Algorithme SQP/OQS

I Orientation (Chap 14)

— On cherche à résoudre numériquement

$$(P_{EI}) \begin{cases} \min f(x) \\ c_E(x) = 0 & \Leftrightarrow \quad c_i(x) = 0, \ \forall i \in E \\ c_I(x) \leq 0 & \Leftrightarrow \quad c_i(x) \leq 0, \ \forall i \in I \end{cases}$$

- $E \cap I = \phi$, $E \cup I = [1:m]$
- $f : \mathbb{E} \to \mathbb{R}$ ($\mathbb{E}$ un espace vectoriel)
- $c : \mathbb{E} \to \mathbb{R}^m$, composantes c_i
- pour $v \in \mathbb{R}^m$, on note $v^\# \in \mathbb{R}$ défini par

$$(v^\#)_i = \begin{cases} v_i & \text{si } i \in E \\ v_i^+ = \max(0, v_i) & \text{si } i \in I \end{cases}$$

alors $c(x)^\# = 0 \Leftrightarrow x$ admissible pour (P_{EI})

— Méthodes de résolution de (P_{EI})

- primales (pénalisation → cours précédent)
- duales (méthodes de dualité → cours précédent)
- primale - duale (SQP / OQS, aujourd'hui)

— OQS = Optimisation Quadratique Successive
SQP = Sequential Quadratic Programming

C'est 1 algorithme du type Newton en $(x, \lambda) \in \mathbb{E} \times \mathbb{R}^m$ (donc primal - dual)

— Il a fallu longtemps pour le montrer :
30 ans (1945 → 1975)

- essai d'étendre le simplexe (PL) à l'ONL
- pénalisation } ont précédé
- dualité

— Difficultés à surmonter

- comprendre comment intervient la courbure des contraintes d'égalité ($\rightarrow$ de $\nabla^2_{xx} \ell$)
- comprendre comment prendre en compte les inégalités.
- accepter une itération + coûteuse que la résolution d'un SL

II L'algorithme local

Beaucoup de manières de le présenter :

- dans le syllabus : il est vu comme un algorithme de Josephy-Newton sur les conditions d'optimalité, interprété comme une inclusion fonctionnelle ou un problème de complémentarité

  (c'est compliqué [beaucoup de notions non vues], mais c'est ce qui conduit au résultat de convergence locale le plus fort)

- ici (comme dans beaucoup de manuels) on suit une procédure "naturelle", une heuristique.

(A) Un algorithme à éviter

→ <u>À ne pas faire</u> : traiter (par l'algorithme de Newton) l'optimalité et l'admissibilité <u>séparément</u>

• $\min f(x) \xrightarrow{\text{Newton}} \begin{cases} \min\limits_d \cancel{f(x)} + \nabla f(x)^T d + \frac{1}{2} d^T \nabla^2 f(x) d \\ x_+ = x + d \end{cases}$

• $c(x) = 0 \longrightarrow \begin{cases} c(x) + c'(x)\, d = 0 \\ x_+ = x + d \end{cases}$

• $\begin{cases} \min f(x) \\ c(x) = 0 \end{cases} \longrightarrow \left[\begin{array}{l} \begin{cases} \min\limits_d \nabla f(x)^T d + \frac{1}{2} d^T \nabla^2 f(x)\, d \\ c(x) + c'(x)\, d = 0 \end{cases} \\ \text{puis } x_+ = x + d \end{array} \right.$ (0)

un algorithme à éviter (il ne converge pas nécessairement localement)

→ <u>Ce qu'il faut faire</u> : appliquer Newton à un système qui prend en compte simultanément l'optimalité et l'admissibilité ⟹

$\boxed{\text{appliquer Newton aux conditions d'optimalité}}$

- naturel : c'est ce que l'on fait en optimisation SANS contrainte
- nouveau : algo en (x, d)
- découvert "seulement" ~1975

(B) L'algorithme SQP / OQS

- On part de CN1 ($\equiv$ KKT) qui est le système en (x, λ)

$$(KKT) \begin{cases} \nabla f(x) + c'(x)^T \lambda = 0 \\ c_E(x) = 0 \\ 0 \leq \lambda_I \perp c_I(x) \leq 0 \end{cases}$$

et on linéarise toutes les fonctions qui y apparaissent (raisonnable ? dangereux pour la convergence ? $\Rightarrow$ il faudra trouver une démonstration nouvelle) et on garde la structure.

- C'est une linéarisation en $(x, \lambda) \in \mathbb{E} \times \mathbb{R}^m$: le nouvel itéré est donc

$$(x_+, \lambda_+) = (x, \lambda) + (d, \mu)$$

où (d, μ) est solution de

$$\begin{cases} \nabla_x \ell(x,\lambda) + \nabla^2_{xx} \ell(x,\lambda)\, d + c'(x)^T \mu = 0 & (1a) \\ c_E(x) + c_E'(x)\, d = 0 & (1b) \\ 0 \leq (\lambda+\mu)_I \perp \left(c_I(x) + c_I'(x)\, d\right) \leq 0 & (1c) \end{cases}$$

- Comment résoudre ce système ?
- Ça doit être les CN1 d'un problème quadratique !

– légère complication : ce n'est pas le même multiplicateur qui apparaît dans (1a) et (1c) ; mais <u>(1a) peut se récrire autrement</u>

$$\lambda^{QP}$$

$$(KKT')\begin{cases} Df(x) + D^2_{xx}\ell(x,\lambda)\, d + c'(x)^T(\lambda+\mu) = 0 \\ c_E(x) + c'_E(x)\, d = 0 \\ 0 \leqslant (\lambda+\mu)_I \perp (c_I(x) + c'_I(x)\, d) \leqslant 0 \end{cases}$$

$\Rightarrow \lambda+\mu =$ multiplicateur associé aux contraintes linéarisées

– (KKT') sont les CN1 du <u>problème quadratique osculateur</u> suivant

$$(PQO)\begin{cases} \min\limits_d \; Df(x)^T d + \frac{1}{2} d^T D^2_{xx}\ell(x,\lambda)\, d \\ \left.\begin{aligned} c_E(x) + c'_E(x)\, d &= 0 \\ c_I(x) + c'_I(x)\, d &\leqslant 0 \end{aligned}\right\} \to \text{mult } \lambda^{PQ} \end{cases}$$

– <u>Remarques</u>

1. le critère du PQO est <u>hybride</u> : $Df(x)$ pour la partie linéaire, $D^2_{xx}\ell(x,\lambda)$ pour la partie quadratique

2. les contraintes du PQO sont obtenues en linéarisant celle de (P_{EI})

3. ce qui manque à (0) par rapport à (PQO) c'est la courbure des contraintes (= dérivées secondes)

$$D^2 f(x) \longrightarrow D^2_{xx}\ell(x,\lambda)$$

4. $\lambda+\mu = \lambda^{PQ}$

- <u>l'algorithme SQP / OQS</u>

Consiste à résoudre une <u>suite de PQ</u>

$(x,\lambda) \to (x_+, \lambda_+)$

1) si (x,λ) est satisfaisant, on s'arrête

2) résoudre le PQO en $(x,\lambda) \to (d, \lambda^{PQ})$

3) $x_+ = x + d$; $\lambda_+ = \lambda^{PQ}$

- <u>Remarques</u> : • résoudre un PQ est plus simple que résoudre (P_{EI}), mais c'est beaucoup plus coûteux que de résoudre un système linéaire

• si $\nabla^2_{xx} \ell(x,\lambda) \not> 0$

$\Rightarrow$ (PQO) est <u>NP ardu</u>

(PQO) peut avoir des <u>solutions importantes</u>

donc on approche souvent $\nabla^2_{xx} \ell(x,\lambda)$ par une matrice $M > 0$ (méthode de QN)

• le (PQO) peut être <u>non borné</u>

(En fait, il suffit d'avoir un point stationnaire du PQO, mais il n'y a pas d'algo spécialisé dans cette tâche)

(plus de problème si $\nabla^2_{xx} \ell(x,\lambda) \sim M > 0$)

• le (PQO) peut ne <u>pas être réalisable</u>

($\exists$ des techniques pour y faire face)

Il faut se rappeler qu'il n'y a pas d'algorithme pouvant résoudre tous les (P_{EI}) !!

(C) Un résultat de convergence locale

théo

Si
- f, c sont $C^{2,1}$ proche de $x_* \in Sol(P_{EI})$
- $\exists$ unique multiplicateur λ_* associé à x_*
- CS2 est vérifié en (x_*, λ_*)

alors $\exists$ voisi V de (x_*, λ_*) tel que si $(x_1, \lambda_1) \in V$

alors
1) l'algorithme SQP peut générer une suite $\{(x_k, \lambda_k)\} \subset V$ en calculant à chaque itération des points stationnaires de PQO
2) $(x_k, \lambda_k) \to (x_*, \lambda_*)$ quadratiquement

Rmq importante

si il y a complémentarité stricte les contraintes actives en la solution sont celles des PQO (pour V assez petit) $\Rightarrow$ plus de combinatoire dans le PQO et il n'y a plus qu'un seul SL à résoudre à chaque itération

III Pénalisation exacte (§ 12.5)

A Motivation

- On cherche à "globaliser" l'algorithme SQP/OQS (c-à-d à forcer sa convergence même si (x_1, λ_1) n'est pas proche d'1 solution primale-duale), par exemple par recherche linéaire (les régions de confiance fonctionnent de la même manière avec quelques adaptations)

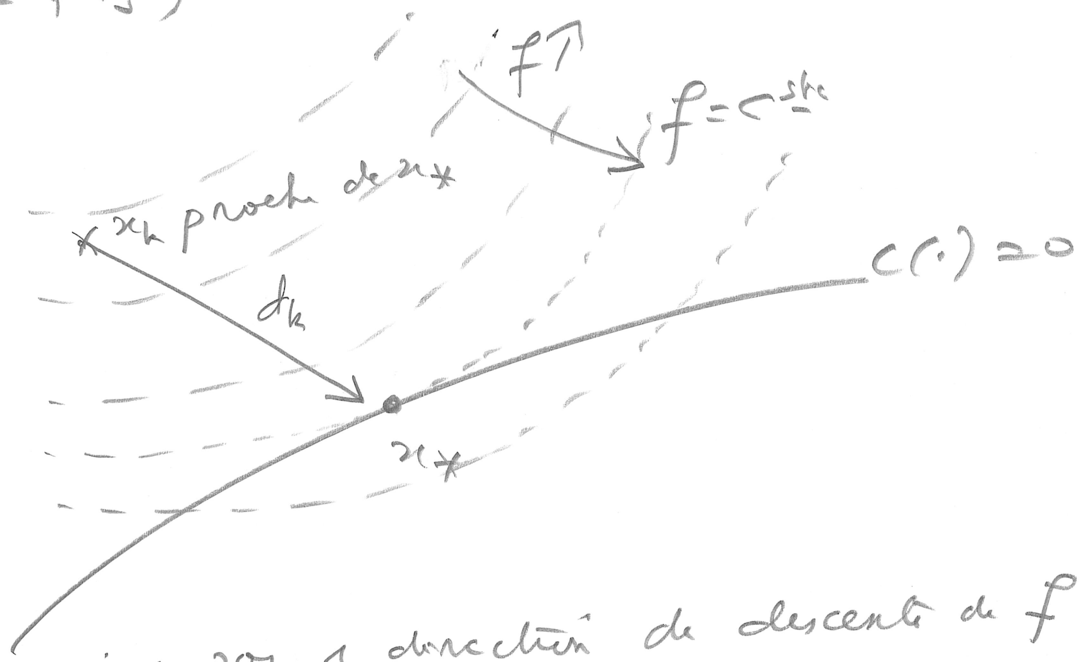
$$x_{k+1} = x_k + \alpha_k d_k$$

où d_k est la direction donnée par le PQO et $\alpha_k > 0$ est un pas.

<u>Rmq</u> : c'est 1 RL en variable primale, alors que SQP/OQS est en variable primale-duale, mais une RL en variable primale-duale n'est pas aisée.

- Comment déterminer α_k ?
- En utilisant quelle fonction ?

- Ça ne peut pas être f :
  (On suppose dans le dessin que $I = \emptyset$ et $E = \{1\}$)



d_k n'est pas 1 direction de descente de f en x_k !!

- On a besoin d'une fonction prenant en compte les deux aspects du problème : optimalité et admissibilité $\Rightarrow$ fonction de pénalisation

- Il faut que x_* soit 1 minimum local de cette fonction de pénalisation sinon la solution ci-dessus se représente (discutable, car on peut jouer avec le facteur de pénalisation) $\Rightarrow$ besoin d'1 fonction de pénalisation exacte

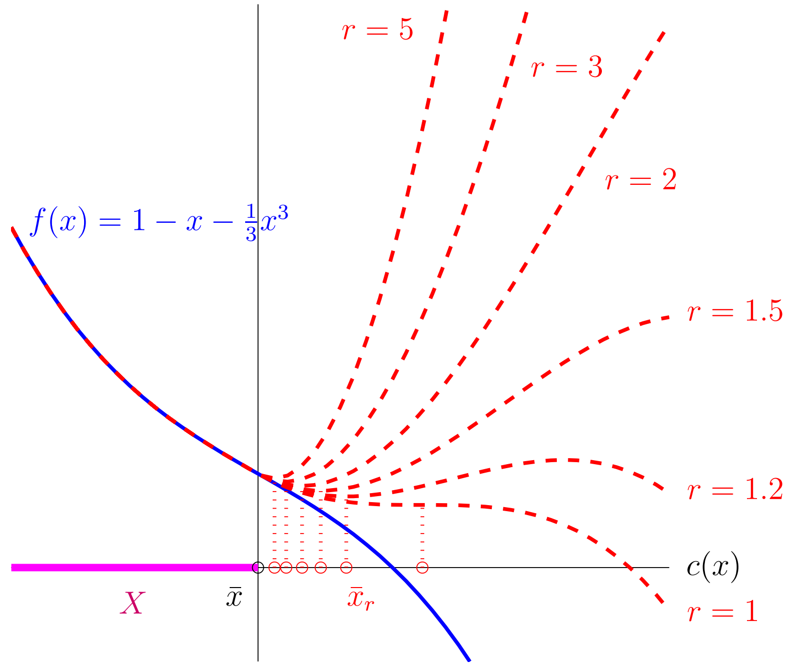
- On prend $\Theta_\sigma : \mathbb{E} \to \mathbb{R}$ dfn en $x \in \mathbb{E}$ par

$$\boxed{\Theta_\sigma(x) = f(x) + \sigma \| c(x)^\# \|}$$

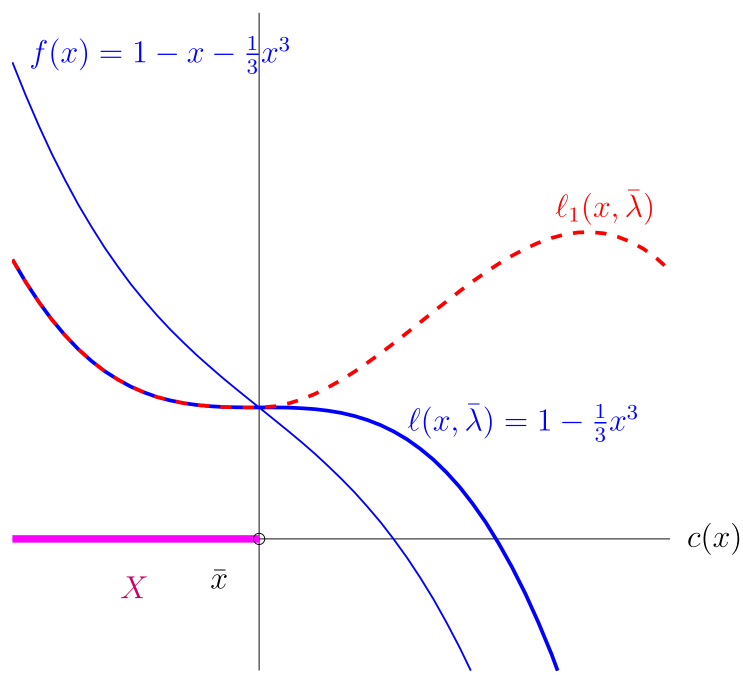
où $\|\cdot\|$ est une norme arbitraire (pour l'instant)

Rappel

10'



Pénalisation quadratique

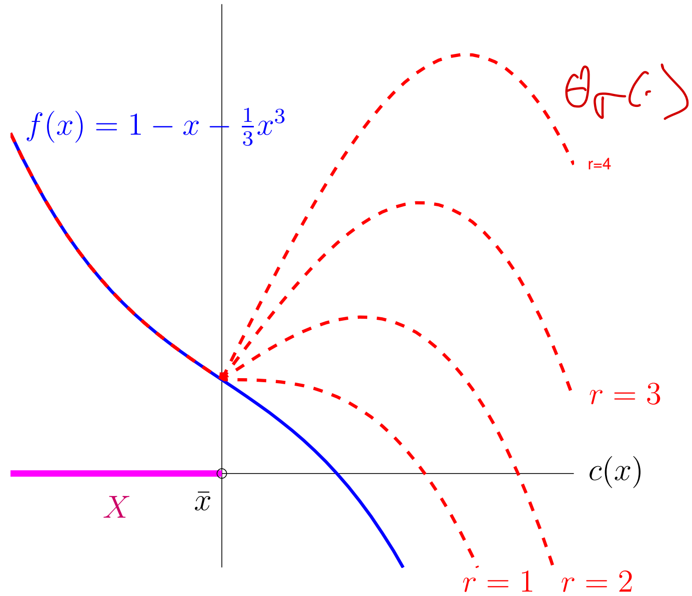


Pénalisation quadratique du lagrangien

(B) Exactitude de Θ_σ

Norme duale $\|\cdot\|_D$ de $\|\cdot\|$:

$$\boxed{\|v\|_D = \sup_{\|u\| \leq 1} v^T u}$$



$$\begin{cases} \min f(x) \\ x \leq 0 \end{cases}$$

Pénalisation exacte non différentiable

La non différentiabilité crée un minimum local en x_* si σ est assez grand.

<u>théor</u>

<u>Si</u>
- x_* min <u>local</u> de (P_E)
- f et c sont C^2 près de x_*
- $c'(x_*)$ surjective (donc $\exists!\ \lambda_*$ optimal)
- $CS2$
- $\sigma \geq \|\lambda_*\|_D$

<u>alors</u> x_* est un <u>min</u> <u>local</u> <u>strict</u> de Θ_σ

Lemme

> Si • φ a des dérivées directionnelles (DD)
> • ψ a des DD et est lipschitz
>
> alors $\psi\circ\varphi$ a des DD

Dém Soient $h \in E$, $t > 0$

$$(\psi\circ\varphi)(x+th)$$
$$= \psi(\varphi(x+th))$$
$$= \psi(\varphi(x)+t\varphi'(x;h)+o(t))$$
$$= \psi(\varphi(x)+t\varphi'(x;h)) + o(t)$$

car ψ est lipschitz

$$= \psi(\varphi(x)) + t\,\underbrace{\psi'(\varphi(x);\varphi'(x;h))}_{(\psi\circ\varphi)'(x;h)} + o(t)$$

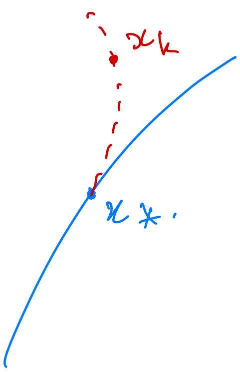
$\square$

Dém du théorème (lorsque $I = \emptyset$) (par l'absurde)

$$\theta_\sigma(x) = f(x) + \sigma \, \| c(x)^\# \| \underset{\uparrow \, I = \emptyset}{=} f(x) + \sigma \|c(x)\|$$

- Si x_* n'est un min local strict de θ_σ :

il n'est pas vrai que $\exists\, V$ de x_*, $\forall x \in V \setminus \{x_*\}$ on a $\theta_\sigma(x) > \theta_\sigma(x_*)$



$\Rightarrow \exists\, (x_k) \to x_*$, $x_k \neq x_*$: $\theta_\sigma(x_k) \leq \theta_\sigma(x_*)$

$$t_k = \|x_k - x_*\| \to 0 \qquad d_k = \frac{x_k - x_*}{t_k} \xrightarrow{\text{s. suite}} d \neq 0$$

$$x_k = x_* + t_k d_k = x_* + t_k d + t_k \underbrace{(d_k - d)}_{\to 0} = x_* + t_k d + o(t_k)$$

DD

- θ_σ est lipschitz avec des

$$\theta_\sigma(x_*) \geq \theta_\sigma(x_k) = \theta_\sigma(x_* + t_k d + o(t_k)) = \theta_\sigma(x_* + t_k d) + o(t_k)$$

$$\Rightarrow \theta_\sigma'(x_*; d) \leq 0$$

$$\| \qquad f'(x_*) \cdot d + \sigma \underbrace{(\|\cdot\| \circ c)'(x_*; d)}_{\|\cdot\|'(c(x_*);\, c'(x_*) d) \ (\text{lemme})}$$

$$\lim_{t \downarrow 0} \frac{1}{t} \Big[\| \underbrace{c(x_*)}_{=0} + t\, c'(x_*) d \| - \| \underbrace{c(x_*)}_{=0} \| \Big]$$

$$\| c'(x_*) d \|$$

$$|\langle u, v \rangle| \leq \|u\| \, \|v\|_D$$

$$0 \geq f'(x_*) \cdot d + \sigma \| c'(x_*) \cdot d \|$$

$$= - d_*^T c'(x_*) d + \sigma \| c'(x_*) d \|$$

$$\sigma \| c'(x_*) d \| \leq d_*^T c'(x_*) d$$

$$\leq \|d_*\|_D \, \| c'(x_*) d \|$$

$$\underbrace{\left(\sigma - \|d_*\|_D \right)}_{>0} \| c'(x_*) d \| \leq 0$$

$$\Rightarrow \quad c'(x_*) d = 0$$

- CS2, $d_* \in \Lambda_*$:

$$d^T \nabla^2_{xx} \ell(x_*, d_*) \, d \geq 0 \qquad \begin{pmatrix} d \neq 0 \\ c'(x_*) d = 0 \end{pmatrix}$$

$$\ell(x_k, d_*) = \ell(x_*, d_*) + \frac{t_k^2}{2} d^T \nabla^2_{xx} \ell(x_*, d_*) d + o(t_k^2)$$

$$\longrightarrow \quad \ell(x_*, d_*) \leq \ell(x_k, d_*), \text{ th proved}$$

- Contradiction

$$\theta_\sigma(x_k) \leq \theta_\sigma(x_*) = f(x_*) = \ell(x_*, d_*)$$

$$\leq \ell(x_k, d_*) = f(x_k) + d_*^T c(x_k)$$

$$\leq f(x_k) + \underbrace{\|d_*\|_D}_{\leq \sigma} \| c(x_k) \|$$

$$\leq \theta_\sigma(x_k) \qquad \text{contradiction} \qquad \square$$

IV $\boxed{\text{Globalisation de l'OQS}}$ $(I = \emptyset)$

On détermine un pas α_k tel que

$$\boxed{\Theta_\sigma(x_k + \alpha_k d_k) \leq \Theta_\sigma(x_k) + \omega\, \alpha_k\, \Theta'_\sigma(x_k ; d_k)}$$

$\omega \in]0, \frac{1}{2}[$ (typiquement 10^{-4})

On se contente de montrer ici que d_k déterminée par le (PQO) est une direction de descente de Θ_σ en x_k, c-à-d

> Si
> - d_k est solution du (PQO) avec $M_k \succ 0$
> - $\sigma > \|\lambda_k^{QP}\|_D$
> - x_k n'est pas un point stationnaire
>
> alors $\Theta'_\sigma(x_k ; d_k) < 0$

Dém

$$\Theta_\sigma(x) = f(x) + \sigma(\|\cdot\| \circ c)(x)$$

$$\Theta'_\sigma(x_k ; d_k) = \nabla f(x_k)^T d_k + \sigma \underbrace{(\|\cdot\|)'(c(x_k) ; c'(x_k) d_k)}$$

(PQO) (KKT)

$$\begin{cases} \nabla f(x_k) + M_k d_k + c'(x_k)^T \lambda_k^{PQ} = 0 \\ c(x_k) + c'(x_k) d_k = 0 \end{cases}$$

$$\lim_{t \downarrow 0} \frac{1}{t}\left[\|c(x_k) + t \underbrace{c'(x_k) d_k}_{= -c(x_k)}\| - \cancel{\|c(x_k)\|}\right]$$

$(\cancel{1} - t)\|c(x_k)\|$

$= -\|c(x_k)\|$

$$= \nabla f(x_k)^T d_k - \sigma \|c(x_k)\|$$

$$= -d_k^T M_k d_k - (d_k^{PQ})^T \underbrace{c'(x_k) d_k}_{\underbrace{-c(x_k)}_{(d_k^{PQ})^T c(x_k)}} - \sigma \|c(x_k)\|$$

$$\leq \|d_k^{PQ}\|_D \; \|c(x_k)\|$$

$$\leq -\underbrace{d_k^T M_k d_k}_{\geq 0 \; (M_k \succ 0)} - \underbrace{(\underbrace{\sigma - \|d_k^{PQ}\|_D}_{> 0}) \|c(x_k)\|}_{< 0}$$

$$\leq 0$$

Supposons que $\theta'(x_k ; d_k) = 0$

$\Rightarrow d_k = 0$ et $c(x_k) = 0$

CO de PQO $\Rightarrow \begin{cases} \nabla f(x_k) + c'(x_k)^T \lambda_k^{PQ} = 0 \\ c(x_k) = 0 \end{cases}$

$\Rightarrow (x_k, \lambda_k)$ est stationnaire

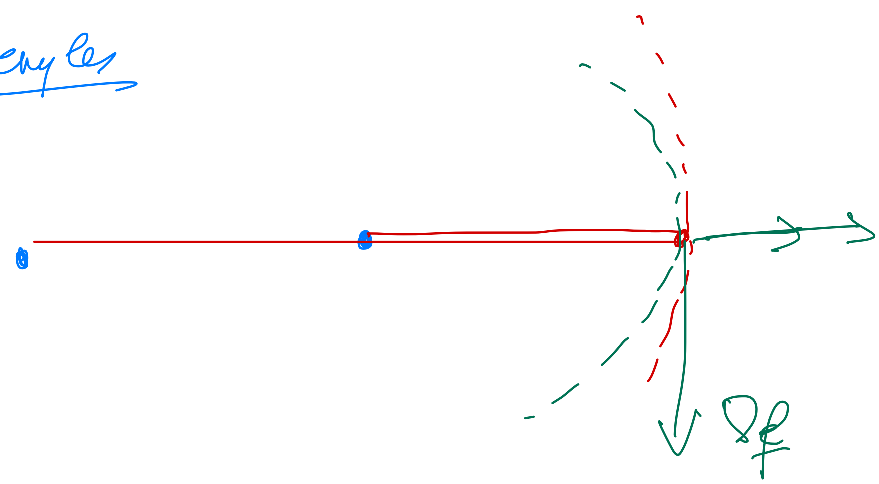
$\square$

Rmq sur les TP

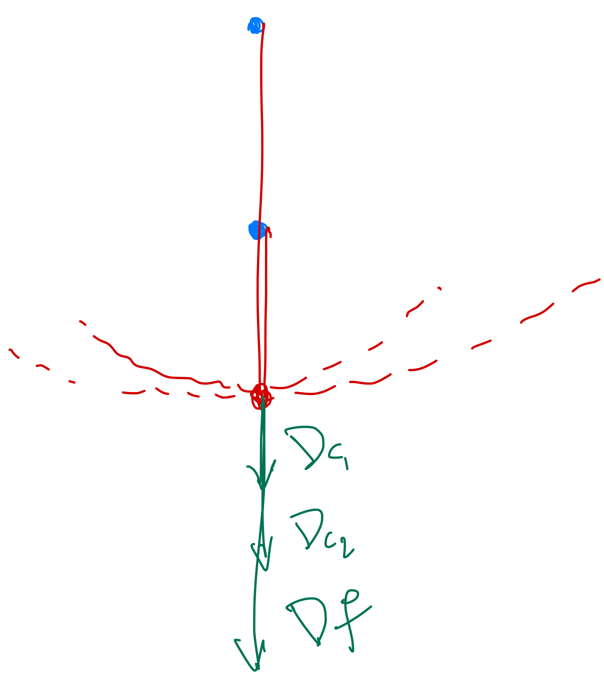
① ∃ de multiplicateurs (cond d'=)

$c'(x_*)$ surjective $\Rightarrow$ (contraintes qualifiées) $\Rightarrow \exists \lambda_*$

Exemples



$$\nabla f(x) + \sum \lambda_i \nabla c_i(x) = 0$$



$$\nabla c(x) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ x \\ x \\ x \end{pmatrix} = \begin{pmatrix} 0 \\ X \end{pmatrix}$$

2

x_* sol $\Rightarrow$ CN2

x_* sol. $\Leftarrow$ CS2

CS2 $L_* \succcurlyeq 0$ dans $N(c'(x_*))$

$d^T L_* d \geq 0$, $\underline{\forall d : c'(x_*) d = 0}$

$Z := \underline{\text{null}}(c'(x))$ $n-m$

$\Leftrightarrow \forall u \in \mathbb{R}^{n-m}$ $u^T \boxed{Z^T L_* Z} u \geq 0$

⚠ il ne suffit pas de regarder à

$(e^i)^T Z^T L_* Z \underline{e^i} \geq 0$, $\forall i$

$$(e^i)^T \begin{pmatrix} \varepsilon & 1 \\ 1 & \varepsilon \end{pmatrix} e^i = \varepsilon > 0$$

(3) $d_1 = ?$

$$\nabla f(x_1) + \underbrace{c'(x_1)^T}_{n \times m} \underbrace{d_1}_{m} \neq 0$$

$\rightarrow$ $d = -(c'(x)\, c'(x)^T)^{-1}\, c'(x)\, \nabla f(x)$

$\rightarrow$ $\min \| \nabla f(x_1) + c'(x_1)^T d_1 \|_2^2$

$$d = -c'(x) \setminus \nabla f(x)$$

(N) $\nabla f(x) + \nabla^2 f(x)\, d = 0$

$\nabla^2 f(x) \rightarrow I$ (algorithme de gradient)

$\nabla f(x) + d = 0 \Rightarrow d = -\nabla f(x)$