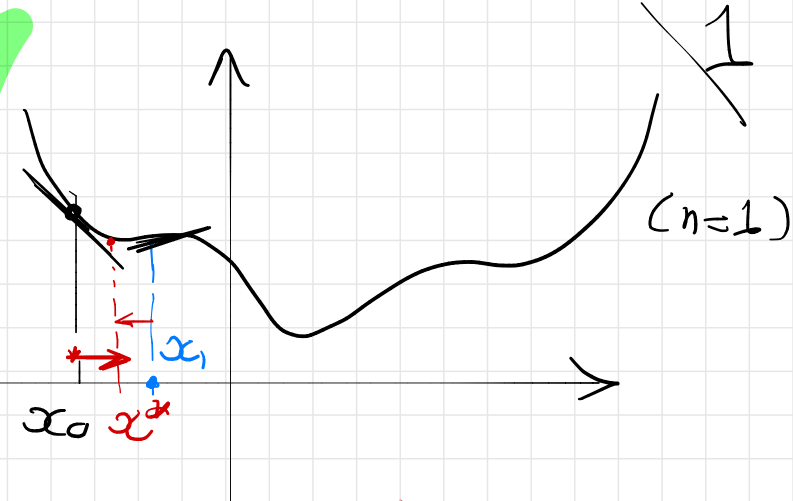
« Optimisation numérique »

Chap. 2 : algorithme d'optimisation locale sans contraintes

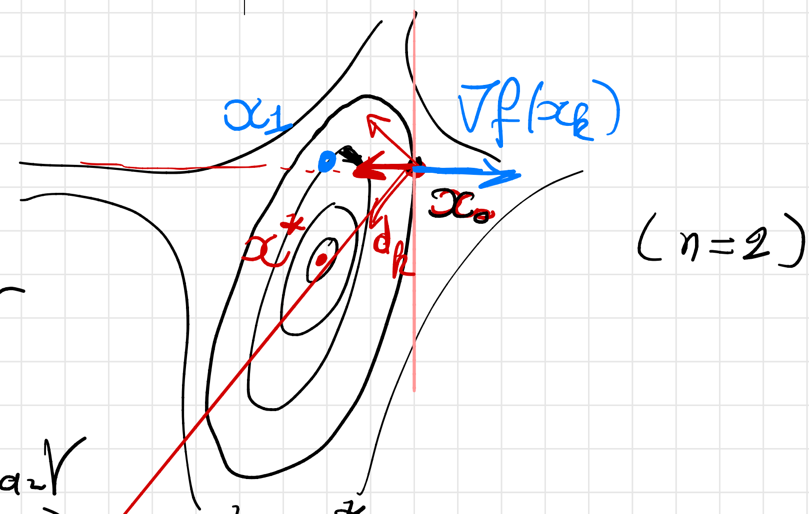


2.1) Principe général des méthodes de descente

Pb à résoudre :

$f : \mathbb{R}^n \to \mathbb{R}$, C^1, voire C^2.

$x_0 \in \mathbb{R}^n$ On cherche à approcher un minimum local de f situé à proximité de x_0, en construisant une suite d'approximations $(x_k)_{k \in \mathbb{N}}$ de x^*.



Une méthode de descente impose d'avoir

$$\forall k \in \mathbb{N},\ f(x_{k+1}) \leq f(x_k)$$

(strict si $x_{k+1} \neq x_k$)

Plus précisément, on construit (x_k) de manière itérative à un pas, en se donnant

* une direction de descente $d_k \in (\mathbb{R}^n)^*$
* un pas dans la direction d_k : $\alpha_k \in \mathbb{R}_+^*$

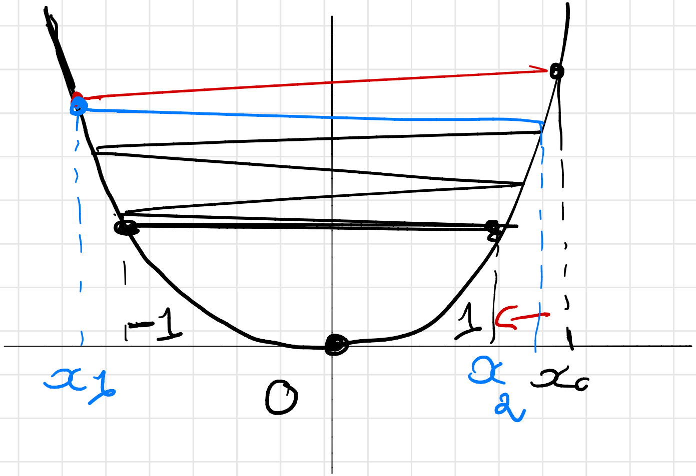
Ainsi, on a l'itération :

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

L'objectif va consister à construire de telles méthodes et à justifier leur convergence éventuelle vers un minimum local de f.

Il ne suffit pas d'imposer une descente stricte pour obtenir la convergence vers x^*.

Deux "contre-exemples" en dimension $n = 1$ en attestent.

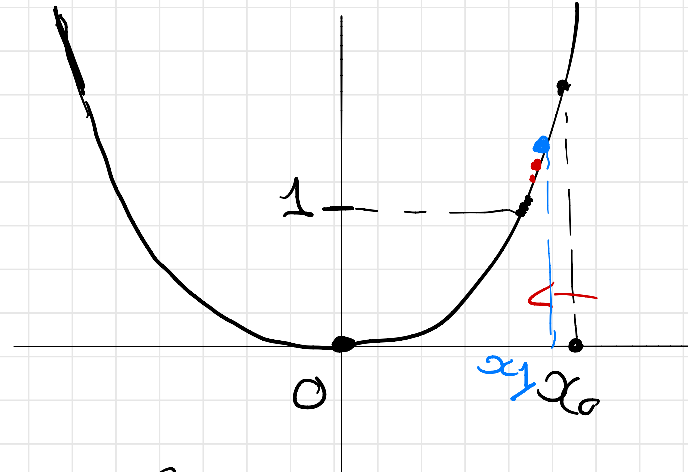


$f(x) = x^2$.

($x_0 = 2$, $x_{k+1} = x_k + \alpha_k d_k$ avec $d_k = (-1)^{k+1}$ et $\alpha_k = 2 + 3/2^{k+1}$)

Pas trop grand (*) par rapport à la descente effectuée.

((*) par une descente trop faible)



$f(x) = x^2$

($x_0 = 2$, $x_{k+1} = x_k + \alpha_k d_k$ avec $d_k = -1$, $\alpha_k = \frac{1}{2^{k+1}}$)

Pas trop petit.

Dans les 2 cas, (x_k) ne converge pas vers le minimum $x^* = 0$ de f.

Pour assurer un résultat de convergence, il faut définir une stratégie de recherche linéaire pour une direction donnée (pas ni trop grand, ni trop petit)

Le choix de la direction de descente est très important. La direction opposée au gradient :

$$d_k = -\nabla f(x_k)$$

est une possibilité (on parle dans ce cas de méthode du gradient ou de plus forte pente) mais ce n'est pas forcément la meilleure direction.

Remarque : le seul cas où on ne peut pas définir de direction de descente est lorsque $\nabla f(x_k) = 0$. L'algorithme s'arrête dans ce cas.

2.2) Un exemple de recherche linéaire : "backtracking + Armijo"

→ Principe de backtracking

* On suppose qu'on se donne à chaque étape une direction de descente (d_k)

* On se donne un pas de référence $\alpha_{init} > 0$ qu'on teste dans un premier temps :

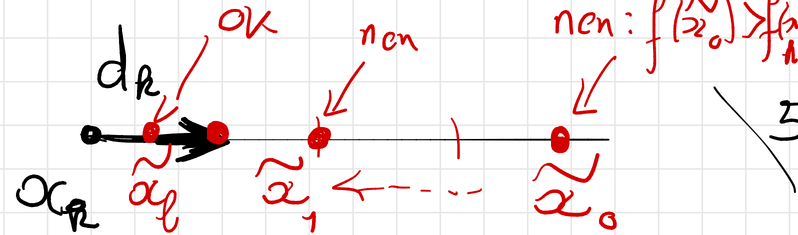
$$\tilde{x}_0 = x_k + \alpha_{init}\, d_k .$$

* Si $f(\tilde{x}_0) < f(x_k)$, on garde $x_{k+1} = \tilde{x}_0$ (et $\alpha_k = \alpha_{init}$)

• Sinon, on teste une nouvelle valeur de α_k : $\tilde{\alpha} = \tau\, \alpha_{init}$ où $\tau \in]0,1[$ fixé, et de $\tilde{x}$:

$$\tilde{x}_1 = x_k + \tilde{\alpha}\, d_k$$

et ainsi de suite jusqu'à obtenir une descente effective avec $\tilde{x}_\ell$ $(\ell \geq 1)$



$(\alpha_{init} = 4,\ \tau = \frac{1}{2})$

Comme d_k est une direction de descente, cet algorithme s'arrête en temps fini

$$(f(\tilde{x}_\ell) < f(x_k))$$

Cet algorithme de backtracking permet de se prémunir des pas trop petits (voir Th.)

→ <u>Condition d'Armijo</u> :

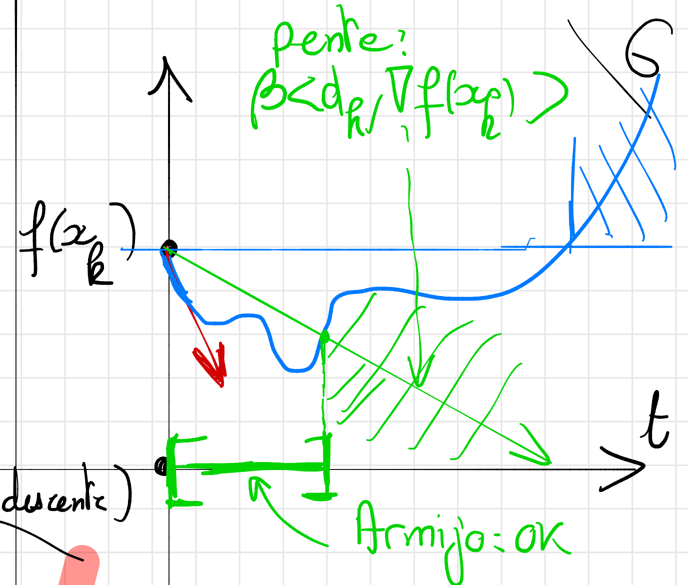
la condition d'Armijo permet d'assurer une descente suffisante dans le cas du pas grands :

on remplace la condition de descente

$$f(x_{k+1}) < f(x_k) \quad \text{par :} \quad \overbrace{<0 \;(d_k : \text{descente})}$$

$$f(x_k+\alpha_k d_k) \leq f(x_k) + \beta \alpha_k \langle d_k, \nabla f(x_k) \rangle$$

(où $\beta \in]0,1[$ fixé, en général petit)



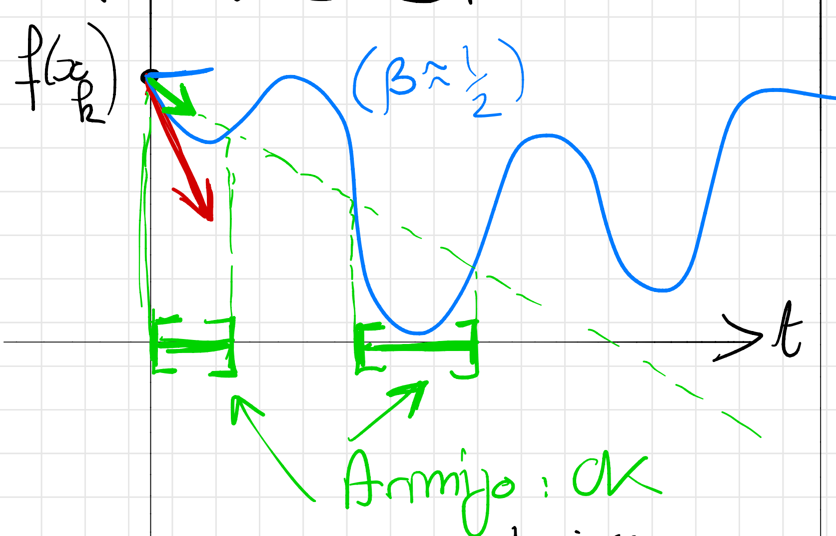
$h'(t) = \langle d_k, f(x_k + t d_k) \rangle$

et $h'(0) = \langle d_k, f(x_k) \rangle < 0$

Interprétation géométrique : on considère

$$h : \begin{pmatrix} [0,+\infty[\to \mathbb{R} \\ t \mapsto f(x_k + t d_k) \end{pmatrix}$$

Autre exemple :



Cette condition donne toujours des valeurs admissibles de pas, en particulier il existe toujours un intervalle initial admissible.

Proposition soit $f : C^1 : \mathbb{R}^n \to \mathbb{R}$ telle que $g = \nabla f : \mathbb{R}^n \to \mathbb{R}^n$ est Lipschitz :

$$\|g(y) - g(x)\| \leq \gamma \|y - x\|.$$

Alors la condition d'Armijo en un point $x \in \mathbb{R}^n$ et pour une direction de descente $d \in \mathbb{R}^n \setminus \{0\}$ est satisfaite si

$$t \in \left[0, \frac{2\underbrace{(\beta - 1)}_{<0} \overbrace{\langle d, \nabla f(x) \rangle}^{<0}}{\gamma \|d\|^2}\right]$$

preuve : en utilisant Taylor et l'hypothèse sur ∇f, on montre que

$$f(x + td) \leq f(x) + t\langle d, \nabla f(x)\rangle + \frac{1}{2}\gamma t^2 \|d\|^2$$

(à faire en exercice)

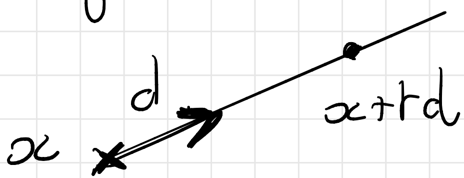
Si $t \leq \dfrac{2(\beta-1)\langle d, \nabla f(x)\rangle}{\gamma \|d\|^2}$

alors

$$f(x+td) \leq f(x) + t\langle d, \nabla f(x)\rangle + \underbrace{t\, \frac{1}{\cancel{2}}\cancel{\|d\|^2}\cancel{\gamma}\, \frac{\cancel{2}(\beta-1)\langle d, \nabla f(x)\rangle}{\cancel{\|d\|^2}\cancel{\gamma}}}$$

$$\leq f(x) + t\beta\langle d, \nabla f(x)\rangle$$

et t vérifie bien la condition d'Armijo



On dispose donc d'une stratégie de recherche linéaire "backtracking + Armijo" (paramètres : α_{init}, $\bar{\sigma}$, β)

On montre que cette stratégie permet de construire des suites d'approximation $(x_k)_{k\in\mathbb{N}}$ qui convergent vers un minimum local de f.

Théorème : soit f C^1 et $g = \nabla f$ Lipschitz continue. Alors la suite d'approximation (x_k) telle que $x_{k+1} = x_k + \alpha_k d_k$ avec

- d_k : direction de descente
- α_k : pas fixé suivant la stratégie "backtracking + Armijo"

est telle que :

soit (i) $\nabla f(x_k) = 0$ par un certain k

soit (ii) $\lim\limits_{k \to -\infty} f(x_k) = -\infty$

soit (iii)

$$\lim_{k \to +\infty} \mathrm{Min}\left(|\langle d_k, \nabla f(x_k) \rangle|, \frac{|\langle d_k, \nabla f(x_k) \rangle|}{\|d_k\|^2} \right) = 0$$

Corollaire : on prend les mêmes hypothèses avec du plus f minorée et $d_k = -\nabla f(x_k)$. Alors

soit $\nabla f(x_k) = 0$ pour un certain k, soit

$$\lim_{k \to +\infty} \nabla f(x_k) = 0$$

(preuve du corollaire immédiate : $\lim \left(\mathrm{Min}(\|d_k\|^2, 1) \right) = 0$)

Démonstration du théorème (partie 1)

(iii) $$\lim_{k\to+\infty} \mathrm{Min}\left(|\langle d_k, \nabla f(x_k)\rangle|,\ \frac{|\langle d_k, \nabla f(x_k)\rangle|}{\|d_k\|^2}\right) = 0$$

(en particulier : $d_k = -\nabla f(x_k)$: $\nabla f(x_k) \to 0$)

preuve (*) on a $f(x_{k+1}) \leq f(x_k) + \beta \alpha_k \langle d_k, \nabla f(x_k)\rangle$

$$\sum_{k=0}^{k} \implies f(x_{k+1}) \leq f(x_0) + \beta \sum_{k=0}^{k} \alpha_k \underbrace{\langle d_k, \nabla f(x_k)\rangle}_{<0} < 0$$

$$C \leq f(x_{k+1})$$

La série à termes négatifs est minorée, elle est donc convergente. On en déduit $\lim_{k\to+\infty} \alpha_k \langle d_k, \nabla f(x_k)\rangle = 0$

(*) on suppose non (ii) et non (i)

(partie 2)

* 2 cas pour α_k sont possibles :

* si

$$\alpha_{init} \leqslant \overbrace{\frac{2(\beta-1)\langle d_k, \nabla f(x_k)\rangle}{\gamma \|d_k\|^2}}^{\alpha_a}$$

alors $\alpha_k = \alpha_{init}$.

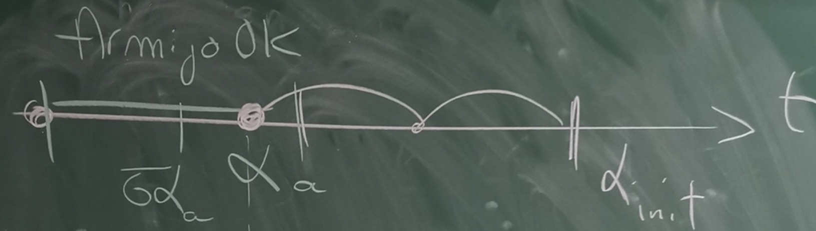
et $|\alpha_k \langle d_k, \nabla f(x_k)\rangle| = \alpha_{init}\, |\langle d_k, \nabla f(x$

* si $\alpha_{init} \geqslant (\quad " \quad)$, alors

$$\alpha_k \geqslant \sigma \, \frac{2(\beta-1)\langle d_k, \nabla f(x_k)\rangle}{\gamma \|d_k\|^2}$$

(backtracking)

(partie 3)



Alors

$$\underbrace{|\alpha_k \langle d_k, \nabla f(x_k)\rangle|}_{\downarrow \atop 0} \geq \frac{2\sigma(\beta-1)}{\gamma} \underbrace{\frac{|\langle d_k, \nabla f(x_k)\rangle|}{\|d_k\|^2}}_{\downarrow \atop 0}$$

Ex 1) TD3 (un exemple d'itération de cette méthode)

Un exemple d'itération de cette méthode :

Ex 1) TD 3

$J(x,y,z) = x^2 + 2y^2 + z^2 - x + y - z$ à minimiser sur $\mathbb{R}^3$

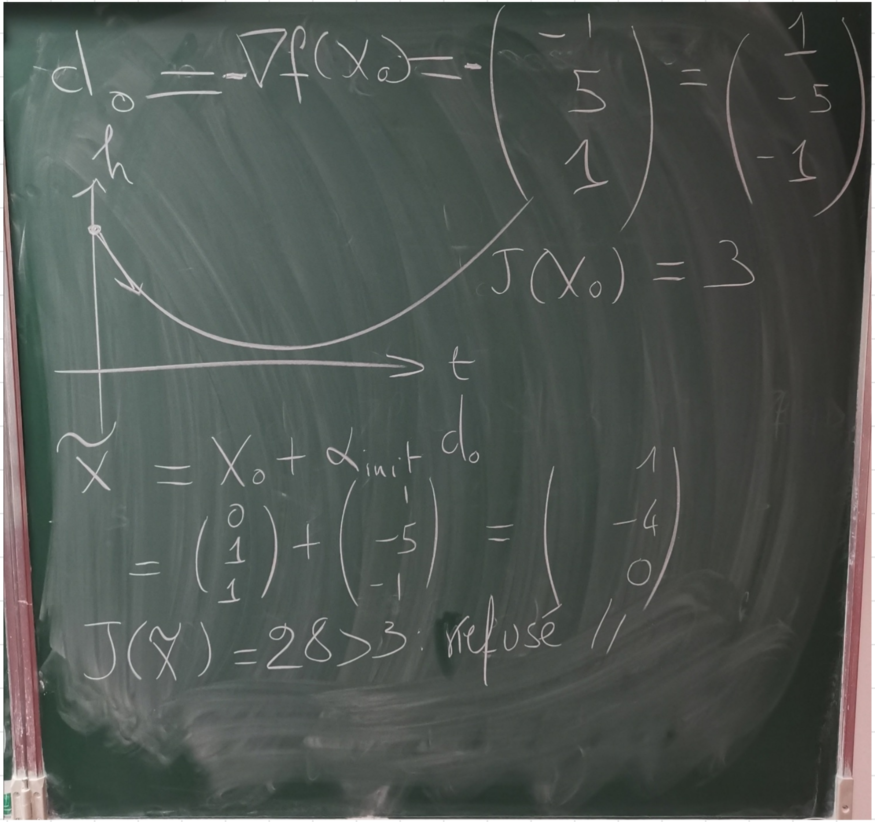
1) Solution exacte ? $\nabla J(x,y,z) = \begin{pmatrix} 2x-1 \\ 4y+1 \\ 2z-1 \end{pmatrix} = 0 \iff \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{4} \\ \frac{1}{2} \end{pmatrix}$

$HJ(x,y,z) = \begin{pmatrix} 2 & & 0 \\ & 4 & \\ 0 & & 2 \end{pmatrix} >> 0 \implies \left(\frac{1}{2}, -\frac{1}{4}, \frac{1}{2}\right)$ min global de J.

2) On part de $X_0 = (0,1,1)$. On prend $\alpha_{init} = 1$; $\sigma = 0{,}1$; $\beta = 0{,}1$

Que vaut X_1 ? (avec direction de plus forte pente : $d_0 = -\nabla J(X_0)$)

(suite)


$d_0 = -\nabla f(X_0) = -\begin{pmatrix} -1 \\ 5 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -5 \\ -1 \end{pmatrix}$
h
$J(X_0) = 3$
t
$\tilde{X} = X_0 + \alpha_{init} d_0$
$= \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ -5 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -4 \\ 0 \end{pmatrix}$
$J(\tilde{X}) = 28 > 3$: refusé //

(suite et fin)

$$\tilde{X}_1 = \begin{pmatrix}0\\1\\1\end{pmatrix} + \frac{1}{10}\begin{pmatrix}1\\-5\\-1\end{pmatrix} = \begin{pmatrix}0,1\\0,5\\0,9\end{pmatrix}$$

$$J(\tilde{X}_1) = 0,01 + 0,5 + 0,81 - 0,1 + 0,5 - 0,9$$
$$= 0,82$$

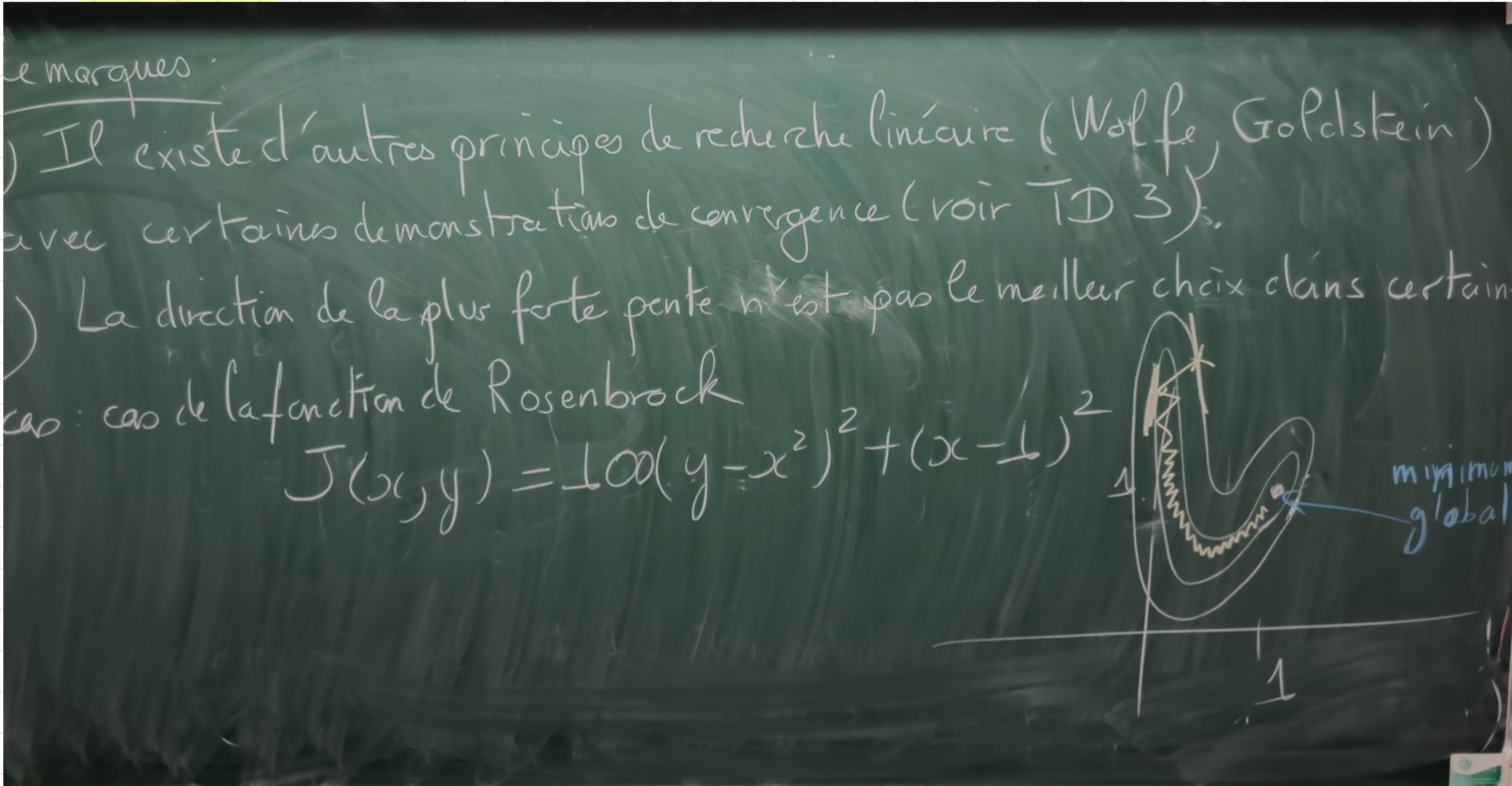
et

$$J(X_0) + \beta \frac{1}{10} \langle d_0, \nabla f(X_0) \rangle =$$

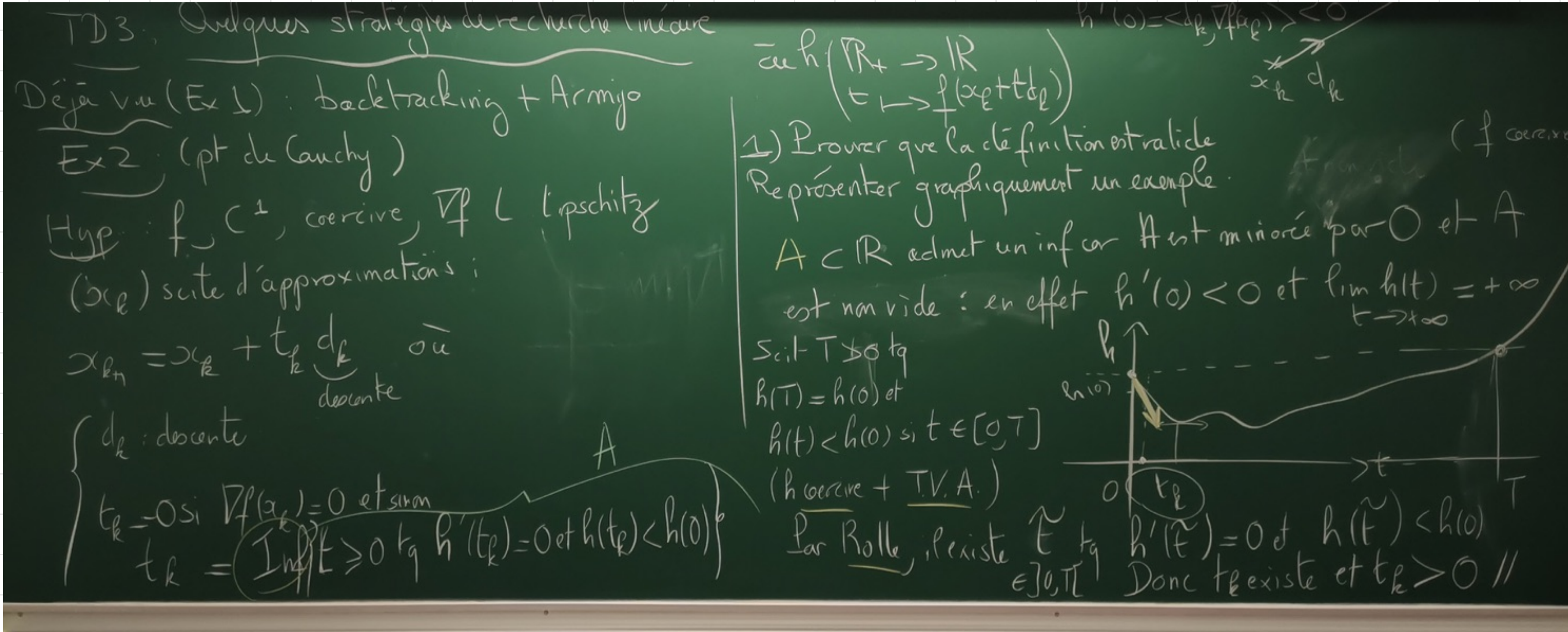
$$3 + \frac{1}{100}(-27) \geq 0,82$$

On a donc $X_1 = \begin{pmatrix}0,1\\0,5\\0,9\end{pmatrix}$

Remarques


emarques :
) Il existe d'autres principes de recherche linéaire (Wolfe, Goldstein)
avec certaines demonstrations de convergence (voir TD 3).
) La direction de la plus forte pente n'est pas le meilleur choix dans certain
cas : cas de la fonction de Rosenbrock
$J(x,y) = 100(y-x^2)^2 + (x-1)^2$
1
1
minimum global

TD3) Ex 2) (pt de Cauchy)



TD3 : Quelques stratégies de recherche linéaire

Déjà vu (Ex 1) : backtracking + Armijo

Ex 2 : (pt de Cauchy)

Hyp : f, C^1, coercive, ∇f L lipschitz

(x_k) suite d'approximations :

$x_{k+1} = x_k + t_k \underbrace{d_k}_{\text{descente}}$ où

$$\begin{cases} d_k : \text{descente} \\ t_k = 0 \text{ si } \nabla f(x_k) = 0 \text{ et sinon} \\ t_k = \operatorname{Inf} \underbrace{\{t \geq 0 \text{ tq } h'(t_k) = 0 \text{ et } h(t_k) < h(0)\}}_{A} \end{cases}$$

où $h : \begin{pmatrix} \mathbb{R}_+ \to \mathbb{R} \\ t \mapsto f(x_k + t d_k) \end{pmatrix}$

$h'(0) = \langle d_k, \nabla f(x_k) \rangle < 0$

x_k d_k

(f coercive

1) Prouver que la définition est valide. Représenter graphiquement un exemple.

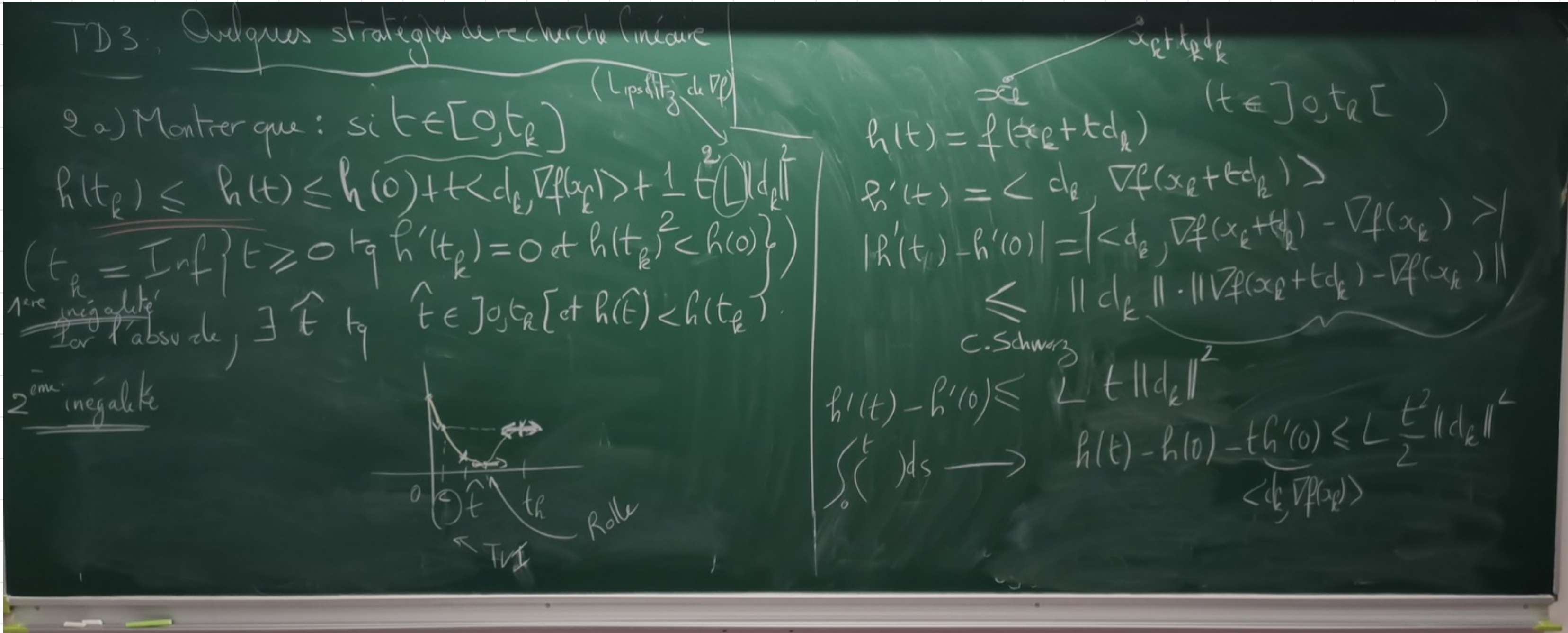
$A \subset \mathbb{R}$ admet un inf car A est minorée par 0 et A est non vide : en effet $h'(0) < 0$ et $\lim_{t \to +\infty} h(t) = +\infty$

Soit $T > 0$ tq $h(T) = h(0)$ et $h(t) < h(0)$ si $t \in]0, T[$ (h coercive + T.V.A.)

Par Rolle, il existe $\tilde{t} \in]0,T[$ tq $h'(\tilde{t}) = 0$ et $h(\tilde{t}) < h(0)$

Donc t_k existe et $t_k > 0$ //

(suite Ex 2)


TD3 : Quelques stratégies de recherche linéaire
(Lipschitz de ∇f)
2 a) Montrer que : si t ∈ [0, t_k]
h(t_k) ≤ h(t) ≤ h(0) + t<d_k, ∇f(x_k)> + 1/2 t² L ||d_k||²
(t_k = Inf {t ≥ 0 tq h'(t_k) = 0 et h(t_k) < h(0)})
1ère inégalité
Par l'absurde, ∃ t̂ tq t̂ ∈]0, t_k[et h(t̂) < h(t_k)
2ème inégalité
0
t̂
t_k
Rolle
TVI
x_k
x_k + t_k d_k
(t ∈]0, t_k[ )
h(t) = f(x_k + td_k)
h'(t) = < d_k, ∇f(x_k + td_k) >
|h'(t) − h'(0)| = |< d_k, ∇f(x_k + td_k) − ∇f(x_k) >|
≤ ||d_k|| · ||∇f(x_k + td_k) − ∇f(x_k)||
C.Schwarz
h'(t) − h'(0) ≤ L t ||d_k||²
∫₀ᵗ ( ) ds
h(t) − h(0) − t h'(0) ≤ L t²/2 ||d_k||²
<d_k, ∇f(x_k)>

PARTIE 1

2b) Trouver t_{min} tel que l'expression de droite soit minimale

La dérivée de l'expression de droite (parabole) s'annule en t_{min} :

$$\langle d_k, \nabla f(x_k)\rangle + t_{min} L\|d_k\|^2 = 0 \Rightarrow$$

$$t_{min} = -\frac{\langle d_k, \nabla f(x_k)\rangle}{L\|d_k\|^2}$$

2c) Trouver

$$\boxed{0 \leq \langle d_k, \nabla f(x_k)\rangle + t_k L\|d_k\|^2} \quad (*)$$

$h'(t_k) = 0$ soit $\langle d_k, \nabla f(x_k + t_k d_k)\rangle = 0$

et ainsi $\langle d_k, \nabla f(x_k + t_k d_k) - \nabla f(x_k)\rangle \leq L t_k \|d_k\|^2$

$$\Rightarrow 0 \leq \langle d_k, \nabla f(x_k)\rangle + L t_k \|d_k\|^2$$

PARTIE 2

TD3 : Quelques stratégies de recherche linéaire

En déduire que

$$f(x_{k+1}) \leq f(x_k) - \frac{\langle d_k, \nabla f(x_k)\rangle^2}{2L\|d_k\|^2}$$

On a

$$\boxed{t_{min} \leq t_k}$$

En effet, en t_k, la dérivée de la parabole est positive (voir (*)).

Avec 2a) appliqué en $t = t_{min}$

$$h(t_k) \leq h(0) + t_{min}\langle d_k, \nabla f(x_k)\rangle + \frac{1}{2} t_{min}^2 L\|d_k\|^2$$

t_{min}

t_k

(suite Ex2)

PARTIE1

4) On suppose $d_k = -\nabla f(x_k)$. Montrer que

$$\lim_{k \to +\infty} \nabla f(x_k) = 0$$

On a dans ce cas

$$f(x_{k+1}) \leq f(x_k) - \frac{\langle d_k, -d_k \rangle^2}{2L \|d_k\|^2}$$

soit

$$f(x_{k+1}) \leq f(x_k) - \frac{1}{2L} \|d_k\|^2$$

soit par somme de 0 à K :

$$f(x_{K+1}) \leq f(x_0) - \frac{1}{2L} \sum_{k=0}^{K} \|d_k\|^2$$

$$\iff \frac{1}{2L} \sum_{k=0}^{K} \|d_k\|^2 \leq \underbrace{f(x_0) - f(x_{K+1})}_{\text{majoré}}$$

minoré (f coerc.)

PARTIE2

série à termes positifs $\sum \|d_k\|^2$ est

orée, elle est donc convergente. Ainsi,

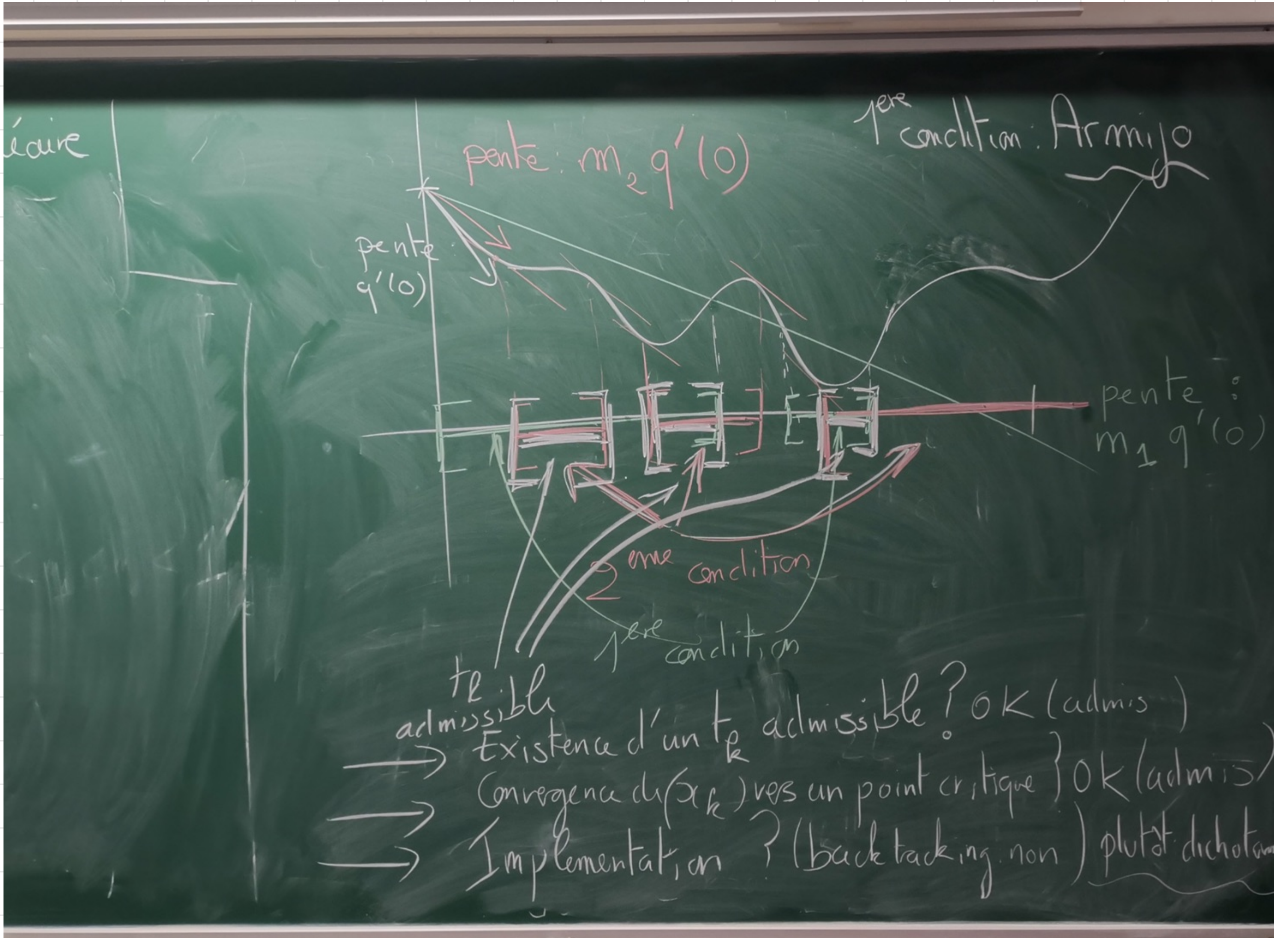
terme général tend vers 0 :

$$\lim_{k \to +\infty} d_k = \lim_{k \to +\infty} \nabla f(x_k) = 0$$

3. Quelques stratégies de recherche linéaire

Ex3)

(Wolfe)


éaire
pente : m2 q'(0)
1ère condition : Armijo
pente q'(0)
pente : m1 q'(0)
2ème condition
1ère condition
tk admissible
Existence d'un tk admissible ? OK (admis)
Convergence de (xk) vers un point critique ? OK (admis)
Implementation ? (backtracking non) plutôt dichotom