write

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

We assume that $rg(A) = p$.

# Basis and non basis

Let $B \subset [\![1; n]\!]$ of size $p$ and $N = [\![1; n]\!] \backslash B$.
We write
$A_B$ = the columns of $A$ for which the index is in $B$.
$A_N$ = the columns of $A$ for which the index is in $N$.
$x_B$ = the variables of $x$ for which the index is in $B$.
$x_N$ = the variables of $x$ for which the index is in $N$.
$\nabla f_B$ = the coefficients of $\nabla f(x)$ for which the index is in $B$.
$\nabla f_N$ = the coefficients of $\nabla f(x)$ for which the index is in $N$.

We can rewrite $A$, $x$ and $\nabla f(x)$ this way:

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

# Reduction, reduced gradient

**Lemma**

If $A_B$ is invertible, it is possible to rewrite $f(x)$ with only $x_N$ with

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

Let $\bar{f} : \mathbb{R}^{n-p} \to \mathbb{R}$ such that $\bar{f}(x_N) = f(x_B, x_N)$.

**Lemma**

The *reduced gradient* is the gradient of $\bar{f}$.

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

# Direction to follow

## Definition

Given $\nabla \bar{f}(x_N)$, we define the direction $d$, divided in two parts, $d_B$ and $d_N$, this way

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

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

## Lemma

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

# Case $d = \vec{0}$

**Lemma**

If $d = \vec{0}$ then, the (KKT) conditions are satisfied.

# Translation in the direction $d \neq \vec{0}$

As in the case of the projected gradient, we write

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

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

We move to

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

**Remark**

$\alpha_1$ may possibly be nul.

# Change of basis

**Lemma**

If there exists $s \in B$ such that $y_s = 0$ ; and if we keep the basis for the next iteration, then the next value of $\alpha_1$ may be 0 $\rightarrow$ we are stuck.

**Non-degeneration hypothesis**

The hypothesis assumes that, whatever the basis $B$ is, if $A_B$ is invertible, then $A_B^{-1} \cdot b > 0$.

**Lemma**

If there exists $s \in B$ such that $y_s = 0$, then, under the non-degeneration hypothesis, there exists $r \in N$ such that $y_r \neq 0$ and $A_{B \cup \{r\} \setminus \{s\}}$ is invertible.

# The reduced gradient algorithm

Find $x \in S$ and $B$ such that $A_B$ is invertible

**loop**

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

  **for** $j \in N$ **do**

    **If** $\nabla \bar{f}(x_N)_j > 0$ and $x_j = 0$ **then** $d_j \leftarrow 0$ **else** $d_j \leftarrow -\nabla \bar{f}(x_N)_j$

  $d_B \leftarrow -A_B^{-1} A_N d_N$

  **If** $d = 0$ **then return** $x$

  $\alpha_1 \leftarrow \max\limits_{0 \leq \alpha}(x + \alpha \cdot d \geq 0)$

  $\alpha_2 \leftarrow \arg \min\limits_{0 \leq \alpha \leq \alpha_1} (f(x + \alpha \cdot d))$

  $x \leftarrow x + \alpha_2 \cdot d$

  **if** $\exists s$ such that $x_s = 0$ **then**

    **for** $r \in N$ by decreasing order of $x_r$ **do**

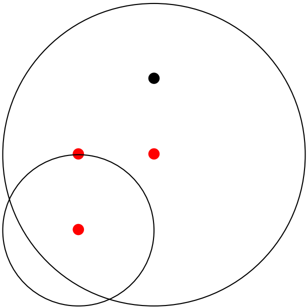
      **if** $A_{B \cup \{r\} \setminus \{s\}}$ is invertible **then**

        $B \leftarrow B \cup \{r\} \setminus \{s\}$

        Restart the loop

# Example

We consider the following MIDA instance
$P = \{p_1, p_2, p_3\} = \{[0,0], [0,5], [5,5]\}$ and $q = [5,10]$.
We start from the solution $r = (r_1, r_2, r_3) = (5, 0, 10)$.



# Example

The formulation of the problem is

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

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

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

$r = (5, 0, 10) \Rightarrow x = (5, 0, 10, 0, 5, 5)$
We start from the basis $B = (1, 3, 6)$ et $N = (2, 4, 5)$.

# Example : Iteration 1

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

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

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

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

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

# Example : Iteration 1

Other method: rewrite $f$ as a function of $x_N = (r_2, s_1, s_2) = (0, 0, 5)$

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

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

# Example : Iteration 1

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

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

# Example : Iteration 1

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

$$= \max_{0 \leq \alpha} \begin{pmatrix} 5 - 30 \cdot \alpha \geq 0 \\ 0 + 30 \cdot \alpha \geq 0 \\ 10 - 50 \cdot \alpha \geq 0 \\ 0 + 0 \cdot \alpha \geq 0 \\ 5 - 20 \cdot \alpha \geq 0 \\ 5 - 50 \cdot \alpha \geq 0 \end{pmatrix}$$

$$= 0.1$$

# Example : Iteration 1

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

We move to $x + 0.1d = \begin{pmatrix} 2 \\ 3 \\ 5 \\ 0 \\ 3 \\ 0 \end{pmatrix}$

# Example : Change of basis

$s_3 = 0$, however $s_3 \in B$ then: change of basis.
$r_2 = 3, s_1 = 0, s_2 = 3$
We try $B' = B - \{s_3\} + \{r_2\} = (1, 2, 3)$

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

Thus we replace $B$ by $B'$.

# Example : Iteration 2

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

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

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

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

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

# Example : Iteration 2

Other method: rewrite $f$ as a function of $x_N = (s_1, s_2, s_3) = (0, 3, 0)$

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

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

# Example : Iteration 2

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

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

# Example : Iteration 2

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

$$= \max_{0\leq\alpha} \begin{pmatrix} 2 + 2 \cdot \alpha \geq 0 \\ 3 - 2 \cdot \alpha \geq 0 \\ 5 + 0 \cdot \alpha \geq 0 \\ 0 + 0 \cdot \alpha \geq 0 \\ 3 - 2 \cdot \alpha \geq 0 \\ 0 + 0 \cdot \alpha \geq 0 \end{pmatrix}$$

$$= 1.5$$

# Example : Iteration 2

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

We move to $x + 0.25d = \begin{pmatrix} 2.5 \\ 2.5 \\ 5 \\ 0 \\ 2.5 \\ 0 \end{pmatrix}$

# Example : Iteration 2; Change of basis

No $j \in B$ such that $x_j = 0$, then no change of basis.

# Example : Iteration 3

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

$$x_B = (2.5, 2.5, 5) \qquad x_N = (0, 2.5, 0)$$

$$^t\nabla f_B = (5, 5, 10) \qquad ^t\nabla f_N = (0, 0, 0)$$

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

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

# Example : Iteration 3

Other method: rewrite $f$ as a function of
$x_N = (s_1, s_2, s_3) = (0, 2.5, 0)$

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

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

# Example : Iteration 3

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

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

$d = 0 \Rightarrow$ (KKT) satisfied. We return $x^* = (2.5, 2.5, 5, 0, 2.5, 0)$ with $f(x^*) = 31.25$.
Moreover, $f$ is convex, the $g_i$ (the positivity constraints) are convex thus linear and the $h_j$ are linear, consequently, the sufficient conditions of (KKT) show that $x^*$ is a **global minimum**.