ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Optimisation non linéaire sans contraintes

Recherche opérationnelle

GC-SIE

Méthodes de descente

Directions de descente

- Problème :

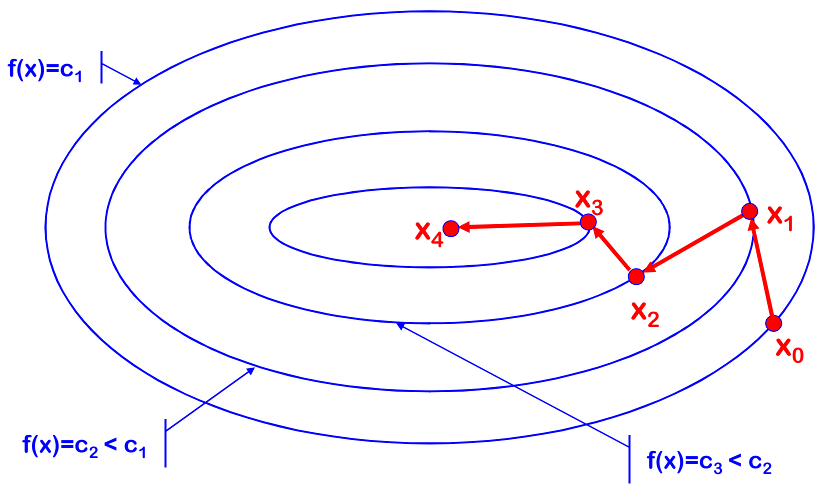
$$\min f : IR^n \rightarrow IR$$

f continûment différentiable

- Idée :
  - On démarre d'un point x_0
  - On génére des vecteurs x_1, x_2,… tels que la valeur de f décroit à chaque itération :

$$f(x_{k+1}) < f(x_k) \quad k=1,2,\ldots$$

Directions de descente



Directions de descente

- Soit $x \in \mathbb{R}^n$ tel que $\nabla f(x) \neq 0$.
- Condidérons la demi-droite

$$x_\alpha = x - \alpha \nabla f(x)$$

- Théorème de Taylor (1er ordre)

$$f(x+s) = f(x) + \nabla f(x)^T s + o(\|s\|)$$

avec $s = x_\alpha - x$

$$\begin{aligned} f(x_\alpha) &= f(x) + \nabla f(x)^T (x_\alpha - x) + o(\|x_\alpha - x\|) \\ &= f(x) - \alpha \|\nabla f(x)\|^2 + o(\alpha \|\nabla f(x)\|) \\ &= f(x) - \alpha \|\nabla f(x)\|^2 + o(\alpha) \end{aligned}$$

Directions de descente

$$f(x_\alpha) = f(x) - \alpha \|\nabla f(x)\|^2 + o(\alpha)$$

- Si α est petit, on peut négliger $o(\alpha)$
- Donc, pour α positif mais petit,

$$f(x_\alpha) < f(x)$$

Théorème :

- Il existe δ tel que, pour tout $\alpha \in \,]0,\delta[$

$$f(x - \alpha \nabla f(x)) < f(x)$$

Directions de descente

- Gradient = plus forte pente
- Question : y a-t-il d'autres directions de descente que $-\nabla f(x)$?
- Appliquons le même raisonnement avec $d \neq 0$.

Directions de descente

- Condidérons la demi-droite

$$x_\alpha = x + \alpha\, d$$

- Théorème de Taylor (1er ordre)

$$f(x+s) = f(x) + \nabla f(x)^T s + o(\|s\|)$$

avec $s = x_\alpha - x$

$$\begin{aligned} f(x_\alpha) &= f(x) + \nabla f(x)^T (x_\alpha - x) + o(\|x_\alpha - x\|) \\ &= f(x) + \alpha\, \nabla f(x)^T d + o(\alpha\, \|d\|) \\ &= f(x) + \alpha\, \nabla f(x)^T d + o(\alpha) \end{aligned}$$

Directions de descente

$$f(x_\alpha) = f(x) + \boldsymbol{\alpha\, \nabla f(x)^T d} + o(\alpha)$$

- Si α est petit, on peut négliger $o(\alpha)$
- Pour avoir $f(x_\alpha) < f(x)$, il faut

$$\nabla f(x)^T d < 0$$

Théorème :

- Soit d tel que $\nabla f(x)^T d < 0$. Il existe δ tel que, pour tout $\alpha \in \,]0,\delta[$

$$f(x+\alpha d) < f(x)$$

Directions de descente

Définition :

- Soit $f:IR^n \rightarrow IR$, une fonction continûment différentiable, et x un vecteur de IR^n. Le vecteur $d \in IR^n$ est appelé **direction de descente** de f en x ssi

$$\nabla f(x)^T d < 0$$

Méthodes de descente

Algorithme de base :

- Soit $x_0 \in IR^n$. Poser k=0.
- Tant que $\nabla f(x_k) \neq 0$
  - Choisir d_k tel que $\nabla f(x_k)^T d_k < 0$
  - Choisir $\alpha_k > 0$
  - Poser $x_{k+1} = x_k + \alpha_k d_k$

Méthodes de descente

Notes :

- Il y a beaucoup de choix possibles
- En général, on choisit α_k tel que
  $$f(x_k+\alpha_k d_k) < f(x_k)$$
  mais il y a des exceptions
- Il n'y a aucune garantie de convergence pour l'algorithme de base

Choix de la direction

- Ecrivons $d_k = -D_k \nabla f(x_k)$
  où D_k est une matrice n x n
- La condition $\nabla f(x_k)^T d_k < 0$ s'écrit

$$\nabla f(x_k)^T D_k \nabla f(x_k) > 0$$

- Si D_k est définie positive, cette condition est toujours vérifiée.
- Le choix de la direction revient donc au choix d'une matrice définie positive.

Choix de la direction

Quelques exemples souvent utilisés

- Méthode de la plus forte pente
  - $D_k = I$
  - $x_{k+1} = x_k - \alpha_k \nabla f(x_k)$
- Méthode de Newton
  - $D_k = (\nabla^2 f(x_k))^{-1}$
  - $x_{k+1} = x_k - \alpha_k (\nabla^2 f(x_k))^{-1} \nabla f(x_k)$

Attention: il faut que $\nabla^2 f(x_k)$ soit inversible et déf. pos.

Choix de la direction

- Mise à l'échelle diagonale
  - $D_k = \begin{pmatrix} d_{k1} & 0 & 0 & \cdots & 0 \\ 0 & d_{k2} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & d_{kn} \end{pmatrix}$
  - $d_{ki} > 0$ pour tout i
  - Exemple : $d_{ki} = \left(\frac{\partial^2 f(x_k)}{\partial x_i^2}\right)^{-1}$

Choix de la direction

- Méthode de Newton modifiée
  - $D_k = (\nabla^2 f(x_0))^{-1}$
  - $x_{k+1} = x_k - \alpha_k (\nabla^2 f(x_0))^{-1} \nabla f(x_k)$
- etc…

Choix du pas

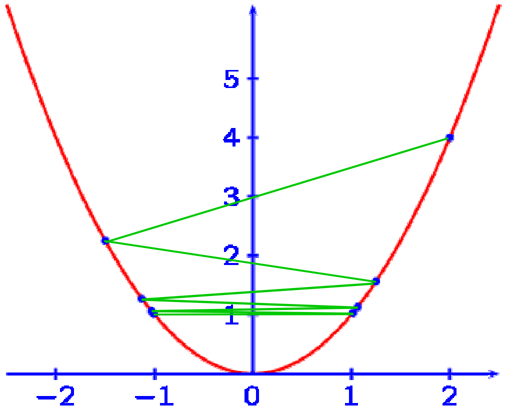
1. Règle de minimisation
   - Problème à une seule variable
2. Règle d'approximation
   - Minimisation prend du temps
   - Section d'or nécessite l'unimodalité
   - A-t-on besoin du minimum exact ?
   - Idée :
     - Choisir un pas qui diminue suffisamment la valeur de la fonction.

Choix du pas : approximation

- Démarche:
  - Voyons d'abord ce que sont de « mauvais » pas.
  - Déterminons des règles empêchant les « mauvais » pas.
- Prenons
  - $f(x) = x^2$
  - $x^* = 0$

Choix du pas : approximation

- Algorithme
  - $d_k = (-1)^{k+1}$
  - $\alpha_k = 2+3(2^{-(k+1)})$
  - $x_k=(-1)^k(1+2^{-k})$
- **Lorsque k est grand**
  - $d_k = (-1)^{k+1}$
  - $\alpha_k \approx 2$
  - $x_k \approx (-1)^k$



Choix du pas : approximation

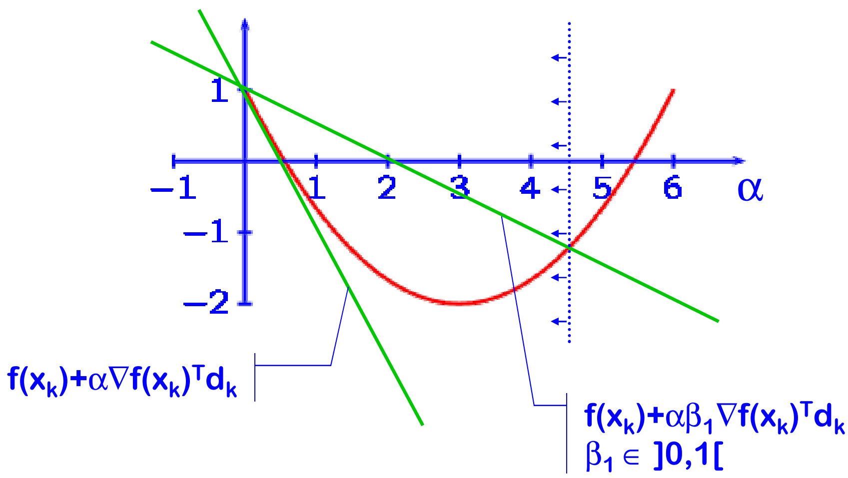
- Algorithme
  - $d_k = (-1)^{k+1}$
  - $\alpha_k = 2+3(2^{-(k+1)})$
  - $x_k=(-1)^k(1+2^{-k})$
- **Lorsque k est grand**
  - $d_k = (-1)^{k+1}$
  - $\alpha_k \approx 2$
  - $x_k \approx (-1)^k$

k	x_k	d_k	α_k	$f(x_k)$
0	2.000	-1	3.500	4.000
1	-1.500	1	2.750	2.250
2	1.250	-1	2.375	1.563
3	-1.125	1	2.188	1.266
4	1.063	-1	2.094	1.129
5	-1.031	1	2.047	1.063
10	1.001	-1	2.001	1.002
999	-1.000	1	2.000	1.000
1000	1.000	-1	2.000	1.000

Choix du pas : approximation

- Problème :
  - très petites diminutions de f relativement à la longueur des pas
- Solution :
  - exiger une diminution suffisante de f

Choix du pas : approximation



Choix du pas : approximation

- On choisit α_k tel que

$$f(x_k+\alpha_k d_k) \leq f(x_k)+\alpha_k\beta_1\nabla f(x_k)^T d_k$$

$$\beta_1 \in \,]0,1[$$

- Reprenons l'exemple (k grand et impair)
- $f(x)=x^2$, $x_k = -1$, $d_k=1$, $\alpha_k=2$, $f'(x)=2x$

$$f(-1+2\bullet 1) \leq 1+2\beta_1(-2 \bullet 1)$$

$$1 \leq 1 - 4\,\beta_1$$

$$4\,\beta_1 \leq 0$$

Impossible

L'algorithme sera rejeté par cette règle

Choix du pas : approximation

- Conditions d'Armijo-Goldstein

$$f(x_k+\alpha_k d_k) \leq f(x_k)+\alpha_k\beta_1\nabla f(x_k)^T d_k$$

$$\beta_1 \in \,]0,1[$$

Choix du pas : approximation

Algorithme de recherche linéaire

- Soient $g(\alpha)$, $\beta_1, \lambda \in]0,1[$, $\alpha_0 > 0$
- Pour k=1,2,…
  - Si $f(x_k+\alpha_k d_k) \leq f(x_k)+\alpha_k \beta_1 \nabla f(x_k)^T d_k$ alors $\alpha^*=\alpha_k$ STOP
  - $\alpha_{k+1} = \lambda \alpha_k$

Convergence

Concept de non orthogonalité

- Supposons que d_k est déterminée de manière unique par x_k.
- On dit que la suite $(d_k)_{k=0,1,\ldots}$ est en relation gradient avec la suite $(x_k)_{k=0,1,\ldots}$ si la propriété suivante est vérifiée :
- Pour toute sous-suite $(x_k)_{k\in\mathcal{K}}$ qui converge vers un point non stationnaire, la sous-suite correspondante $(d_k)_{k\in\mathcal{K}}$ est bornée et vérifie :

$$\limsup_{\{k\to\infty, k\in\mathcal{K}\}} \nabla f(x_k)^T d_k < 0$$

Convergence

Notes :

- On peut souvent garantir a priori que (d_k) est en relation-gradient.
- En particulier, c'est le cas si
  - $d_k = -D_k \nabla f(x_k)$, et
  - les valeurs propres de D_k sont bornées, i.e. pour c1 et c2 > 0, on a

$$c_1 \|z\|^2 \leq z^T D_k z \leq c_2 \|z\|^2, \forall z \in \mathbb{R}^n, k=0,1,\ldots$$

Convergence

Théorème :

- Si (d_k) est en relation-gradient avec (x_k)
- Si le pas est choisi
  - soit par la règle de minimisation
  - soit par la règle d'Armijo-Goldstein
- Alors tous les points limites de (x_k) sont stationnaires.

Critère d'arrêt

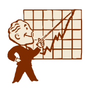
- En général, ces méthodes ne permettent pas de trouver la solution en un nombre fini d'itérations.
- Quand arrête-t-on les itérations ?

Critère 1:

$$\|\nabla f(x_k)\| < \varepsilon, \text{ avec } \varepsilon > 0 \text{ petit.}$$

- Problèmes :
  - Supposons $\varepsilon=10^{-3}$, et $f(x)\in[10^{-7},10^{-5}]$. Il est probable que toutes les valeurs de x vérifieront la condition d'arrêt.
  - Par contre, si $f(x) \in [10^5,10^7]$, cela n'arrivera peut-être jamais.

Critère d'arrêt



Critère 2 :

$$\|r(x)\|_\infty \leq \varepsilon, \text{ avec } \varepsilon > 0 \text{ petit,}$$

avec $r(x)_i = (\nabla f(x)_i \, x_i) / f(x)$.

r(x) est le gradient relatif en x.

Ce critère est indépendant de changement d'unités en f et en x.

Attention si f ou x est proche de 0.

Critère d'arrêt

Critère 3 :

$$\max_{1 \leq i \leq n} \left| \frac{\nabla f(x_k)_i \max(|(x_k)_i|, tx_i)}{\max(|f(x_k)|, tf)} \right| \leq \varepsilon$$

où

- **ε > 0 est petit.**
- **tx_i est une valeur typique de x_i.**
- **tf est une valeur typique de f.**