

Techniques d'optimisation

Sommaire

1. [Bases théoriques](#)
2. [Optimisation sans contraintes](#)
3. **Optimisation avec contraintes**
 - 3.1 [Simplexe](#)
 - 3.2 [Point intérieur](#)
 - 3.3 [Gradient projeté](#)
 - 3.4 [Lagrangien augmenté](#)
 - 3.5 [Programmation quadratique séquentielle](#)
 - 3.6 [Convergence](#)
4. [Optimisation discrète](#)
5. [Optimisation fonctionnelle](#)

3 Optimisation avec contraintes

Problème non linéaire sous contraintes

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{sous} \quad \begin{cases} c_E(x) = 0 \\ c_I(x) \leq 0 \end{cases} \quad \rightarrow \text{Problème noté (PO)}$$

Catégories de problèmes

- **Programmation linéaire** $\rightarrow$ Fonctions f, c_E, c_I linéaires
- **Programmation non linéaire** $\rightarrow$ Fonctions f, c_E, c_I quelconques

Traitement des contraintes

- Méthodes de contraintes actives $\rightarrow$ Identification des inégalités actives
Transformation en un **problème avec contraintes égalité**
Respect des contraintes à chaque itération
- Méthodes de point intérieur $\rightarrow$ Fonction barrière (pénalisation intérieure)
Suivi d'un **chemin central intérieur aux contraintes**
- Méthodes de pénalisation $\rightarrow$ Critère augmenté (pénalisation extérieure)
Transformation en un **problème sans contraintes**

3 Optimisation avec contraintes

Problème non linéaire sous contraintes

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{sous} \quad \begin{cases} c_E(x) = 0 \\ c_I(x) \leq 0 \end{cases} \quad \rightarrow \text{Problème noté (PO)}$$

Classification des méthodes

	Méthode primale	Méthode primale-duale	Méthode duale
Problème traité	problème primal	problème primal	problème dual
Objectif	min f - méthode directe - point stationnaire	solution KKT - méthode indirecte - point stationnaire	max w - méthode indirecte - point col
Itérations	admissibles	admissibles ou non	non admissibles
Variables	primales x	primales x , duales λ	primales x , duales λ
Algorithmes	- simplexe (LP) - gradient projeté - pénalisation	- point intérieur (LP, NLP) - séquentiel quadratique	- simplexe dual (LP) - lagrangien augmenté - Uzawa

Sommaire

1. [Bases théoriques](#)
2. [Optimisation sans contraintes](#)
3. **Optimisation avec contraintes**
 - 3.1 [Simplexe](#)
 - 3.2 [Point intérieur](#)
 - 3.3 **Gradient projeté**
 - 3.3.1 [Principes](#)
 - 3.3.2 [Direction de déplacement](#)
 - 3.3.3 [Restauration](#)
 - 3.3.4 [Algorithme](#)
 - 3.4 [Lagrangien augmenté](#)
 - 3.5 [Programmation quadratique séquentielle](#)
 - 3.6 [Convergence](#)
4. [Optimisation discrète](#)
5. [Optimisation fonctionnelle](#)

3.3.1 Gradient projeté

Problème avec contraintes égalité

$$\min_{x \in \mathbb{R}^n} f(x) \text{ sous } \begin{cases} c_E(x) = 0 \\ c_I(x) \leq 0 \end{cases} \Leftrightarrow \min_{x \in \mathbb{R}^n} f(x) \text{ sous } c(x) = 0 \rightarrow m \text{ contraintes actives}$$

Etapes principales

A chaque itération :

- Construction d'une direction de descente d_k à partir du point x_k
- Réglage du pas de déplacement s_k suivant d_k

Direction de descente

On construit la direction d_k dans l'hyperplan tangent aux contraintes (= espace nul) en x_k

- **Gradient projeté** → projection du gradient sur l'hyperplan tangent
- **Gradient réduit** → réduction du gradient sur une base de l'espace nul

Pas de déplacement

- Recherche linéaire suivant d_k → pas s_k
- Restauration de l'admissibilité → méthode de Newton
- Règles d'acceptation du pas → Armijo, Goldstein, Wolfe

3.3.2 Direction de déplacement

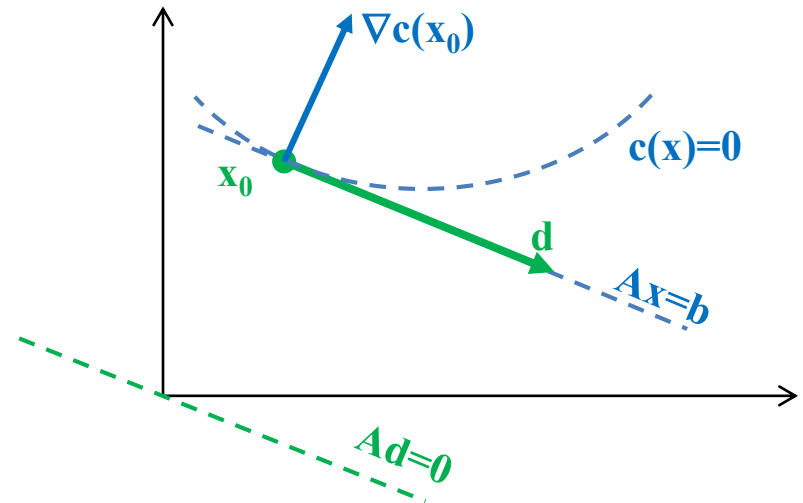
- ☐ [Hyperplan tangent aux contraintes](#)
- ☐ [Gradient projeté](#)
- ☐ [Gradient réduit](#)
- ☐ [Direction de déplacement](#)
- ☐ [Exemple](#)

3.3.2 Hyperplan tangent

Problème avec contraintes linéaires

$$\min_{x \in \mathbb{R}^n} f(x) \text{ sous } Ax = b$$

- x_0 point admissible $\rightarrow Ax_0 = b$
- d déplacement admissible à partir de x_0 $\rightarrow A(x_0 + d) = b$ $\left. \vphantom{\begin{matrix} \rightarrow Ax_0 = b \\ \rightarrow A(x_0 + d) = b \end{matrix}} \right\} \rightarrow Ad = 0$
- $Ad = 0$ définit l'espace nul des contraintes
= hyperplan des contraintes
- Le déplacement $d \in \mathbb{R}^n$ est admissible si d est dans l'hyperplan des contraintes.



Problème avec contraintes non linéaires

$$\min_{x \in \mathbb{R}^n} f(x) \text{ sous } c(x) = 0$$

- On définit l'**espace nul tangent** ou **hyperplan tangent** en x_0 avec $A = \nabla c(x_0)^T \rightarrow Ad = 0$
- On cherche un déplacement $d \in \mathbb{R}^n$ dans l'hyperplan tangent : $\nabla c(x_0)^T d = 0$
Un déplacement complémentaire est ensuite nécessaire pour restaurer l'admissibilité.

3.3.2 Gradient projeté

Définition

$$\min_{x \in \mathbb{R}^n} f(x) \text{ sous } c(x) = 0$$

Le gradient projeté est la **projection orthogonale** du gradient de f sur l'hyperplan tangent.

Expression du gradient projeté

- Hyperplan tangent aux contraintes en x_0 admissible
 $Ad = 0$ avec $A = \nabla c(x_0)^T$

- Matrice de projection sur l'hyperplan tangent

$$P = I - A^T (AA^T)^{-1} A$$

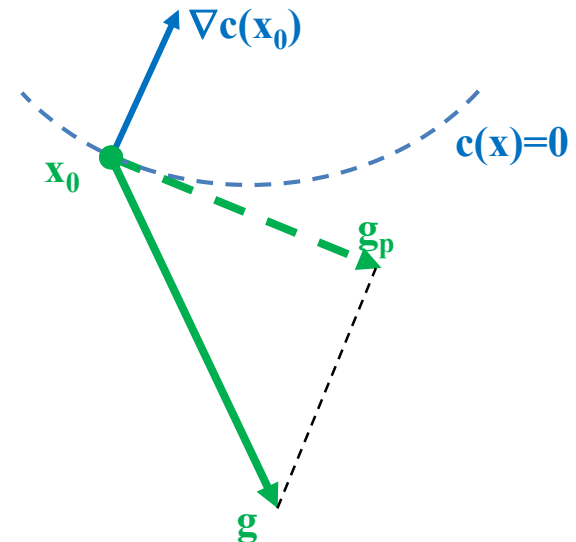
- Notations**

$$g(x_0) \text{ gradient de } f \text{ en } x_0 \rightarrow g(x_0) = \nabla f(x_0)$$

$$g_p(x_0) \text{ gradient projeté} \rightarrow g_p(x_0) = Pg(x_0)$$

$$\rightarrow g_p = \left(I - A^T (AA^T)^{-1} A \right) g \text{ avec } A = \nabla c(x_0)^T$$

- g_p vérifie : $\begin{cases} Ag_p = 0 \\ g_p^T (g - g_p) = 0 \end{cases} \rightarrow g_p \in \text{hyperplan tangent}$
 $\rightarrow g - g_p \perp \text{hyperplan tangent}$



$$\text{car } \begin{cases} P^T = P \\ P^2 = P \end{cases}$$

3.3.2 Gradient projeté

Direction de descente

- La direction du gradient projeté est la **direction de plus forte pente dans l'hyperplan tangent**
 = direction dans l'hyperplan qui maximise la dérivée directionnelle de f

Preuve

La direction d dans l'hyperplan maximisant la dérivée directionnelle de f est solution de

$$\min_{d \in \mathbb{R}^n} g^T d \quad \text{sous} \quad \begin{cases} Ad = 0 & \rightarrow d \in \text{hyperplan tangent} \\ \|d\| = 1 \Leftrightarrow d^T d = 1 & \rightarrow \text{norme} = 1 \end{cases}$$

Lagrangien : $L(d, \lambda, \mu) = g^T d + \lambda^T Ad + \mu(d^T d - 1)$ avec $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}$

$$\text{Conditions KKT : } \begin{cases} g + A^T \lambda + 2\mu d = 0 & \rightarrow d = -(g + A^T \lambda) / (2\mu) \\ Ad = 0 & \rightarrow Ag + AA^T \lambda = 0 \rightarrow \lambda = -(AA^T)^{-1} Ag \\ \|d\| = 1 & \rightarrow 2\mu = \pm \|g + A^T \lambda\| \end{cases}$$

d est bien un vecteur normé colinéaire à $(I - A^T (AA^T)^{-1} A)g$

- La méthode du gradient projeté équivaut à la méthode de plus forte pente appliquée dans l'espace nul des contraintes $\rightarrow$ méthode d'ordre 1 peu efficace
 $\rightarrow$ amélioration par méthode de quasi-Newton

3.3.2 Gradient réduit

Définition

$$\min_{x \in \mathbb{R}^n} f(x) \text{ sous } c(x) = 0$$

Le gradient réduit est le gradient de la **fonction réduite** sur une base de l'espace nul tangent.

Expression du gradient réduit

- Base Z de l'espace nul tangent aux contraintes : $AZ = 0$ avec $A = \nabla c(x_0)^T$
- Décomposition du déplacement : $p = Yp_Y + Zp_Z$ avec $\begin{cases} AZ = 0 \\ AY \text{ inversible} \end{cases}$
- Déplacement admissible : $Ap = 0 \Rightarrow \begin{cases} p_Y = 0 \\ p = Zp_Z \end{cases}$
- Fonction réduite f_r : $\min_{p \in \mathbb{R}^n} f(x_0 + p) \text{ sous } A(x_0 + p) = b \Leftrightarrow \min_{p_Z \in \mathbb{R}^{n-m}} f_r(p_Z) = f(x_0 + Zp_Z)$
- **Notations**
 - $g(x_0)$ gradient de f en $x_0 \rightarrow g(x_0) = \nabla f(x_0) \rightarrow g \in \mathbb{R}^n$
 - $g_r(x_0)$ gradient réduit $\rightarrow g_r(x_0) = \nabla f_r(p_Z=0) \rightarrow g_r \in \mathbb{R}^{n-m}$ (m = nombre de contraintes)
 - $f_r(p_Z) = f(x_0 + Zp_Z) \Rightarrow \nabla f_r(p_Z) = Z^T \nabla f(x_0 + Zp_Z) \rightarrow \boxed{g_r = Z^T g}$ en $p_Z = 0$
 - $\rightarrow g_r$ est le gradient de la fonction réduite f_r (= fonction de $n-m$ variables p_Z)

3.3.2 Gradient réduit

Direction de descente

- Le déplacement à partir du point x_0 admissible est décomposé en

$$p = Yp_Y + Zp_Z \quad \text{avec} \quad \begin{cases} AZ = 0 \\ AY \text{ inversible} \end{cases} \quad Ap = 0 \Rightarrow \begin{cases} p_Y = 0 \\ p = Zp_Z \end{cases}$$

- Le gradient réduit g_r donne la direction **de plus forte pente suivant les variables p_Z** .
La direction de déplacement dans R^n est : **$d = Zg_r$** .

- On peut choisir les matrices Y et Z
 - à partir de matrices orthogonales $\rightarrow$ factorisation QR de A
 - à partir d'une base de A $\rightarrow B \in R^{m \times m}$ (= m colonnes indépendantes de A)

Gradient réduit sur une base B de A

$$AE = \begin{pmatrix} m & n-m \\ B & N \end{pmatrix} \Rightarrow g = \begin{pmatrix} g_B \\ g_N \end{pmatrix} \quad Y = \begin{pmatrix} B^{-1} \\ 0 \end{pmatrix}_{n-m}^m \quad Z = \begin{pmatrix} -B^{-1}N \\ I \end{pmatrix}_{n-m}^m$$

(E = matrice de permutation de colonnes de A)

- Le gradient réduit par rapport à la base B est : $\boxed{g_r = Z^T g = g_N - (B^{-1}N)^T g_B}$

3.3.2 Direction de déplacement

Problème avec contraintes égalité

$\min_{x \in \mathbb{R}^n} f(x)$ sous $c(x) = 0 \rightarrow m$ contraintes actives

- On construit la direction de déplacement $d \in \mathbb{R}^n$ dans l'hyperplan tangent aux contraintes en x_0
 $Ad = 0$ avec $A = \nabla c(x_0)^T$
 $\rightarrow$ 2 méthodes de construction de la direction d

- Méthode du gradient projeté**

La direction d est celle du gradient projeté : $d = g_p$

$$d = Pg \quad \text{avec} \quad P = I - A^T (AA^T)^{-1} A \quad (P = \text{matrice de projection sur l'hyperplan tangent})$$

- Méthode du gradient réduit**

La direction d est obtenue à partir du gradient de la fonction réduite : $d = Zg_r$ avec $g_r = Z^T g$

$$d = ZZ^T g \quad \text{avec} \quad AZ = 0 \quad (Z = \text{base de l'hyperplan tangent})$$

- On cherche un pas de déplacement suivant $-d$ pour minimiser f .
 Un déplacement complémentaire est nécessaire pour restaurer l'admissibilité.

3.3.2 Exemple

Exemple

- Minimisation de $f(x) = x_1 + x_2$ sous $c(x) = x_1^2 + (x_2 - 1)^2 - 1 = 0$

$$\nabla f(x) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \nabla c(x) = \begin{pmatrix} 2x_1 \\ 2(x_2 - 1) \end{pmatrix}$$

- Changement de variables (coordonnées polaires)

$$\begin{cases} x_1 = r \cos \theta \\ x_2 = r \sin \theta + 1 \end{cases} \rightarrow \begin{cases} f(r, \theta) = r(\cos \theta + \sin \theta) + 1 \\ c(r, \theta) = r^2 - 1 \end{cases}$$

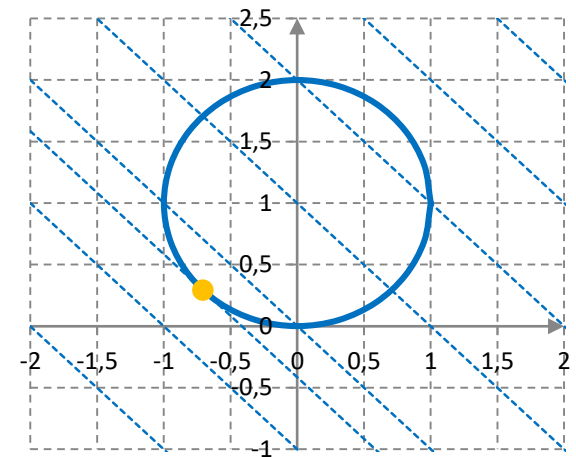
- Elimination variable r

$$r = 1 \rightarrow f(\theta) = \cos \theta + \sin \theta + 1$$

- Minimum**

$$\left\{ \begin{array}{l} f'(\theta) = -\sin \theta + \cos \theta = 0 \rightarrow \tan \theta = 1 \rightarrow \theta = \frac{\pi}{4} \text{ ou } \frac{\pi}{4} + \pi \\ f''(\theta) = -\cos \theta - \sin \theta \geq 0 \rightarrow -\cos \theta(1 + \tan \theta) \geq 0 \rightarrow -\cos \theta \geq 0 \end{array} \right\} \rightarrow \theta^* = \frac{\pi}{4} + \pi$$

$$\rightarrow \begin{cases} x_1^* = -1/\sqrt{2} \approx -0.70711 \\ x_2^* = 1 - 1/\sqrt{2} \approx 0.29289 \end{cases}$$



3.3.2 Exemple

Gradient projeté

- Minimisation de $f(x) = x_1 + x_2$ sous $c(x) = x_1^2 + (x_2 - 1)^2 - 1 = 0$

Point admissible $x_0 (r_0=1, \theta_0) \rightarrow \nabla f(x_0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\nabla c(x_0) = \begin{pmatrix} 2x_1 \\ 2(x_2 - 1) \end{pmatrix} = 2r_0 \begin{pmatrix} \cos \theta_0 \\ \sin \theta_0 \end{pmatrix}$

- Gradient projeté au point x_0

$$g_p = Pg \text{ avec } \begin{cases} A = \nabla c(x_0)^T, g = \nabla f(x_0) \\ P = I - A^T(AA^T)^{-1}A \end{cases}$$

$$A = 2r_0(\cos \theta_0 \quad \sin \theta_0)$$

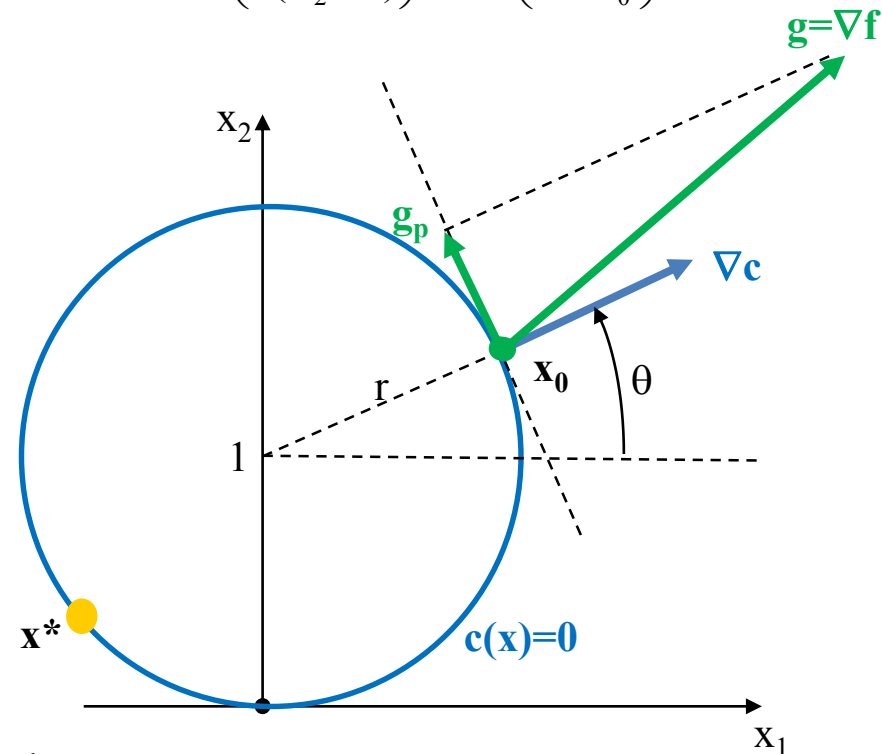
$$P = \begin{pmatrix} \sin^2 \theta_0 & -\sin \theta_0 \cos \theta_0 \\ -\sin \theta_0 \cos \theta_0 & \cos^2 \theta_0 \end{pmatrix}$$

$$\rightarrow g_p = (\cos \theta_0 - \sin \theta_0) \begin{pmatrix} -\sin \theta_0 \\ \cos \theta_0 \end{pmatrix}$$

- Direction de descente au point x_0

$$d = -\frac{g_p}{\|g_p\|} = \pm \begin{pmatrix} -\sin \theta_0 \\ \cos \theta_0 \end{pmatrix}$$

$\rightarrow$ **tangente** au cercle en x_0



3.3.2 Exemple

Gradient réduit

- Minimisation de $f(x) = x_1 + x_2$ sous $c(x) = x_1^2 + (x_2 - 1)^2 - 1 = 0$

Point admissible $x_0 (r_0=1, \theta_0) \rightarrow \nabla f(x_0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \nabla c(x_0) = \begin{pmatrix} 2x_1 \\ 2(x_2 - 1) \end{pmatrix} = 2r_0 \begin{pmatrix} \cos \theta_0 \\ \sin \theta_0 \end{pmatrix}$

- Gradient réduit au point x_0

$$g_r = Z^T g \text{ avec } \begin{cases} A = \nabla c(x_0)^T, & g = \nabla f(x_0) \\ AE = (B \quad N), & Z = \begin{pmatrix} -B^{-1}N \\ I \end{pmatrix} \end{cases}$$

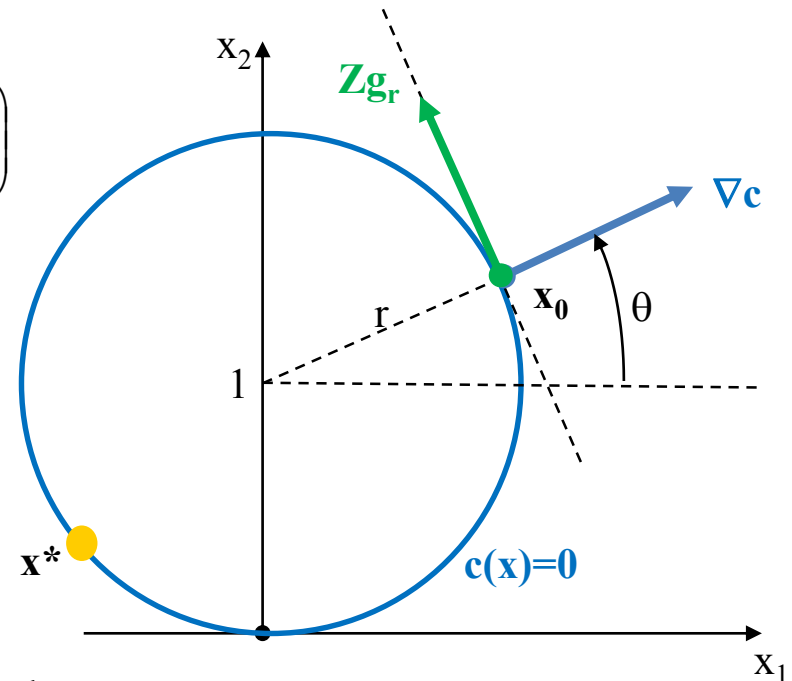
$$A = 2r_0(\cos \theta_0 \quad \sin \theta_0) \quad Z = \begin{pmatrix} -\tan \theta_0 \\ 1 \end{pmatrix}$$

$$\rightarrow g_r = 1 - \tan \theta_0$$

$$\rightarrow Zg_r = \frac{\cos \theta_0 - \sin \theta_0}{\cos \theta_0} \begin{pmatrix} -\sin \theta_0 \\ \cos \theta_0 \end{pmatrix}$$

- Direction de descente au point x_0

$$d = -\frac{Zg_r}{\|Zg_r\|} = \pm \begin{pmatrix} -\sin \theta_0 \\ \cos \theta_0 \end{pmatrix} \rightarrow \text{tangente au cercle en } x_0$$



3.3.3 Restauration

- ☐ [Point initial](#)
- ☐ [Itérations admissibles](#)
- ☐ [Méthode de restauration](#)

3.3.3 Restauration

Itérations admissibles

- La méthode du gradient projeté ou réduit construit une suite de solutions admissibles
- point initial admissible
 - restauration de la faisabilité à chaque itération

Point initial

- On peut construire un point initial admissible du problème $\min_{x \in \mathbb{R}^n} f(x)$ sous $\begin{cases} c_E(x) = 0 \\ c_I(x) \leq 0 \end{cases}$ en résolvant le **problème préliminaire sans contrainte** $\min_{x \in \mathbb{R}^n} \|c_E(x)\|_2 + \|\max(0, c_I(x))\|_2$
- La solution x_0 de ce problème préliminaire est admissible si le coût est nul.

$$\|c_E(x_0)\|_2 + \|\max(0, c_I(x_0))\|_2 = 0 \Rightarrow \begin{cases} c_E(x_0) = 0 \\ \max(0, c_I(x_0)) = 0 \end{cases} \Rightarrow \begin{cases} c_E(x_0) = 0 \\ c_I(x_0) \leq 0 \end{cases}$$

Restauration de la faisabilité

- La direction de descente d est dans l'hyperplan tangent aux contraintes au point courant.
- Si les contraintes ne sont pas linéaires, un pas s suivant d donne un point non admissible.
 → **Il faut restaurer la faisabilité avant d'évaluer si le pas s est acceptable.**

3.3.3 Restauration

Déplacement admissible

$$\min_{x \in \mathbb{R}^n} f(x) \text{ sous } c(x) = 0 \quad \rightarrow \text{ m contraintes actives}$$

On construit le déplacement p à partir du point initial x_0 en 2 étapes : $p = p_1 + p_2$

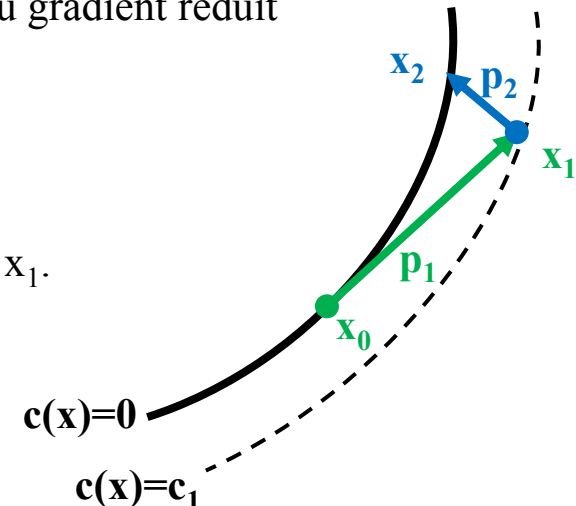
- Le déplacement p_1 est suivant la direction de descente d dans l'hyperplan tangent : $p_1 = -sd$
 $d \in \mathbb{R}^n$ = direction construite à partir du gradient projeté ou du gradient réduit
 $s > 0$ = pas de déplacement suivant $-d$ (pour minimisation)

On obtient un point $x_1 = x_0 + p_1$ dans l'hyperplan tangent

→ non admissible si contraintes non linéaires

- Le déplacement p_2 restaure un point admissible à partir du point x_1 .
 → linéarisation des contraintes en x_1
 → résolution d'un système sous-déterminé

On obtient un point $x_2 = x_1 + p_2$ admissible.



Recherche linéaire

- On évalue le point x_2 correspondant au pas s de recherche linéaire suivant d .
- Le pas s est modifié par dichotomie jusqu'à trouver un point $x_2(s)$ acceptable.
 → règles d'Armijo, Goldstein, Wolfe,...

3.3.3 Restauration

Méthode de restauration

Le déplacement p_2 doit vérifier : $A_1 p = b_1$ avec $\begin{cases} A_1 = \nabla c(x_1)^T \approx \nabla c(x_0)^T = A_0 \\ b_1 = -c(x_1) = -c_1 \end{cases}$

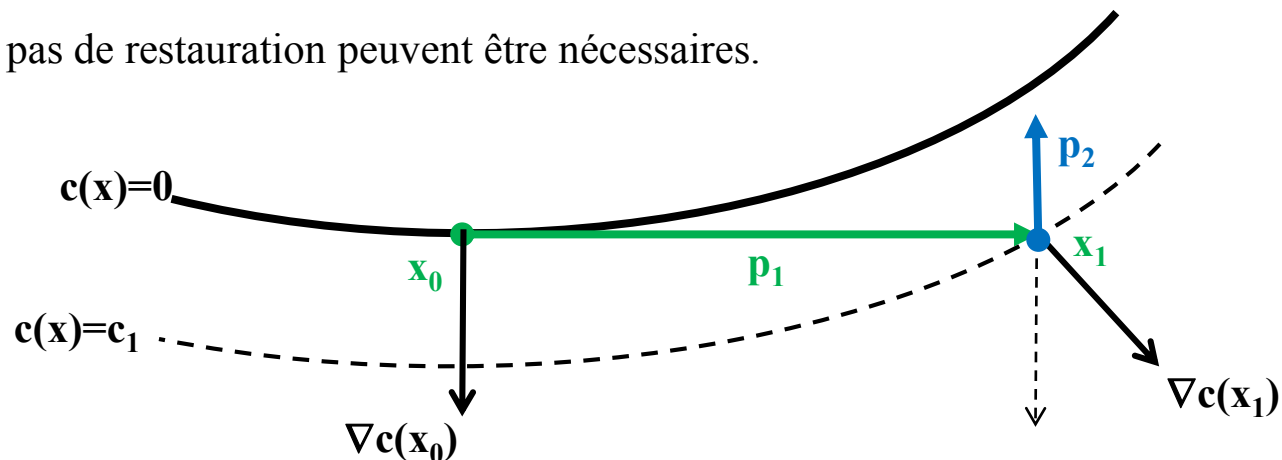
- **Solution de norme minimale** → projection sur l'hyperplan tangent aux contraintes actives (cf §1.2.4)

$$\min_{p \in \mathbb{R}^n} \|p\| \text{ sous } A_1 p = b_1 \quad \rightarrow \quad p_2 = A_1^T (A_1 A_1^T)^{-1} b_1$$

- **Solution de base** → pour ne pas dégrader la minimisation apportée par p_1 (cf §1.2.3)

$$A_1 (Y p_Y + Z p_Z) = b_1 \Rightarrow p_Y = (A_1 Y)^{-1} b_1 \rightarrow p_2 = Y (A_1 Y)^{-1} b_1$$

- Plusieurs pas de restauration peuvent être nécessaires.

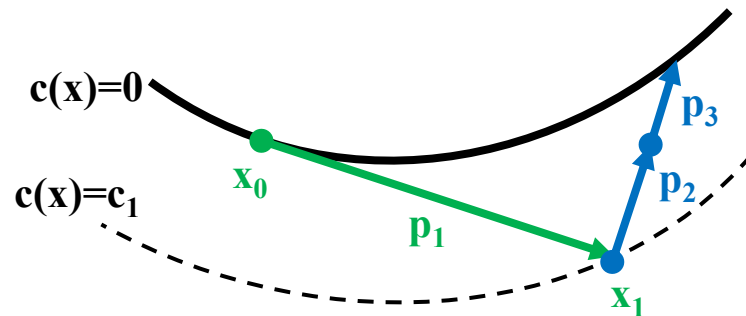


Techniques d'optimisation

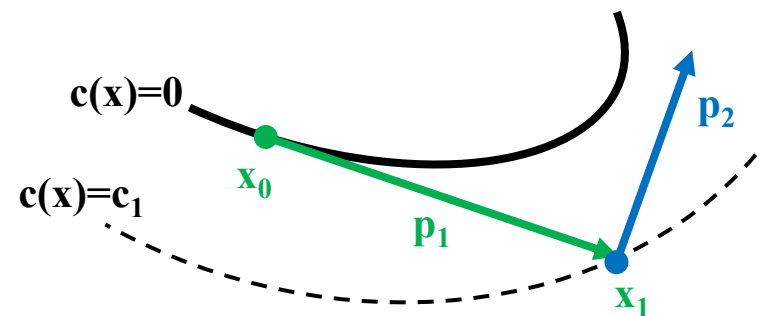
3.3.3 Restauration

Illustrations

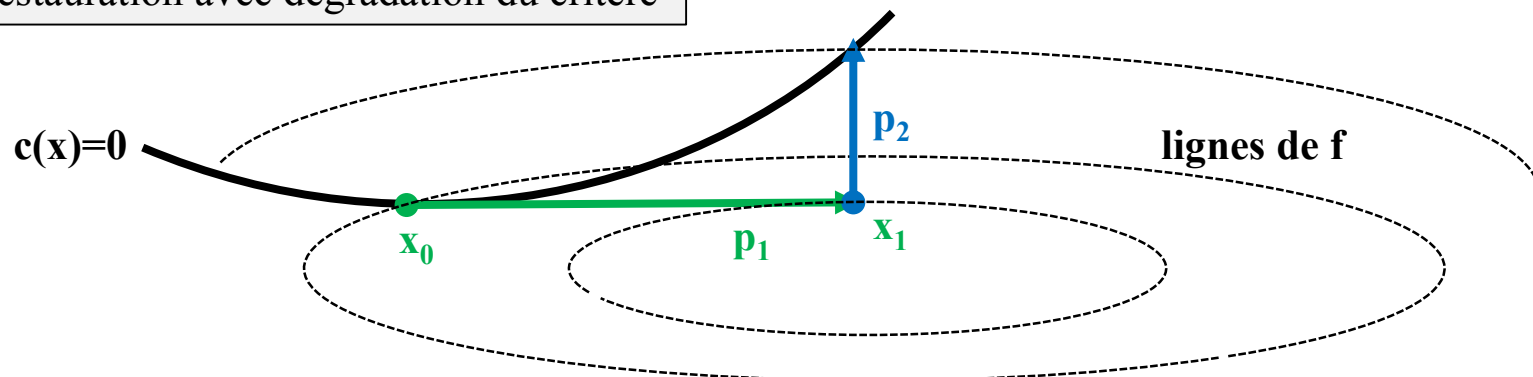
Restauration en plusieurs itérations : p_2, p_3



Restauration infructueuse (non linéarité)



Restauration avec dégradation du critère



3.3.4 Algorithme

- ❑ [Algorithme de gradient projeté / réduit](#)
- ❑ [Exemple](#)

3.3.4 Algorithme

Algorithme de gradient projeté/réduit

Les principales étapes d'une itération de gradient projeté/réduit sont

- construire la direction de descente au point courant
- effectuer une recherche linéaire avec restauration

Direction de descente

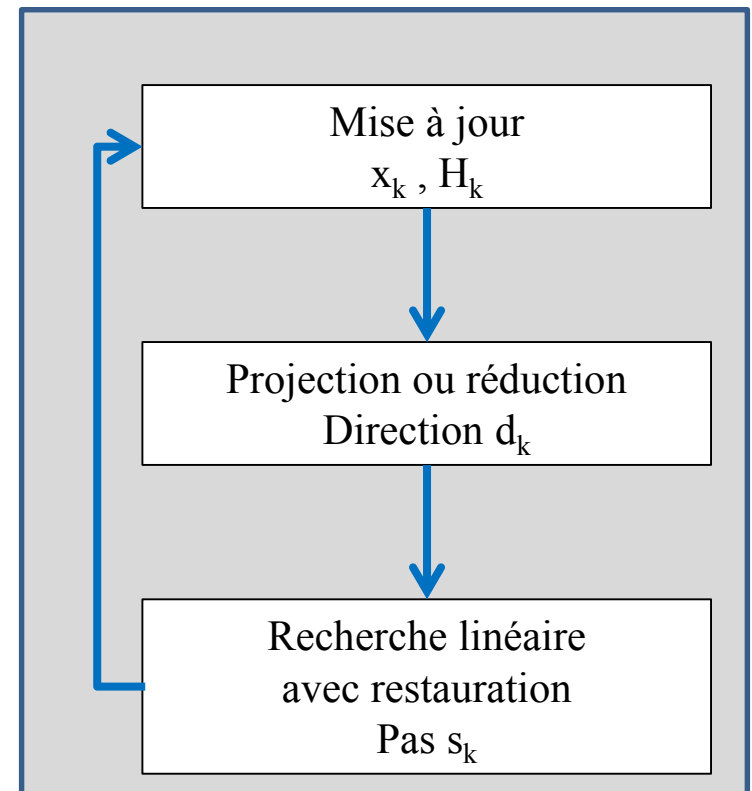
- Sélection des contraintes actives
- Projection ou réduction dans l'hyperplan tangent
- Mise à jour du hessien (quasi-Newton)

Recherche linéaire

- Méthode de dichotomie sur le pas de déplacement
- Restauration avant évaluation du pas
- Règles d'acceptation (Armijo,...)

Principales difficultés

- Amélioration critère → grands pas
- Restauration contraintes → petits pas
→ difficultés sur problèmes très non-linéaires
→ réglages à adapter au cas par cas



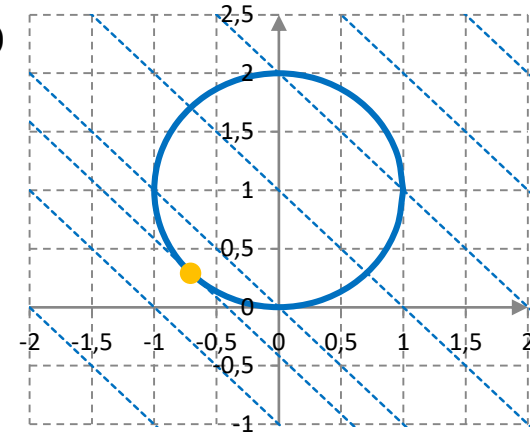
3.3.4 Exemple

Exemple

- Minimisation de $f(x) = x_1 + x_2$ sous $c(x) = x_1^2 + (x_2 - 1)^2 - 1 = 0$
- Solution :
$$\begin{cases} x_1^* = -1/\sqrt{2} \approx -0.70711 \\ x_2^* = 1 - 1/\sqrt{2} \approx 0.29289 \end{cases}$$

Itérations

- Point courant : $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} r \cos \theta \\ r \sin \theta + 1 \end{pmatrix}$
- Descente : $x' = x - s_1 d_1$ avec $d_1 = \pm \begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix} \leftarrow g_p = (\cos \theta - \sin \theta) \begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix}$
 → **pas s_1 suivant le gradient projeté**
- Restauration : $x'' = x' - s_2 d_2$ avec $d_2 = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} \leftarrow \nabla c(x) = \begin{pmatrix} 2x_1 \\ 2(x_2 - 1) \end{pmatrix}$
 → **pas s_2 suivant le gradient des contraintes**
- Réglage des pas : s_2 est calculé pour restaurer $c(x'') = 0$
 s_1 est choisi pour vérifier une décroissance suffisante $f(x'') < f(x)$



3.3.4 Exemple

Exemple

- Minimisation de $f(x) = x_1 + x_2$ sous $c(x) = x_1^2 + (x_2 - 1)^2 - 1 = 0$
- Point initial : $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0.1 \\ 1 \end{pmatrix} \rightarrow$ Restauration initiale : $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

Itération	x_1	x_2	$f(x)$	$c(x)$	Descente s_1	x_1'	x_2'	$c(x')$	Restauration s_2
1	0,10000	1,00000	1,10000	-0,99000	0,00000	0,10000	1,00000	-0,99000	4,50000
2	1,00000	1,00000	2,00000	0,00000	1,00000	1,00000	0,00000	1,00000	-0,50000
3	0,00000	0,00000	0,00000	0,00000	0,50000	-0,50000	0,00000	0,25000	-0,06699
4	-0,50000	0,13397	-0,36603	0,00000	0,18301	-0,65849	0,22548	0,03349	-0,00844
5	-0,65005	0,24011	-0,40994	0,00000	5,492E-02	-0,69178	0,27581	3,016E-03	-7,547E-04
6	-0,69080	0,27696	-0,41385	0,00000	1,612E-02	-0,70246	0,28809	2,599E-04	-6,497E-05
7	-0,70237	0,28819	-0,41418	0,00000	4,722E-03	-0,70573	0,29150	2,230E-05	-5,576E-06
8	-0,70572	0,29151	-0,41421	0,00000	1,383E-03	-0,70670	0,29249	1,913E-06	-4,783E-07
9	-0,70670	0,29249	-0,41421	0,00000	4,051E-04	-0,70699	0,29277	1,641E-07	-4,103E-08
10	-0,70699	0,29277	-0,41421	0,00000	1,187E-04	-0,70707	0,29286	1,408E-08	-3,520E-09
11	-0,70707	0,29286	-0,41421	0,00000	3,475E-05	-0,70710	0,29288	1,208E-09	-3,020E-10
12	-0,70710	0,29288	-0,41421	0,00000					

Techniques d'optimisation

3.3.4 Exemple

Exemple

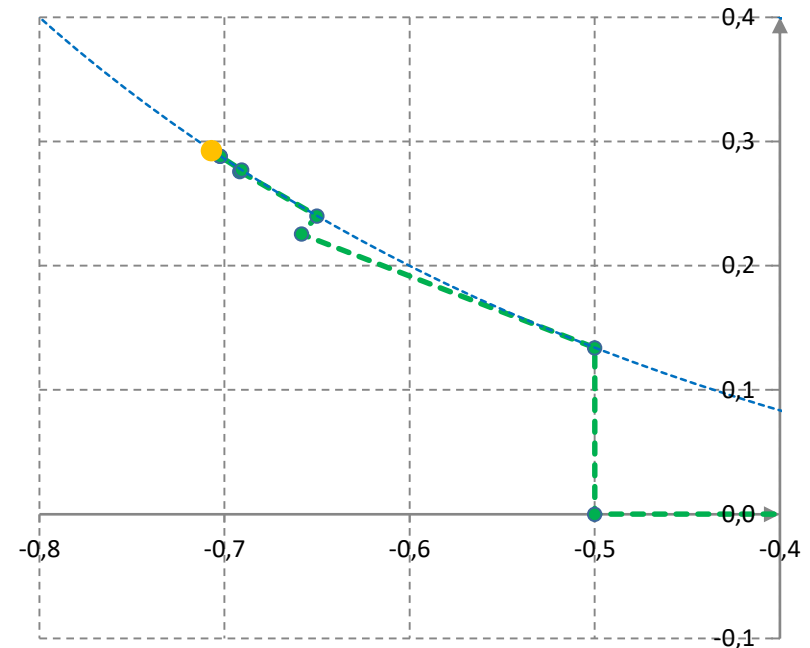
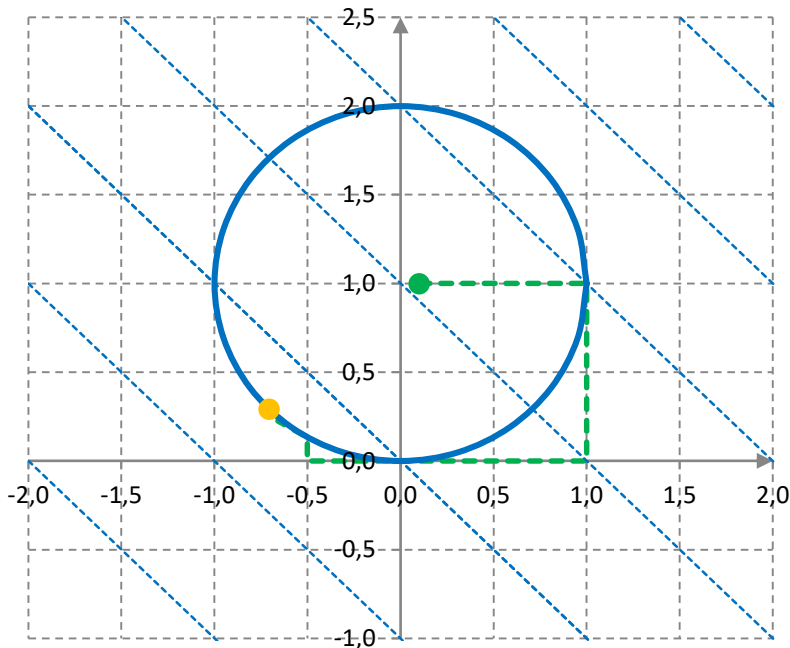
Minimisation de $f(x) = x_1 + x_2$ sous $c(x) = x_1^2 + (x_2 - 1)^2 - 1 = 0$

Point initial

Point initial admissible

Solution

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0.1 \\ 1 \end{pmatrix} \xrightarrow{\text{restauration initiale}} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \xrightarrow{\text{itérations}} \begin{pmatrix} x_1^* \\ x_2^* \end{pmatrix} = \begin{pmatrix} -1/\sqrt{2} \approx -0.70711 \\ 1 - 1/\sqrt{2} \approx 0.29289 \end{pmatrix}$$



Sommaire

1. [Bases théoriques](#)
2. [Optimisation sans contraintes](#)
3. **Optimisation avec contraintes**
 - 3.1 [Simplexe](#)
 - 3.2 [Point intérieur](#)
 - 3.3 [Gradient projeté](#)
 - 3.4 **Lagrangien augmenté**
 - 3.4.1 [Principes](#)
 - 3.4.2 [Pénalisation](#)
 - 3.4.3 [Méthode duale](#)
 - 3.4.4 [Algorithme](#)
 - 3.5 [Programmation quadratique séquentielle](#)
 - 3.6 [Convergence](#)
4. [Optimisation discrète](#)
5. [Optimisation fonctionnelle](#)

3.4.1 Lagrangien augmenté

Problème avec contraintes égalité

$$\min_{x \in \mathbb{R}^n} f(x) \text{ sous } \begin{cases} c_E(x) = 0 \\ c_I(x) \leq 0 \end{cases} \Leftrightarrow \min_{x \in \mathbb{R}^n} f(x) \text{ sous } c(x) = 0 \rightarrow \text{contraintes actives}$$

La difficulté de résolution vient des 2 objectifs antagonistes :

- Minimiser le critère $f(x)$
- Satisfaire les contraintes $c(x)=0$

Méthodes de pénalisation

Les contraintes sont ajoutées à la fonction coût avec une pondération :

- Critère augmenté $\rightarrow$ pondération = pénalisation des contraintes
- Lagrangien $\rightarrow$ pondération = multiplicateurs de Lagrange
- Lagrangien augmenté $\rightarrow$ pondération = pénalisation + multiplicateurs

$\rightarrow$ On se ramène à un **problème sans contraintes** plus simple

Les difficultés viennent du réglage de la pondération :

- Le problème pénalisé sans contraintes doit être équivalent au problème avec contraintes.
- Le problème pénalisé est mal conditionné lorsque la pénalisation est grande.

3.4.2 Pénalisation

- ☐ [Critère augmenté](#)
- ☐ [Pénalisation quadratique](#)
- ☐ [Pénalisation exacte](#)
- ☐ [Mise en œuvre](#)
- ☐ [Lagrangien augmenté](#)

3.4.2 Critère augmenté

Problème avec contraintes égalité

$$\min_{x \in \mathbb{R}^n} f(x) \text{ sous } \begin{cases} c_E(x) = 0 \\ c_I(x) \leq 0 \end{cases} \Leftrightarrow \min_{x \in \mathbb{R}^n} f(x) \text{ sous } c(x) = 0 \rightarrow \text{contraintes actives}$$

On note x^* la solution du problème avec contraintes.

Critère augmenté

On ajoute au critère un terme positif fonction de la violation des contraintes avec un coefficient de pénalisation $\rho > 0 \rightarrow 2$ méthodes usuelles de pénalisation

- **Pénalisation en norme 2** (pénalisation quadratique)

$$f_\rho(x) = f(x) + \frac{1}{2} \rho \left(\|c_E(x)\|_2^2 + \|\max(0, c_I(x))\|_2^2 \right) \Leftrightarrow f_\rho(x) = f(x) + \frac{1}{2} \rho \|c(x)\|_2^2$$

- **Pénalisation en norme 1**

$$f_\rho(x) = f(x) + \rho \left(\|c_E(x)\|_1 + \|\max(0, c_I(x))\|_1 \right) \Leftrightarrow f_\rho(x) = f(x) + \rho \|c(x)\|_1$$

Problème sans contraintes

$$\min_{x \in \mathbb{R}^n} f_\rho(x) \rightarrow \text{solution } x_\rho$$

3.4.2 Pénalisation quadratique

Problème pénalisé l_2

$$\min_{x \in \mathbb{R}^n} f_p(x) \quad \text{avec} \quad f_p(x) = f(x) + \frac{1}{2} \rho \left(\|c_E(x)\|_2^2 + \|\max(0, c_I(x))\|_2^2 \right)$$

$$\Leftrightarrow f_p(x) = f(x) + \frac{1}{2} \rho \|c(x)\|_2^2 \quad \rightarrow \text{contraintes actives}$$

Le critère l_2 est différentiable deux fois pour un problème avec contraintes égalité.
 On peut appliquer les algorithmes d'optimisation sans contraintes à base de gradient.

Méthode de résolution

- On résout une suite de problèmes pénalisés avec des valeurs croissantes de la pénalisation ρ .
- Chaque problème $k+1$ est initialisé avec la solution précédente x_k .
- Problème k avec pénalisation ρ_k : $\min_{x \in \mathbb{R}^n} f_{\rho_k}(x) \rightarrow \text{solution } x_k$
- Il faut vérifier que la suite des solutions x_k converge vers la solution x^* du problème initial

$$\lim_{k \rightarrow \infty} x_k = x^* \quad \text{si} \quad \lim_{k \rightarrow \infty} \rho_k = +\infty$$

$\rightarrow$ 2 résultats de convergence selon que x_k est une solution exacte ou approchée

3.4.2 Pénalisation quadratique

Problème pénalisé l_2

- Problème avec contraintes : $\min_{x \in \mathbb{R}^n} f(x)$ sous $c(x) = 0 \rightarrow$ solution x^*
- Problème k avec pénalisation ρ_k : $\min_{x \in \mathbb{R}^n} f_{\rho_k}(x) \rightarrow$ solution x_k
 $\lim_{k \rightarrow \infty} \rho_k = +\infty \quad \lim_{k \rightarrow \infty} x_k = x_\infty$

Convergence

- Si x_k est le **minimum global exact**, alors $\lim_{k \rightarrow \infty} x_k = x^*$
- Si x_k est un **minimum local approché** : $\|\nabla f_{\rho_k}(x_k)\| \leq \varepsilon_k$ avec $\lim_{k \rightarrow \infty} \varepsilon_k = 0$
 $\rightarrow$ précision de résolution ε_k décroissante

alors la limite x_∞ est : - soit un point non admissible qui minimise $\|c(x)\|_2^2$
 - soit un point x^* vérifiant les conditions KKT du problème initial

On a dans ce 2^{ème} cas : $\begin{cases} \lim_{k \rightarrow \infty} x_k = x^* \\ \lim_{k \rightarrow \infty} \rho_k c(x_k) = \lambda^* \end{cases} \rightarrow$ minimum local
 $\rightarrow$ multiplicateurs des contraintes actives

La solution exacte x^* n'est obtenue qu'à la limite lorsque la pénalisation ρ tend vers l'infini.

3.4.2 Pénalisation quadratique

Eléments de la démonstration

$$f_{\rho}(x) = f(x) + \frac{1}{2} \rho \|c(x)\|^2 \quad \Rightarrow \quad \nabla f_{\rho}(x) = \nabla f(x) + \rho \nabla c(x) c(x)$$

- Critère d'arrêt sur x_k : $\|\nabla f_{\rho_k}(x_k)\| \leq \varepsilon_k$

$$\|\nabla f_{\rho_k}(x_k)\| = \|\nabla f(x_k) + \rho_k \nabla c(x_k) c(x_k)\| \geq \|\rho_k \nabla c(x_k) c(x_k)\| - \|\nabla f(x_k)\| \quad \text{car } \|a+b\| \geq \|a\| - \|b\|$$

$$\Rightarrow \|\rho_k \nabla c(x_k) c(x_k)\| \leq \varepsilon_k + \|\nabla f(x_k)\| \quad \Rightarrow \quad \|\nabla c(x_k) c(x_k)\| \leq \frac{\varepsilon_k + \|\nabla f(x_k)\|}{\rho_k} \xrightarrow{k \rightarrow \infty} 0$$

$$\Rightarrow \begin{cases} \nabla c(x^*) = 0 & \rightarrow \min \|c(x)\| \\ \text{ou} & \text{si les gradients sont linéairement indépendants} \\ c(x^*) = 0 & \rightarrow \text{admissible} \end{cases}$$

- Multiplicateurs de Lagrange

$$\|\nabla f_{\rho_k}(x_k)\| = \|\nabla f(x_k) + \rho_k \nabla c(x_k) c(x_k)\| \leq \varepsilon_k \xrightarrow{k \rightarrow \infty} 0$$

$$\rightarrow \nabla f(x^*) + \nabla c(x^*) \lambda^* = 0 \quad \text{avec} \quad \lambda^* = \lim_{k \rightarrow \infty} \rho_k c(x_k)$$

3.4.2 Pénalisation exacte

Problème pénalisé l_1

$$\min_{x \in \mathbb{R}^n} f_\rho(x) \quad \text{avec} \quad f_\rho(x) = f(x) + \rho \left(\|c_E(x)\|_1 + \|\max(0, c_I(x))\|_1 \right)$$

$$\Leftrightarrow f_\rho(x) = f(x) + \rho \|c(x)\|_1 \quad \rightarrow \text{contraintes actives}$$

Le critère l_1 n'est pas différentiable.

Méthode de résolution

- On résout une suite de problèmes pénalisés avec des valeurs croissantes de la pénalisation ρ .
- Chaque problème $k+1$ est initialisé avec la solution précédente x_k .
- Problème k avec pénalisation ρ_k : $\min_{x \in \mathbb{R}^n} f_{\rho_k}(x) \rightarrow \text{solution } x_k$

Convergence

- Si $\rho > \rho^* = \|\lambda^*\|_\infty = \max |\lambda_i|$ alors **x^* est un minimum local de f_ρ** avec la pénalisation l_1 .
- La pénalisation l_1 est **exacte** si ρ est supérieur au plus grand multiplicateur.
 → ne nécessite pas d'augmenter indéfiniment ρ pour obtenir la solution exacte x^*

3.4.2 Mise en oeuvre

Méthodes avec critère augmenté

- Type de pénalisation
 - Pénalisation l_2 → différentiable, mais nécessite une pénalisation forte pour approcher x^*
 - Pénalisation l_1 → exacte, mais non différentiable
- Réglage de la pénalisation
 - Trop faible → risque de divergence (pas de minimum du problème pénalisé)
 - Trop forte → mauvais conditionnement, difficultés numériques
- Utilisation du critère augmenté
 - Difficultés pratiques si l'on veut obtenir une bonne précision sur la solution x^*
 - Le critère augmenté peut servir de fonction mérite dans le cadre d'autres algorithmes pour évaluer la progression vers l'optimum.

Méthodes avec lagrangien augmenté

On cherche à se ramener à une suite de problèmes sans contraintes

- en conservant un critère différentiable
- en évitant le mauvais conditionnement dû à une pénalisation trop forte
 - utilisation des multiplicateurs de Lagrange pour réduire la pénalisation
 - **méthode duale** ou **lagrangienne**

3.4.2 Lagrangien augmenté

Problème pénalisé I_2

- La méthode de pénalisation consiste à minimiser le critère augmenté.

$$\min_{x \in \mathbb{R}^n} f_p(x) = f(x) + \frac{1}{2} \rho \|c(x)\|^2$$

- La convergence est obtenue pour des valeurs croissantes de pénalisation.

$$\lim_{k \rightarrow \infty} \rho_k = +\infty \rightarrow \begin{cases} \lim_{k \rightarrow \infty} x_k = x^* & \rightarrow \text{minimum local} \\ \lim_{k \rightarrow \infty} \rho_k c(x_k) = \lambda^* & \rightarrow \text{multiplicateurs des contraintes actives} \end{cases}$$

- La solution x_k ne respecte qu'approximativement les contraintes : $c(x_k) \approx \frac{\lambda^*}{\rho_k}$
- Pour respecter précisément les contraintes, il faut augmenter fortement la pénalisation.
→ cause de mauvais conditionnement et de difficultés numériques
- On peut appliquer la méthode de pénalisation **au problème équivalent**

$$\min_{x \in \mathbb{R}^n} L(x, \lambda^*) \text{ sous } c(x) = 0 \Leftrightarrow \min_{x \in \mathbb{R}^n} f(x) \text{ sous } c(x) = 0$$

→ si l'on connaît λ^*

$$\rightarrow \min_{x \in \mathbb{R}^n} L_p(x, \lambda^*) = L(x, \lambda^*) + \frac{1}{2} \rho \|c(x)\|_2^2 = f(x) + \lambda^{*T} c(x) + \frac{1}{2} \rho \|c(x)\|^2$$

3.4.2 Lagrangien augmenté

Lagrangien augmenté

- La méthode de lagrangien augmenté consiste à résoudre une suite de problèmes :

$$\min_{x \in \mathbb{R}^n} L_{\rho_k}(x, \lambda_k) = L(x, \lambda_k) + \frac{1}{2} \rho_k \|c(x)\|_2^2 = f(x) + \lambda_k^T c(x) + \frac{1}{2} \rho_k \|c(x)\|_2^2$$

avec ρ_k = valeur de pénalisation du problème k

λ_k = estimation des multiplicateurs pour le problème k

- Si $\lim_{k \rightarrow \infty} \lambda_k = \lambda^*$ et $\lim_{k \rightarrow \infty} \rho_k = +\infty$ les problèmes deviennent équivalents.

$$\left. \begin{array}{l} \min_{x \in \mathbb{R}^n} L(x, \lambda^*) \text{ sous } c(x) = 0 \Leftrightarrow \min_{x \in \mathbb{R}^n} f(x) \text{ sous } c(x) = 0 \\ \Leftrightarrow \min_{x \in \mathbb{R}^n} L_{\rho}(x, \lambda^*) \Leftrightarrow \min_{x \in \mathbb{R}^n} f_{\rho}(x) \end{array} \right\} \rightarrow \lim_{k \rightarrow \infty} x_k = x^*$$

La solution x_k du problème $\min_{x \in \mathbb{R}^n} L_{\rho_k}(x, \lambda_k)$ converge vers la solution x^* du problème initial.

- La solution x_k vérifie également : $\nabla_x L_{\rho_k}(x_k, \lambda_k) = \nabla f(x_k) + \nabla c(x_k)(\lambda_k + \rho_k c(x_k)) = 0$

à comparer à x^* qui vérifie : $\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) + \nabla c(x^*) \lambda^* = 0$

$$\Rightarrow \lim_{k \rightarrow \infty} \lambda_k + \rho_k c(x_k) = \lambda^*$$

(même démonstration que pour le critère augmenté avec pénalisation quadratique)

3.4.2 Lagrangien augmenté

Lagrangien augmenté

- On peut estimer les multiplicateurs à l'itération k .

$$\lim_{k \rightarrow \infty} \lambda_k + \rho_k c(x_k) = \lambda^* \rightarrow \lambda^* \approx \lambda_k + \rho_k c(x_k) \quad \text{pour } \rho_k \text{ assez grand}$$

$$\rightarrow \lambda_{k+1} = \lambda_k + \rho_k c(x_k) \quad \text{pour l'itération } k+1$$

- La valeur des contraintes à l'itération k est : $c(x_k) \approx \frac{\lambda^* - \lambda_k}{\rho_k} \xrightarrow{\lambda_k \rightarrow \lambda^*} 0$

On peut parvenir à respecter les contraintes sans augmenter indéfiniment la pénalisation si λ_k est une bonne estimation des multiplicateurs $\rightarrow$ **méthode duale**

$\rightarrow$ meilleur conditionnement

$\rightarrow$ convergence plus rapide et précise que la méthode du critère augmenté

$\rightarrow$ **méthode de lagrangien augmenté** appelée aussi **méthode des multiplicateurs**

Convergence

Pour ρ assez grand, **la solution x^* du problème initial est un minimum local du problème**

$$\min_{x \in \mathbb{R}^n} L_\rho(x, \lambda^*) = f(x) + \lambda^{*T} c(x) + \frac{1}{2} \rho \|c(x)\|^2 \rightarrow \text{pénalisation **exacte** si on connaît } \lambda^*$$

$\rightarrow$ ne nécessite pas d'augmenter indéfiniment ρ pour obtenir la solution exacte x^*

3.4.3 Méthode duale

- ☐ [Problème dual](#)
- ☐ [Maximisation duale](#)
- ☐ [Méthode d'Uzawa](#)
- ☐ [Méthode de Newton](#)
- ☐ [Méthode de lagrangien augmenté](#)
- ☐ [Exemple](#)

3.4.3 Problème dual

Problème avec contraintes égalité

- Problème primal : $\min_{x \in \mathbb{R}^n} f(x)$ sous $c(x) = 0$ → m contraintes actives
- Lagrangien : $L(x, \lambda) = f(x) + \lambda^T c(x)$ → $x = n$ variables primales
→ $\lambda = m$ variables duales
- Fonction duale : $w(\lambda) = \min_{x \in \mathbb{R}^n} L(x, \lambda)$ → minimisation par rapport à x
- Problème dual : $\max_{\lambda \in \mathbb{R}^m} w(\lambda)$ → maximisation par rapport à λ

Méthode duale (ou lagrangienne)

Une méthode duale consiste à résoudre le problème dual :

$$\max_{\lambda \in \mathbb{R}^m} w(\lambda)$$

- Le problème est sans contraintes, mais la fonction $w(\lambda)$ est coûteuse à évaluer.
- La solution (x^*, λ^*) est un point col du lagrangien : $\forall x, \lambda, L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*)$
→ **existence non garantie** (saut de dualité)

3.4.3 Maximisation duale

Fonction duale

On note $x_L(\lambda)$ la valeur de x qui minimise $L(x, \lambda)$ à λ fixé.

Condition d'ordre 1 : $\min_{x \in \mathbb{R}^n} L(x, \lambda) \text{ en } x = x_L(\lambda) \Rightarrow \boxed{\nabla_x L(x_L(\lambda), \lambda) = 0}$

- Fonction w : $w(\lambda) = \min_{x \in \mathbb{R}^n} L(x, \lambda) = L(x_L(\lambda), \lambda) \rightarrow$ fonction composée

- Gradient de w : $\nabla w(\lambda) = \nabla_{x_L}(\lambda) \cdot \nabla_x L(x_L(\lambda), \lambda) + \nabla_\lambda L(x_L(\lambda), \lambda)$
 $\quad \quad \quad = 0 \quad \quad \quad = c(x_L(\lambda))$

$$\Rightarrow \boxed{\nabla w(\lambda) = c(x_L(\lambda))}$$

- Hessien de w : $\nabla^2 w(\lambda) = \nabla_\lambda c(x_L(\lambda)) = \nabla_{x_L}(\lambda) \nabla c(x_L(\lambda))$

$$\Rightarrow \boxed{\nabla^2 w(\lambda) = -\nabla c(x_L(\lambda))^T [\nabla_{xx}^2 L(x_L(\lambda), \lambda)]^{-1} \nabla c(x_L(\lambda))}$$

Preuve : on calcule $\nabla x_L(\lambda)$ à partir de $\nabla_x L(x_L(\lambda), \lambda) = 0$

$$\begin{aligned} \nabla_x L(x_L(\lambda), \lambda) = 0, \forall \lambda &\Rightarrow \nabla_\lambda [\nabla_x L(x_L(\lambda), \lambda)] = 0 \Rightarrow \nabla x_L(\lambda) \nabla_{xx}^2 L(x_L(\lambda), \lambda) + \nabla_{\lambda x}^2 L(x_L(\lambda), \lambda) = 0 \\ &\Rightarrow \nabla x_L = -\nabla c^T [\nabla_{xx}^2 L]^{-1} \quad \text{avec } \nabla_\lambda L = c \end{aligned}$$

3.4.3 Maximisation duale

Méthode de maximisation

- La fonction duale est **coûteuse** à évaluer.
 $w(\lambda) = \min_{x \in \mathbb{R}^n} L(x, \lambda) = L(x_L(\lambda), \lambda) \rightarrow$ minimisation en chaque valeur λ pour obtenir $x_L(\lambda)$
 $\rightarrow$ on ne peut pas directement maximiser w
- On réalise à chaque itération :
 - une minimisation par rapport à x : $(x_k, \lambda_k) \rightarrow (x_{k+1}, \lambda_k)$
 - une maximisation par rapport à λ : $(x_{k+1}, \lambda_k) \rightarrow (x_{k+1}, \lambda_{k+1})$

Déplacement

- Le déplacement sur x est construit en minimisant $L(x, \lambda_k)$.
 $\min_{x \in \mathbb{R}^n} L(x, \lambda_k) \rightarrow x_{k+1} = x_L(\lambda_k) \rightarrow$ minimisation exacte ou approchée
- Le déplacement sur λ est construit à partir du gradient et éventuellement du hessien de w .

$$\begin{cases} \nabla w(\lambda_k) = c(x_{k+1}) \\ \nabla^2 w(\lambda_k) = -\nabla c(x_{k+1})^T [\nabla_{xx}^2 L(x_{k+1}, \lambda_k)]^{-1} \nabla c(x_{k+1}) \end{cases} \rightarrow$$
 minimisation approchée (**pas fixé**)
- Méthodes de déplacement sur λ : Uzawa, lagrangien augmenté, Newton

3.4.3 Méthode d'Uzawa

Principe

Le déplacement en λ est construit par une **méthode de plus forte pente**.

$$\lambda_{k+1} = \lambda_k + s_k \nabla w(\lambda_k) \Rightarrow \boxed{\lambda_{k+1} = \lambda_k + s_k c(x_{k+1})}$$

L'efficacité de la méthode dépend du réglage du pas s_k .

Méthode d'Uzawa (1958)

- On choisit un **pas constant** : $s_k = s_0$.
- Difficulté :
 - pas s_0 trop grand $\rightarrow$ oscillations autour de la solution
 - pas s_0 trop petit $\rightarrow$ convergence très lente
- La convergence est linéaire (méthode de plus forte pente).
- La vitesse de convergence dépend du conditionnement du hessien de la fonction duale.

$$\nabla^2 w(\lambda^*) = -\nabla c(x^*)^T \left[\nabla_{xx}^2 L(x^*, \lambda^*) \right]^{-1} \nabla c(x^*)$$

Améliorations

- On peut régler le pas à chaque itération par une recherche linéaire suivant $\nabla w(\lambda_k)$.
- On peut appliquer une **méthode de Newton** au problème dual en utilisant $\nabla^2 w(\lambda_k)$.
 - $\rightarrow$ méthodes nécessitant une globalisation pour contrôler la convergence
 - $\rightarrow$ évaluations coûteuses de $w(\lambda)$

3.4.3 Méthode de Newton

Principe

L'itération de Newton sur le problème dual est définie par :

$$\lambda_{k+1} = \lambda_k - [\nabla^2 w(\lambda_k)]^{-1} \nabla w(\lambda_k) \quad \text{avec} \quad \begin{cases} \nabla w(\lambda_k) = c(x_{k+1}) \\ \nabla^2 w(\lambda_k) = -\nabla c(x_{k+1})^T [\nabla_{xx}^2 L(x_{k+1}, \lambda_k)]^{-1} \nabla c(x_{k+1}) \end{cases}$$

- On applique une **méthode de quasi-Newton** au problème primal : $\min_{x \in \mathbb{R}^n} L(x, \lambda_k)$

pour obtenir une approximation de l'inverse du hessien de L : $H_k \approx \nabla_{xx}^2 L(x_{k+1}, \lambda_k)^{-1}$

- On applique ensuite une **méthode de Newton** au problème dual.

$$\lambda_{k+1} = \lambda_k - [\nabla^2 w(\lambda_k)]^{-1} \nabla w(\lambda_k) \Rightarrow \boxed{\lambda_{k+1} = \lambda_k + [\nabla c(x_{k+1})^T H_k \nabla c(x_{k+1})]^{-1} c(x_{k+1})}$$

Convergence

- La convergence est superlinéaire (quasi-Newton) ou quadratique (Newton).
- Une globalisation est nécessaire pour vérifier la solution de Newton.
→ évaluations supplémentaires de $w(\lambda)$ pouvant être coûteuses
- Le point col n'existe pas nécessairement si le problème n'est pas convexe (saut de dualité)
→ création d'un point col par pénalisation

Techniques d'optimisation

3.4.3 Exemple

Exemple

- Minimisation de $f(x) = x_1 + x_2$ sous $c(x) = x_1^2 + (x_2 - 1)^2 - 2 = 0$
- Lagrangien : $L(x, \lambda) = x_1 + x_2 + \lambda(x_1^2 + (x_2 - 1)^2 - 2)$
- Fonction duale : $w(\lambda) = \min_{x \in \mathbb{R}^2} L(x, \lambda) = 1 - 2\lambda - \frac{1}{2\lambda}$ pour $x_L(\lambda) = \begin{pmatrix} -\frac{1}{2\lambda} \\ -\frac{1}{2\lambda} + 1 \end{pmatrix}$
- Problème dual : $\max_{\lambda \in \mathbb{R}} w(\lambda) \Rightarrow \boxed{\lambda^* = \frac{1}{2} \rightarrow x^* = \begin{pmatrix} -1 \\ 0 \end{pmatrix}} \rightarrow \text{point col}$
- Gradient : $\nabla w(\lambda) = c(x_L(\lambda)) = \frac{1}{2\lambda^2} - 2$
- Hessien : $\nabla^2 w(\lambda) = -\nabla c(x_L(\lambda))^T [\nabla_{xx}^2 L(x_L(\lambda), \lambda)]^{-1} \nabla c(x_L(\lambda)) = -\frac{1}{\lambda^3}$
 avec $\nabla c(x_L(\lambda)) = -\frac{1}{\lambda} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\nabla_{xx}^2 L(x_L(\lambda), \lambda) = 2\lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

3.4.3 Exemple

Méthode d'Uzawa

- Minimisation de $f(x) = x_1 + x_2$ sous $c(x) = x_1^2 + (x_2 - 1)^2 - 2 = 0$
- Itération à pas s fixé :
$$x_{k+1} = \begin{pmatrix} -\frac{1}{2\lambda_k} \\ -\frac{1}{2\lambda_k} + 1 \end{pmatrix} \quad \lambda_{k+1} = \lambda_k + s \left(\frac{1}{2\lambda_k^2} - 2 \right)$$
- Influence du pas** : $s=0,1$ ou $s=0,2$ → oscillations autour de la solution

Uzawa pas $s = 0,1$

Itération	x_1	x_2	λ	s
1	-0,10000	1,00000	1,00000	0,1
2	-0,50000	0,50000	0,85000	0,1
3	-0,58824	0,41176	0,71920	0,1
4	-0,69521	0,30479	0,61587	0,1
5	-0,81186	0,18814	0,54769	0,1
6	-0,91292	0,08708	0,51438	0,1
7	-0,97205	0,02795	0,50335	0,1
8	-0,99334	0,00666	0,50070	0,1
9	-0,99861	0,00139	0,50014	0,1
10	-0,99972	0,00028	0,50003	0,1
11	-0,99994	5,63E-05	0,50001	0,1
12	-0,99999	1,13E-05	0,50000	0,1

Uzawa pas $s = 0,2$

Itération	x_1	x_2	λ	s
1	-0,10000	1,00000	1,00000	0,2
2	-0,50000	0,50000	0,70000	0,2
3	-0,71429	0,28571	0,50408	0,2
4	-0,99190	0,00810	0,49763	0,2
5	-1,00476	-0,00476	0,50145	0,2
6	-0,99711	0,00289	0,49914	0,2
7	-1,00172	-0,00172	0,50052	0,2
8	-0,99896	0,00104	0,49969	0,2
9	-1,00062	-0,00062	0,50019	0,2
10	-0,99963	0,00037	0,49989	0,2
11	-1,00022	-2,24E-04	0,50007	0,2
12	-0,99987	1,34E-04	0,49996	0,2

3.4.3 Exemple

Comparaison Uzawa / Newton

- Minimisation de $f(x) = x_1 + x_2$ sous $c(x) = x_1^2 + (x_2 - 1)^2 - 2 = 0$

- Itération k :
$$x_{k+1} = \begin{pmatrix} -\frac{1}{2\lambda_k} \\ -\frac{1}{2\lambda_k} + 1 \end{pmatrix}$$

$$\lambda_{k+1} = \lambda_k + s \left(\frac{1}{2\lambda_k^2} - 2 \right) \rightarrow \text{Uzawa}$$

$$\lambda_{k+1} = \lambda_k + s\lambda_k^3 \left(\frac{1}{2\lambda_k^2} - 2 \right) \rightarrow \text{Newton}$$
 Globalisation $\rightarrow$ pas s

Uzawa pas s = 0,1

Itération	x_1	x_2	λ	s
1	-0,10000	1,00000	1,00000	0,1
2	-0,50000	0,50000	0,85000	0,1
3	-0,58824	0,41176	0,71920	0,1
4	-0,69521	0,30479	0,61587	0,1
5	-0,81186	0,18814	0,54769	0,1
6	-0,91292	0,08708	0,51438	0,1
7	-0,97205	0,02795	0,50335	0,1
8	-0,99334	0,00666	0,50070	0,1
9	-0,99861	0,00139	0,50014	0,1
10	-0,99972	0,00028	0,50003	0,1
11	-0,99994	5,63E-05	0,50001	0,1
12	-0,99999	1,13E-05	0,50000	0,1

Newton pas s = 1

Itération	x_1	x_2	λ	s
1	-0,10000	1,00000	1,00000	0,25
2	-0,50000	0,50000	0,62500	0,50
3	-0,80000	0,20000	0,53711	1,00
4	-0,93091	0,06909	0,49577	1,00
5	-1,00854	-0,00854	0,49995	1,00
6	-1,00011	-0,00011	0,50000	1,00
7	-1,00000	-1,72E-08	0,50000	1,00
8	-1,00000	0,00E+00	0,50000	1,00

3.4.3 Méthode de lagrangien augmenté

Principe

La méthode de lagrangien augmenté est une **méthode duale appliquée au critère augmenté**.

- Critère augmenté f_ρ :
$$f_\rho(x) = f(x) + \frac{1}{2} \rho \|c(x)\|_2^2$$
- Lagrangien augmenté L_ρ :
$$L_\rho(x, \lambda) = f(x) + \lambda^T c(x) + \frac{1}{2} \rho \|c(x)\|_2^2$$
- **Fonction duale augmentée w_ρ** :
$$w_\rho(\lambda) = \min_{x \in \mathbb{R}^n} L_\rho(x, \lambda)$$

Le problème dual devient :
$$\max_{\lambda \in \mathbb{R}^m} w_\rho(\lambda)$$

Fonction duale augmentée

Le gradient et le hessien de w_ρ sont identiques à ceux de w en remplaçant L par L_ρ .

$$\begin{cases} \nabla w_\rho(\lambda) = c(x_{L_\rho}(\lambda)) \\ \nabla^2 w_\rho(\lambda) = -\nabla c(x_{L_\rho}(\lambda))^T [\nabla_{xx}^2 L_\rho(x_{L_\rho}(\lambda), \lambda)]^{-1} \nabla c(x_{L_\rho}(\lambda)) \end{cases} \quad \text{avec} \quad \min_{x \in \mathbb{R}^n} L_\rho(x, \lambda) \quad \text{en} \quad x = x_{L_\rho}(\lambda)$$

On réalise à chaque itération :

- une minimisation de L_ρ par rapport à x
- un déplacement en λ à partir du gradient et du hessien de w_ρ

3.4.3 Méthode de lagrangien augmenté

Déplacement

- Le déplacement sur x est construit en minimisant $L_\rho(x, \lambda_k)$ avec une précision donnée.

$$\min_{x \in \mathbb{R}^n} L_\rho(x, \lambda_k) \rightarrow x_{k+1} = x_{L_\rho}(\lambda_k)$$

- Le déplacement sur λ est construit par une méthode de plus forte pente.

$$\lambda_{k+1} = \lambda_k + s_k \nabla w_\rho(\lambda_k) \Rightarrow \lambda_{k+1} = \lambda_k + s_k c(x_{k+1})$$

Le pas s_k est choisi à partir de la propriété de convergence : $\lim_{k \rightarrow \infty} \lambda_k + \rho_k c(x_k) = \lambda^*$

$$s_k = \rho_k \rightarrow \boxed{\lambda_{k+1} = \lambda_k + \rho_k c(x_{k+1})}$$

Convergence

- La convergence est linéaire (méthode de plus forte pente).
- La vitesse de convergence dépend du conditionnement des hessiens de L_ρ (primal) et w_ρ (dual).

$$\nabla_{xx}^2 L_\rho = \nabla_{xx}^2 L + \rho \nabla c \cdot \nabla c^T \xrightarrow{\rho \rightarrow \infty} \rho \nabla c \cdot \nabla c^T \rightarrow \text{mal conditionné si } \rho \text{ grand}$$

$$\nabla^2 w_\rho = -\nabla c^T [\nabla_{xx}^2 L_\rho]^{-1} \nabla c \xrightarrow{\rho \rightarrow \infty} -\rho^{-1} I \rightarrow \text{bien conditionné si } \rho \text{ grand}$$

- On peut appliquer une méthode de Newton au dual avec $H_k \approx [\nabla_{xx}^2 L_\rho]^{-1}$ (quasi-Newton primal)
 → convergence superlinéaire + globalisation nécessaire
- La pénalisation peut **créer un point col** qui n'existe pas dans le problème initial.

3.4.3 Exemple

Exemple

- Minimisation de $f(x) = -x_1x_2$ sous $c(x) = x_1 + x_2 - 1 = 0$
- Lagrangien : $L(x, \lambda) = -x_1x_2 + \lambda(x_1 + x_2 - 1)$
- Solution KKT : $\lambda^* = \frac{1}{2}, \quad x_1^* = x_2^* = \frac{1}{2} \rightarrow L(x^*, \lambda^*) = -\frac{1}{4}$
- Fonction duale : $w(\lambda) = \min_{x \in \mathbb{R}^2} L(x, \lambda) = -\infty$ en prenant $\begin{cases} x_1 = -\lambda \\ x_2 \rightarrow \pm\infty \end{cases}$ (signe de $-\lambda$)

La fonction duale vaut $-\infty$ pour toute valeur de λ .

- Problème dual : $\max_{\lambda \in \mathbb{R}} w(\lambda) = -\infty \rightarrow < -\frac{1}{4}$
- Le lagrangien n'admet **pas de point col** (problème non convexe).
 Le problème non pénalisé présente un **saut de dualité**.
 → On ne peut pas appliquer une méthode duale (Uzawa) sur ce problème.
 → On peut créer un point col par pénalisation (méthode de lagrangien augmenté).

3.4.3 Exemple

Point col

- Problème pénalisé : $f_\rho(x) = -x_1x_2 + \frac{1}{2}\rho(x_1 + x_2 - 1)^2$ sous $c(x) = x_1 + x_2 - 1 = 0$
- Lagrangien augmenté : $L_\rho(x, \lambda) = -x_1x_2 + \frac{1}{2}\rho(x_1 + x_2 - 1)^2 + \lambda(x_1 + x_2 - 1)$
- Fonction duale : $w_\rho(\lambda) = \min_{x \in \mathbb{R}^2} L_\rho(x, \lambda)$
- Minimisation de L_ρ par rapport à x

Ordre 1 : $\nabla_x L_\rho = 0 \Rightarrow \begin{cases} -x_2 + \rho(x_1 + x_2 - 1) + \lambda = 0 \\ -x_1 + \rho(x_1 + x_2 - 1) + \lambda = 0 \end{cases} \Rightarrow \boxed{x_1(\lambda) = x_2(\lambda) = \frac{\lambda - \rho}{1 - 2\rho}}$

Ordre 2 : $\nabla_{xx}^2 L_\rho > 0 \Rightarrow \begin{pmatrix} \rho & \rho - 1 \\ \rho - 1 & \rho \end{pmatrix} > 0$

Valeurs propres : $(\rho - \sigma)^2 - (\rho - 1)^2 = 0 \Rightarrow \begin{cases} \sigma_1 = 1 \\ \sigma_2 = 2\rho - 1 \end{cases} \rightarrow \boxed{L_\rho \text{ convexe si } \rho > \frac{1}{2}}$

- On obtient si $\rho > \frac{1}{2}$: $w_\rho(\lambda) = -\left(\frac{\lambda - \rho}{1 - 2\rho}\right)^2 + \frac{1}{2}\rho\left(\frac{2\lambda - 1}{1 - 2\rho}\right)^2 + \lambda\frac{2\lambda - 1}{1 - 2\rho}$

3.4.3 Exemple

Point col

- Fonction duale : $w_\rho(\lambda) = \frac{1}{(1-2\rho)^2} \left[-(\lambda - \rho)^2 + \frac{1}{2}\rho(2\lambda - 1)^2 + (1-2\rho)(2\lambda^2 - \lambda) \right]$

- Problème dual : $\max_{\lambda \in \mathbb{R}} w_\rho(\lambda)$

$$\nabla w_\rho = 0 \Rightarrow -2(\lambda - \rho) + 2\rho(2\lambda - 1) + (1-2\rho)(4\lambda - 1) = 0$$

Solution : $\lambda^* = \frac{1}{2} \Rightarrow w_\rho(\lambda^*) = -\frac{1}{4} \text{ et } x_1(\lambda^*) = x_2(\lambda^*) = \frac{1}{2}$

$$w_\rho(\lambda^*) = L_\rho(x^*, \lambda^*) = -\frac{1}{4}$$

On retrouve la solution du problème primal **sans saut de dualité** → point col

- La pénalisation permet de **créer un point col** pour $\rho > \frac{1}{2}$ en rendant le problème convexe.
 → On peut appliquer une méthode duale au problème pénalisé ($\neq$ problème initial).
 → La méthode de lagrangien augmenté élargit le domaine d'application des méthodes duales.

3.4.4 Algorithme

- ❑ [Algorithme de lagrangien augmenté](#)
- ❑ [Exemple](#)

3.4.4 Algorithme

Méthode de lagrangien augmenté (ou méthode des multiplicateurs)

Les principales étapes d'une itération de lagrangien augmenté sont

- minimiser le lagrangien augmenté
- mettre à jour les paramètres de réglage

Minimisation du lagrangien augmenté

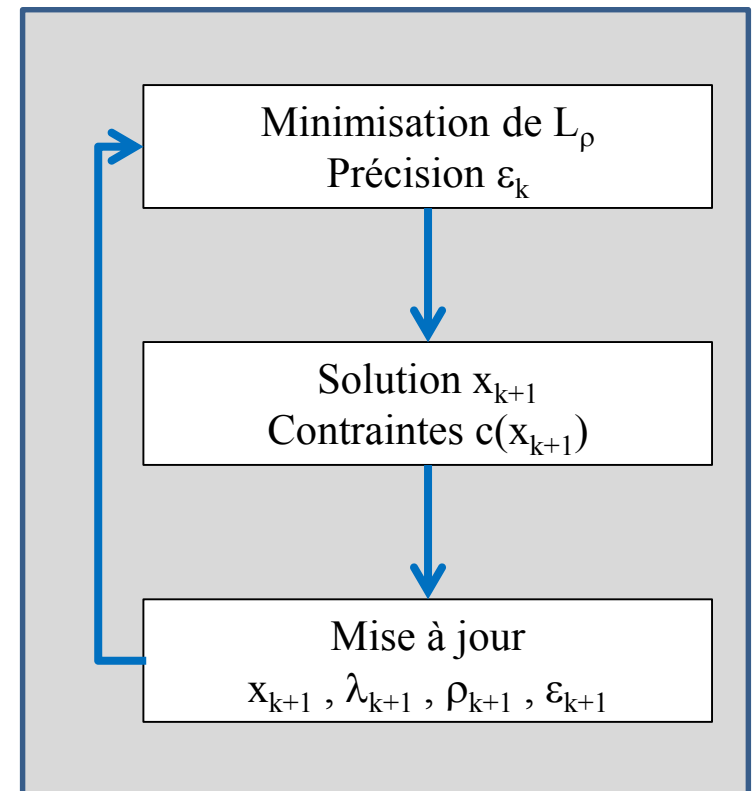
- Méthode de quasi-Newton
- Recherche linéaire ou région de confiance
- **Précision d'arrêt sur gradient**

Paramètres de réglage

- Multiplicateurs
- Pénalisation
- Précisions (gradient, contraintes)

Principales difficultés

- Précision contraintes → pénalisation forte
- Conditionnement → pénalisation faible
→ convergence précise difficile
→ réglages à adapter au cas par cas



3.4.4 Algorithme

Méthode de lagrangien augmenté (ou méthode des multiplicateurs)

- Réglages à l'itération k :
 - multiplicateurs λ_k
 - pénalisation ρ_k
 - précision de résolution ε_k
 - précision sur les contraintes η_k
- Minimisation : $\min_{x \in \mathbb{R}^n} L_{\rho_k}(x, \lambda_k) \rightarrow \text{critère d'arrêt } \|\nabla_x L_{\rho_k}(x_{k+1}, \lambda_k)\| \leq \varepsilon_k$
 initialisation $x_k \rightarrow$ solution $x_{k+1} \rightarrow$ valeurs des contraintes $= c(x_{k+1})$
- Mise à jour des réglages à l'itération $k+1$ en fonction du respect des contraintes
 - Si $\|c(x_{k+1})\| < \eta_k$ $\rightarrow$ **mise à jour des multiplicateurs** $\lambda_{k+1} = \lambda_k + \rho_k c(x_k)$
 (contraintes bien respectées) $\rightarrow$ résolution plus précise $\varepsilon_{k+1} < \varepsilon_k$, $\eta_{k+1} < \eta_k$
 - Si $\|c(x_{k+1})\| > \eta_k$ $\rightarrow$ **augmentation de la pénalisation** $\rho_{k+1} > \rho_k \rightarrow \times 10$
 (contraintes mal respectées) $\rightarrow$ résolution moins précise $\varepsilon_{k+1} > \varepsilon_k$, $\eta_{k+1} > \eta_k$

3.4.4 Algorithme

Estimation des multiplicateurs

- La solution (x^*, λ^*) doit vérifier la condition KKT d'ordre 1.

$$\nabla_x L(x^*, \lambda^*) = 0 \quad \Leftrightarrow \quad \nabla f(x^*) + \nabla c(x^*) \lambda^* = 0$$

- On cherche au point x_0 le multiplicateur qui approche « au mieux » la condition KKT.

$$\min_{\lambda} \|\nabla_x L(x_0, \lambda)\|^2 \quad \Leftrightarrow \quad \min_{\lambda} \|\nabla c(x_0) \lambda + \nabla f(x_0)\|^2$$

- Il s'agit d'un problème de moindres carrés de la forme : $\min_{\lambda} \|A\lambda - b\|^2$ avec $\begin{cases} A = \nabla c(x_0) \\ b = -\nabla f(x_0) \end{cases}$

La solution λ_{MC} vérifie les équations normales : $A^T A \lambda_{MC} = A^T b$

$$\lambda_{MC} = -[\nabla c(x_0)^T \nabla c(x_0)]^{-1} \nabla c(x_0)^T \nabla f(x_0) \rightarrow \text{multiplicateurs « des moindres carrés »}$$

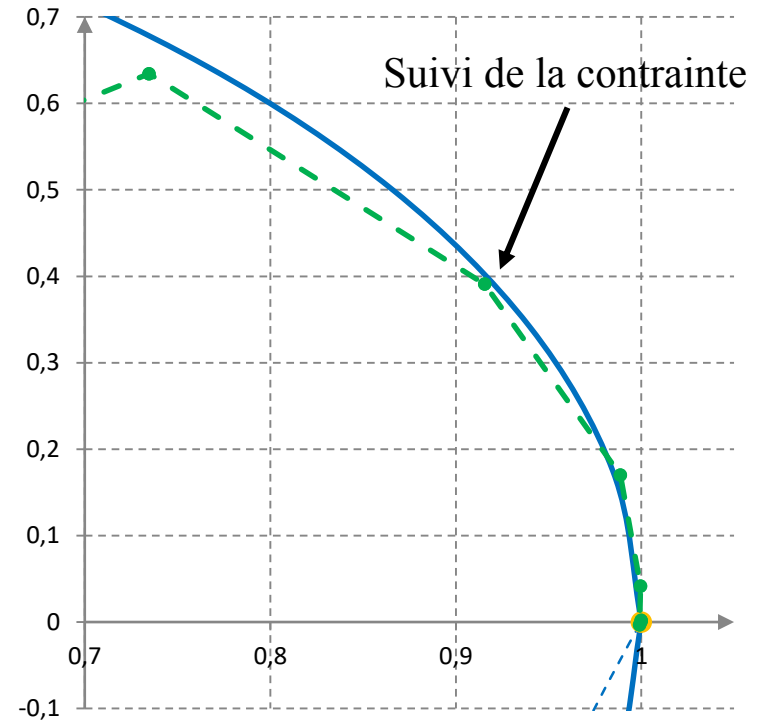
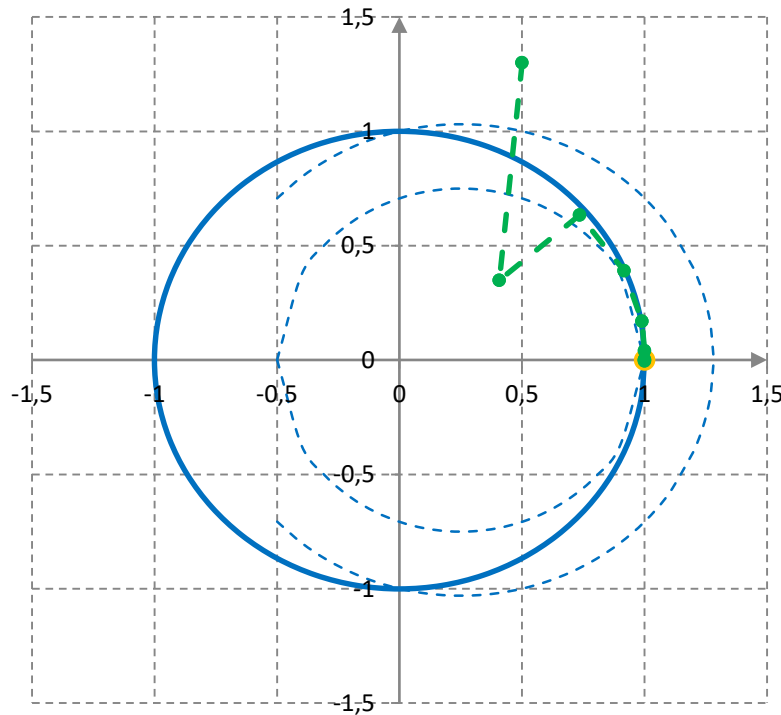
- L'estimation des multiplicateurs par moindres carrés est utile :
 - à la première itération pour **initialiser** λ_0
 - au cours des itérations pour réinitialiser λ_k (si nécessaire).

Techniques d'optimisation

3.4.4 Exemple

Exemple

- Minimisation de $f(x) = 2(x_1^2 + x_2^2 - 1) - x_1$ sous $c(x) = x_1^2 + x_2^2 - 1 = 0$
- Point initial : $x = \begin{pmatrix} 0.5 \\ 1.3 \end{pmatrix}$, $\lambda = 0$ Solution : $x^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\lambda = -\frac{3}{2}$



Techniques d'optimisation

3.4.4 Exemple

Exemple

- Minimisation de $f(x) = 2(x_1^2 + x_2^2 - 1) - x_1$ sous $c(x) = x_1^2 + x_2^2 - 1 = 0$
- Point initial : $x = \begin{pmatrix} 0.5 \\ 1.3 \end{pmatrix}$, $\lambda = 0$ Solution : $x^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\lambda^* = -\frac{3}{2}$

Itération	x_1	x_2	λ	ρ	$c(x)$	$\ \nabla L_\rho(x, \lambda)\ $	Newton
1	0,50000	1,30000	0,00000	1	-0,71238	0,90050	1
2	0,40707	0,34917	-0,71238	10	-0,05788	0,90016	1
3	0,73467	0,63433	-1,29122	10	-0,00905	0,50091	2
4	0,91556	0,39077	-1,38175	10	0,00635	0,41807	2
5	0,98869	0,16985	-1,38175	100	0,00188	0,62061	2
6	0,99953	0,04158	-1,30283	100	-0,00188	0,01728	2
7	0,99905	-0,00320	-1,49103	100	-0,00009	0,00172	1
8	0,99995	0,00171	-1,50003	100	2,06E-06	0,00057	3
9	1,00000	0,00045	-1,50003	100	1,85E-06	0,00031	

↓
 pénalisation

↓
 précision
 contrainte

↓
 précision
 résolution

↓
 itérations
 Newton