Schémas Numériques pour l'Optimisation

Master de Mécanique

Rafic YOUNES

Plan

- **Méthodes de descente**
  - **Choix de la direction**
  - **Choix du pas : minimisation / Approximation**
  - **Convergence / Critères d’arrêt**
- Méthode de Newton
  - Résolution d’une équation non linéaire à une inconnue
  - Résolution d’un système d’équations à plusieurs inconnues
- Optimisation par la Méthode de Newton
  - Minimisation
  - Convergence globale
- Méthodes sécantes ou quasi-Newton
  - Minimisation à une seule variable
  - Minimisation à plusieurs variables

Méthodes de descente

- Problème :

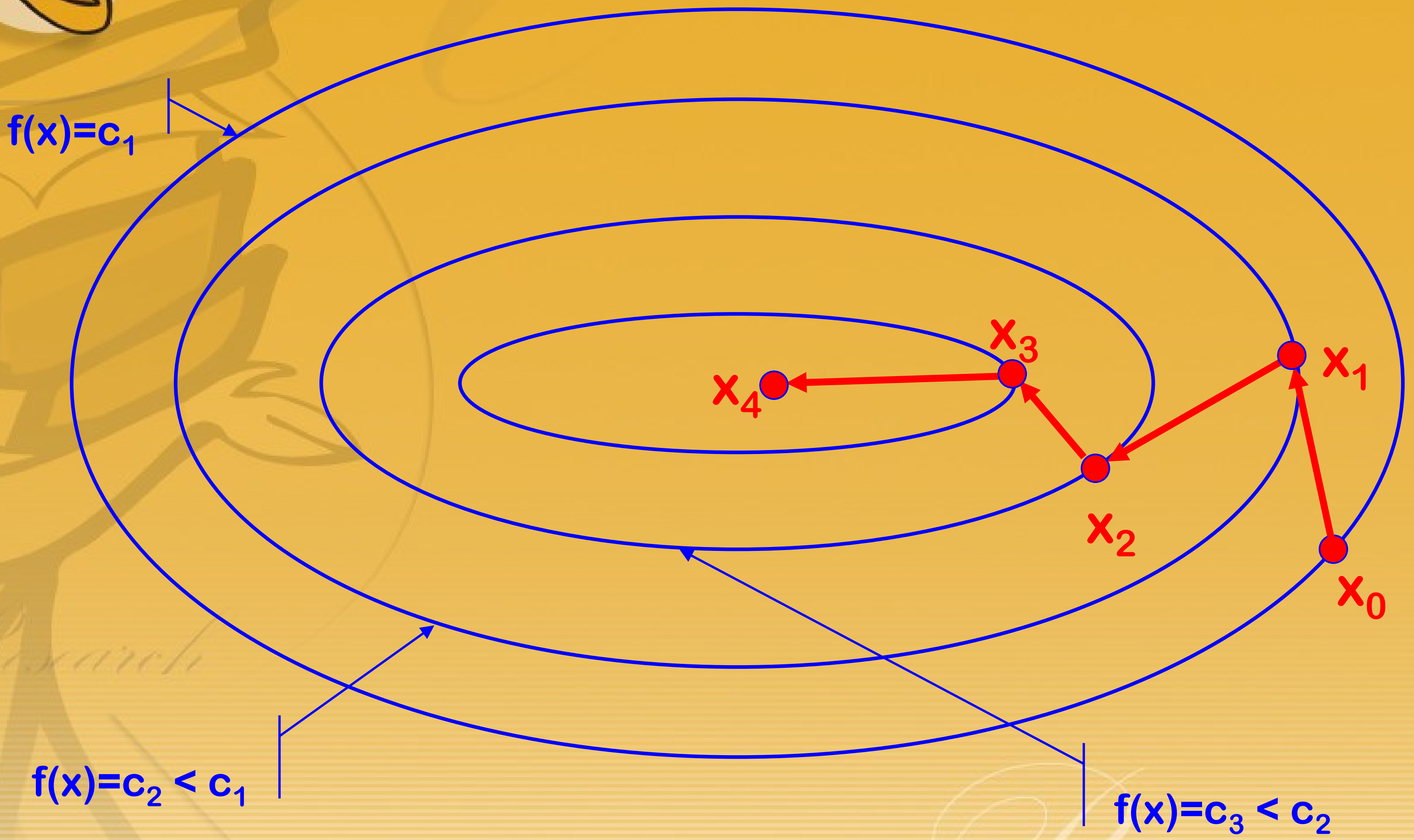
$$\min f : IR^n \rightarrow IR$$

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 :

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

Méthodes de descente



Directions de descente

- Soit $x \in IR^n$ tel que $\nabla f(x) \neq 0$.
- Considérons 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 $\alpha \in \,]0,\delta[$

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

- 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

- Considérons la demi-droite

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

- Théorème de Taylor (1er ordre)

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

avec $\mathbf{s = x_\alpha - x}$

$$\mathbf{f(x_\alpha) = f(x) + \nabla f(x)^T (x_\alpha - x) + o(\|x_\alpha - x\|)}$$

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

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

Directions de descente

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

- 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 $f:IR^n \to IR$, 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

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

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

Choix de la direction

- Beaucoup de choix possibles. On choisit α_k tel que

  $$f(x_k+\alpha_k d_k) < f(x_k)$$

  Aucune garantie de convergence.
- 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 : $d_{ki} = \left(\dfrac{\partial^2 f(x_k)}{\partial x_i^2} \right)^{-1}$

- Méthode de Newton modifiée :

$D_k = (\nabla^2 f(x_0))^{-1}$

  - $x_{k+1} = x_k - \alpha_k (\nabla^2 f(x_0))^{-1} \nabla f(x_k)$

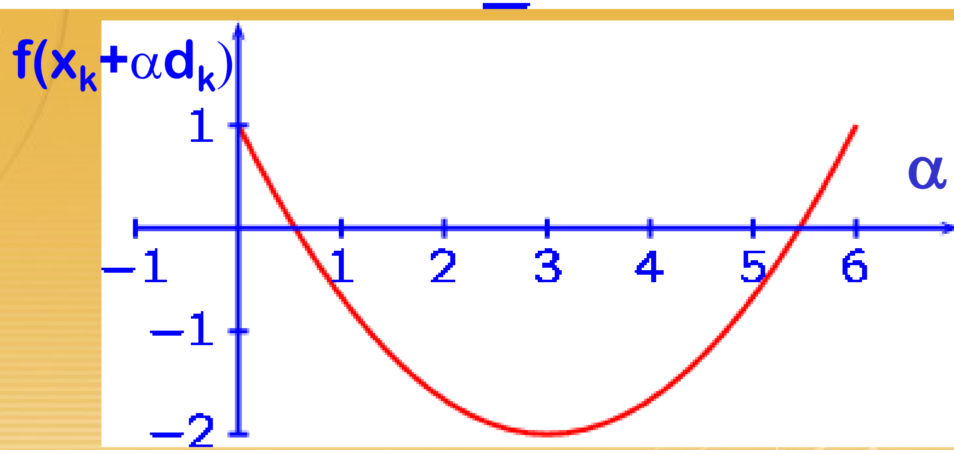
- etc…

Choix du pas : minimisation

Règle de minimisation

- Choisir α_k qui minimise la fonction le long de d_k, c'est-à-dire tel que

$$f(x_k+\alpha_k d_k)=\min_{\alpha\geq 0} f(x_k+\alpha d_k)$$



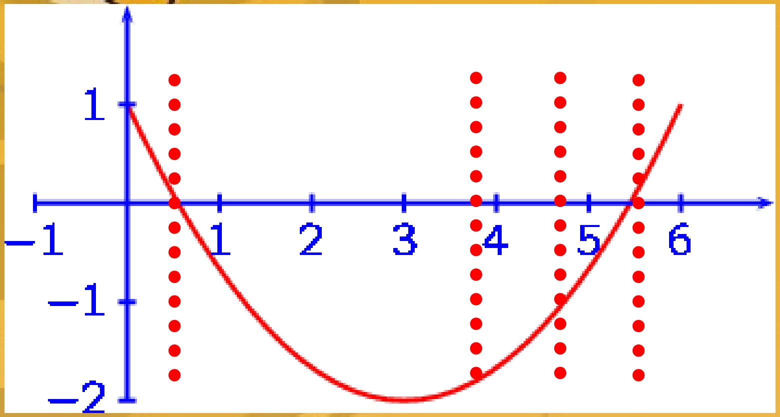
Choix du pas : minimisation

- Minimisation à une variable

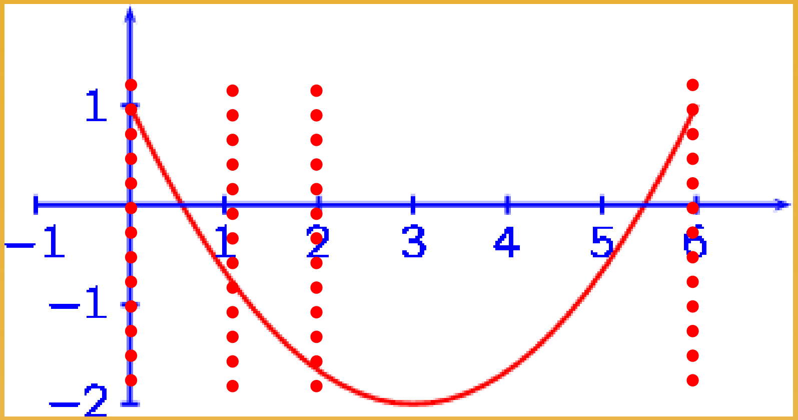
  $$\mathbf{min\ g(\alpha)=f(x_k+\alpha d_k),\ \alpha \in [a_0,b_0]}$$

- Algorithme de la section d'or
- Hypothèse :
  - g est uni-modale dans $[a_0,b_0]$ ssi :
  - g possède un et un seul minimum global α^* dans $[a_0,b_0]$
  - Soient α_1 et $\alpha_2 \in [a_0,b_0]$
    - Si $\alpha_1 < \alpha_2 < \alpha^*$, alors $g(\alpha_1) > g(\alpha_2) > g(\alpha^*)$
    - Si $\alpha^* < \alpha_1 < \alpha_2$, alors $g(\alpha^*) < g(\alpha_1) < g(\alpha_2)$

Choix du pas : minimisation



a_0 a'_0 b'_0 b_0



a_0 a'_0 b'_0 b_0

Si $f(a'_0) < f(b'_0)$ Alors

$\alpha^* \in [a_0, b'_0]$

$[a_1, b_1] = [a_0, b'_0]$

Si $f(a'_0) > f(b'_0)$ Alors

$\alpha^* \in [a'_0, b_0]$

$[a_1, b_1] = [a'_0, b_0]$

Choix du pas : minimisation

Algorithme de la section d'or

a_0 $\quad$ a'_0 $\quad$ b'_0 $\quad$ b_0

a_1 $\quad$ a'_1 $\quad$ b'_1 $\quad$ b_1

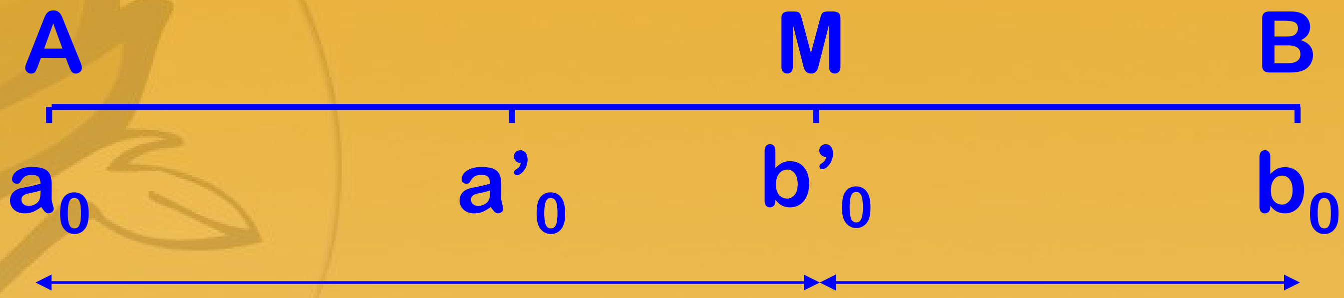
- Itération 0 : $a'_0 - a_0 = b_0 - b'_0 = \rho(b_0 - a_0)$
- Itération 1: $a'_1 - a_1 = b_1 - b'_1 = \rho(b_1 - a_1)$

  $a'_1 - a_0 = b'_0 - a'_0 = \rho(b'_0 - a_0)$
- On montre que

$$\rho = (3-\sqrt{5})/2 \approx 0.382$$

Choix du pas : minimisation

- Notes : anciens Grecs
- Section d'or :



$(1-\rho)(b_0-a_0)=(1-\rho)\lambda \qquad \rho(b_0-a_0)=\rho\lambda$

$$\frac{AB}{AM}=\frac{AM}{MB} \qquad \frac{1}{1-\rho}=\frac{1-\rho}{\rho}$$

Choix du pas : minimisation

Algorithme de la section d'or

- Soient $g(\alpha)$ unimodale sur $[a_0,b_0]$, $\varepsilon \in \mathbb{R}$ et $\rho = (3-\sqrt{5})/2$

Pour k=1,2,…

  - Si $b_k-a_k < \varepsilon$, alors $\alpha^*=(a_k+b_k)/2$ STOP
  - $a'_k = a_k+\rho(b_k-a_k)=(1-\rho)a_k+\rho b_k$
  - $b'_k = b_k-\rho(b_k-a_k) = \rho a_k+(1-\rho)b_k$
  - Si $g(a'_k)\leq g(b'_k)$, alors $a_{k+1}=a_k$, $b_{k+1}=b'_k$
  - Si $g(a'_k)> g(b'_k)$, alors $a_{k+1}=a'_k$, $b_{k+1}=b_k$

Choix du pas : minimisation

Exemple :

- min $g(\alpha)=\alpha^4 - 14\alpha^3 + 60\alpha^2 - 70\alpha$

□ $\alpha \in [0,2]$

k	a_k	a'_k	b'_k	b_k	α^*	$g(a'_k)$	$g(b'_k)$
1	0.0000	0.7639	1.2361	2.0000	1.0000	-24.3607	-18.9582
2	0.0000	0.4721	0.7639	1.2361	0.6180	-21.0985	-24.3607
3	0.4721	0.7639	0.9443	1.2361	0.8541	-24.3607	-23.5925
4	0.4721	0.6525	0.7639	0.9443	0.7082	-23.8374	-24.3607
5	0.6525	0.7639	0.8328	0.9443	0.7984	-24.3607	-24.2879
6	0.6525	0.7214	0.7639	0.8328	0.7426	-24.2579	-24.3607
7	0.7214	0.7639	0.7902	0.8328	0.7771	-24.3607	-24.3669
8	0.7639	0.7902	0.8065	0.8328	0.7984	-24.3669	-24.3495
9	0.7639	0.7802	0.7902	0.8065	0.7852	-24.3696	-24.3669
10	0.7639	0.7740	0.7802	0.7902	0.7771	-24.3681	-24.3696

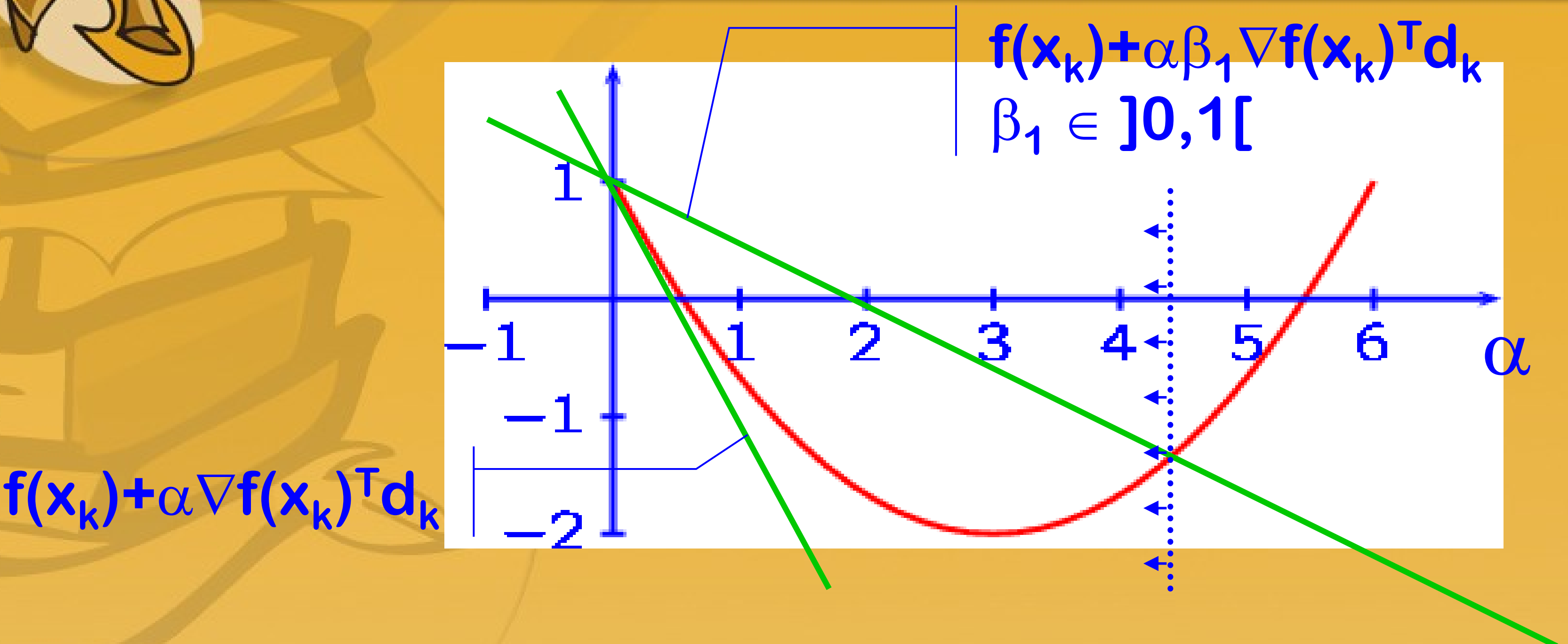
Choix du pas

Règle de minimisation

- Méthode de la section d’or
- Minimisation prend du temps
- Section d’or nécessite l’uni-modalité
- A-t-on besoin du minimum exact ?
- Idée : Règle d’approximation

Choisir un pas qui diminue suffisamment la valeur de la fonction.

Choix du pas : approximation



- Conditions d'Armijo-Goldstein

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

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

Choix du pas : approximation

Algorithme de recherche linéaire

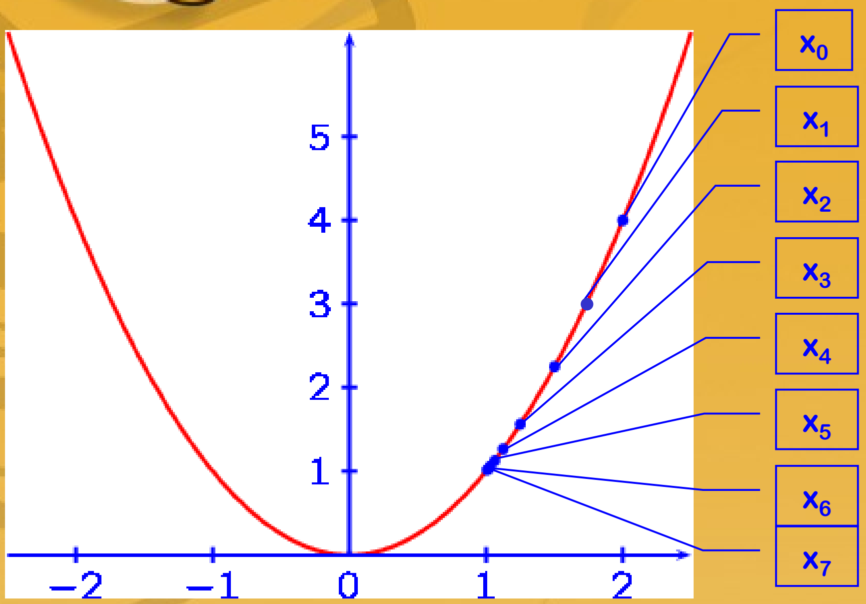
- Soient $g(\alpha)$, $\beta_1, \lambda \in]0:1[$, $\beta_2 \in]\beta_1:1[$, $\alpha_0 > 0$

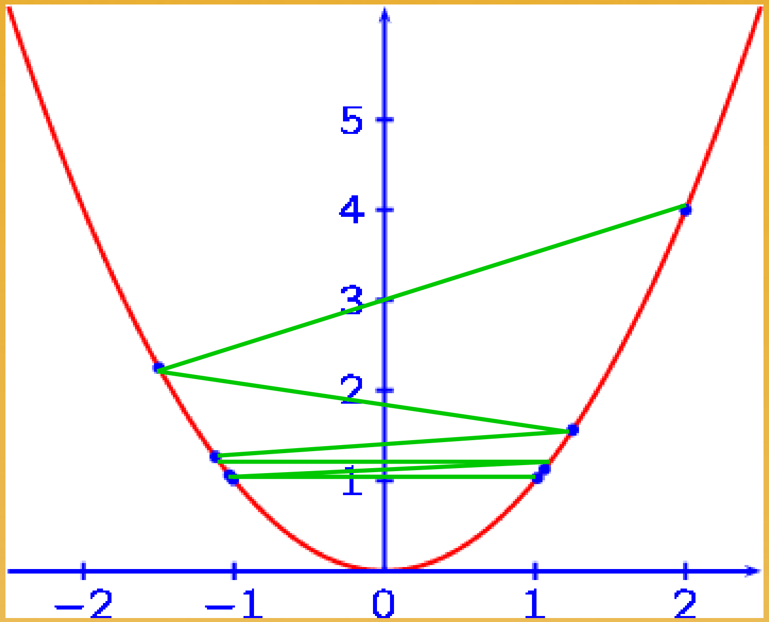
- Pour k=1,2,…
  - Si $f(x_k+\alpha_k d_k) \leq f(x_k)+\alpha_k \beta_1 \nabla f(x_k)^T d_k$ alors $\alpha^*=\alpha_k$ STOP
  - $\alpha_{k+1} = \lambda\ \alpha_k$

Convergence / Critères d’arrêt

- Démarche:
  - Voyons d’abord ce que sont de «mauvais» pas.
  - Déterminons des règles empêchant les « mauvais » pas.
- Prenons
  - $f(x) = x^2$
  - $x_0 = 0$

Convergence / Critères d'arrêt





- **Lorsque k est grand**
  - $d_k = -1$
  - $\alpha_k \approx 0$
  - $x_k \approx 1$

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

Convergence / Critères d'arrêt

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.

Convergence / Critères 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.

Convergence / Critères d'arrêt

Critère 2 :

$||r(x)||_\infty \leq \varepsilon$, avec $\varepsilon > 0$ 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 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 $\varepsilon > 0$ est petit. t_{xi} est une valeur typique de x_i. t_f est une valeur typique de f.

Vitesse de convergence

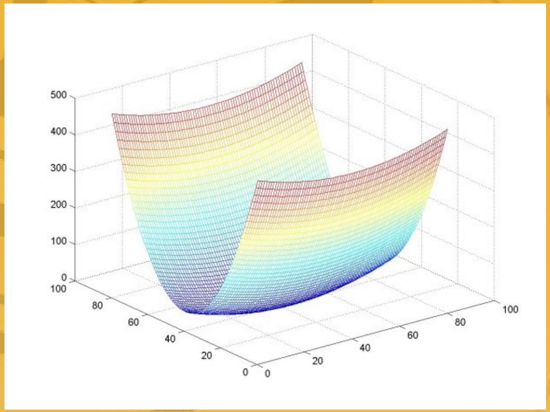
Méthode de la plus forte pente

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$

- Supposons que
  - $f(x) = \frac{1}{2} x^T Q x$ est quadratique
  - Q est définie positive
  - f est minimisé par $x^* = 0$
  - $f(x^*)=0$
- Donc :
  - $\nabla f(x) = Qx$ $\qquad\qquad \nabla^2 f(x) = Q$
  - $x_{k+1} = x_k - \alpha_k Qx$

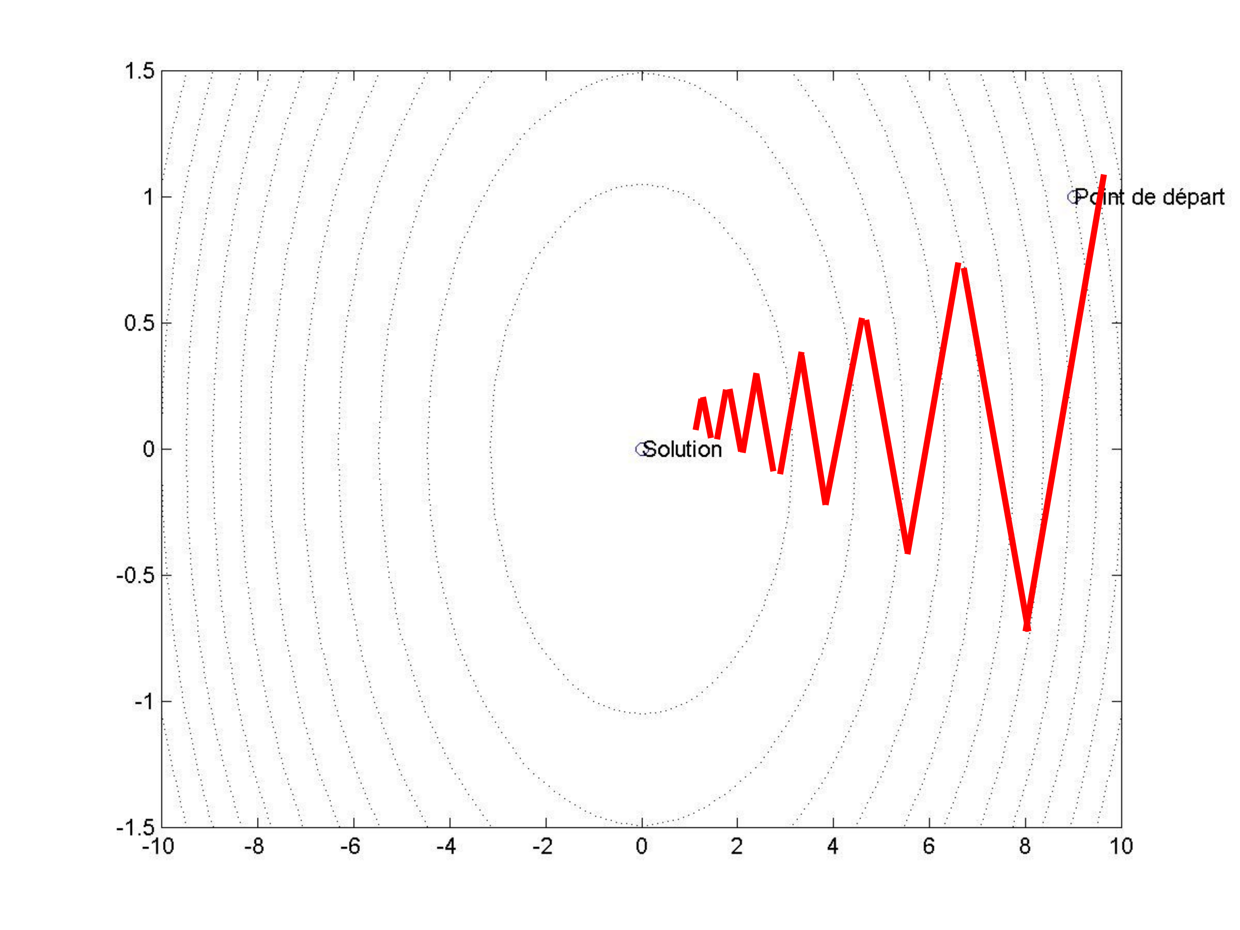
Vitesse de convergence

$$f(x,y)=\frac{1}{2}\cdot x^2+\frac{9}{2}\cdot y^2$$



La direction du gradient est perpendiculaire aux courbes de niveaux

La représentation n'est pas à l'échelle



Plan

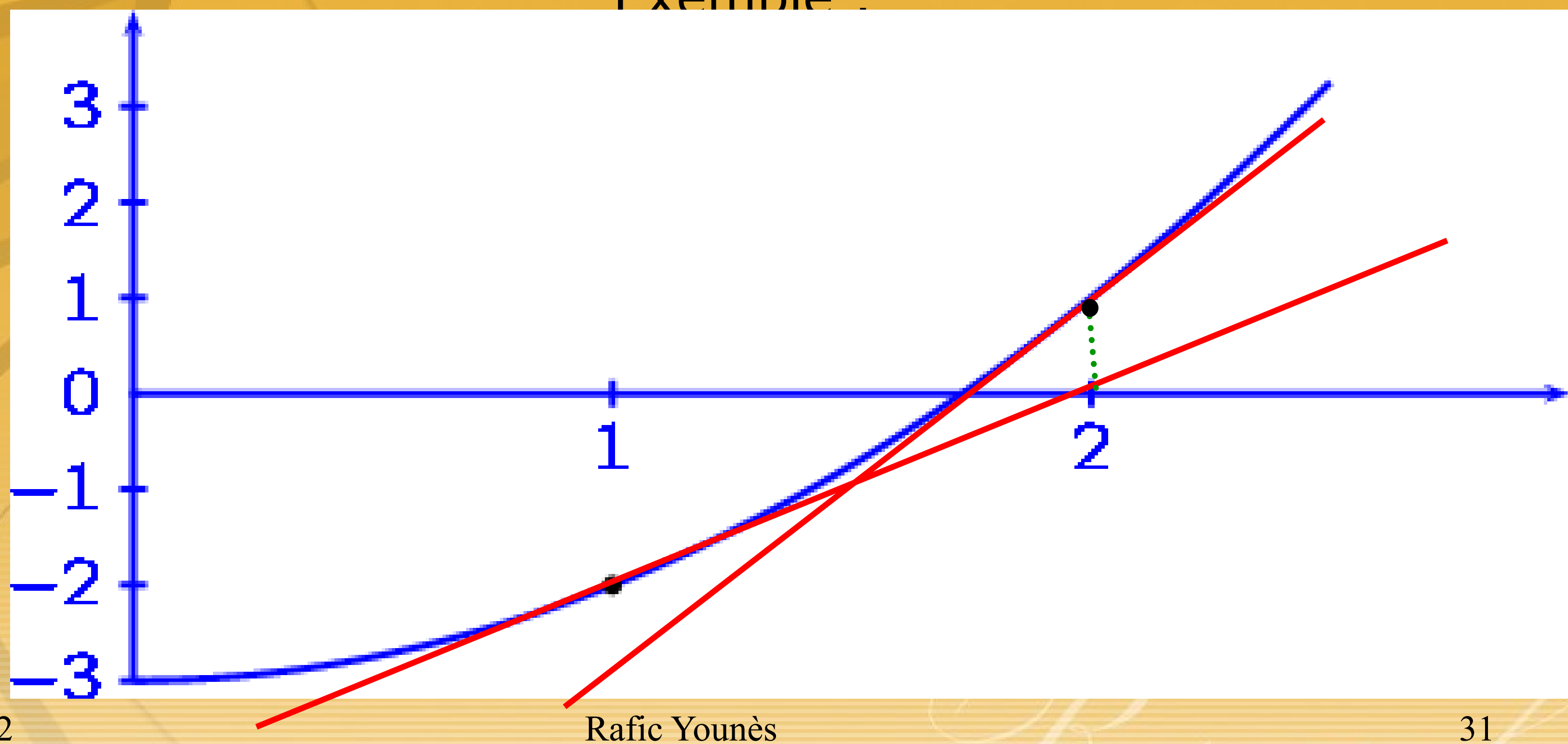
- Méthodes de descente
  - Choix de la direction
  - Choix du pas : minimisation / Approximation
  - Convergence / Critères d’arrêt
- **Méthode de Newton**
  - Résolution d’une équation non linéaire à une inconnue
  - Résolution d’un système d’équations à plusieurs inconnues
- Optimisation par la Méthode de Newton
  - Minimisation
  - Convergence globale
- Méthodes sécantes ou quasi-Newton
  - Minimisation à une seule variable
  - Minimisation à plusieurs variables

Equation à une inconnue

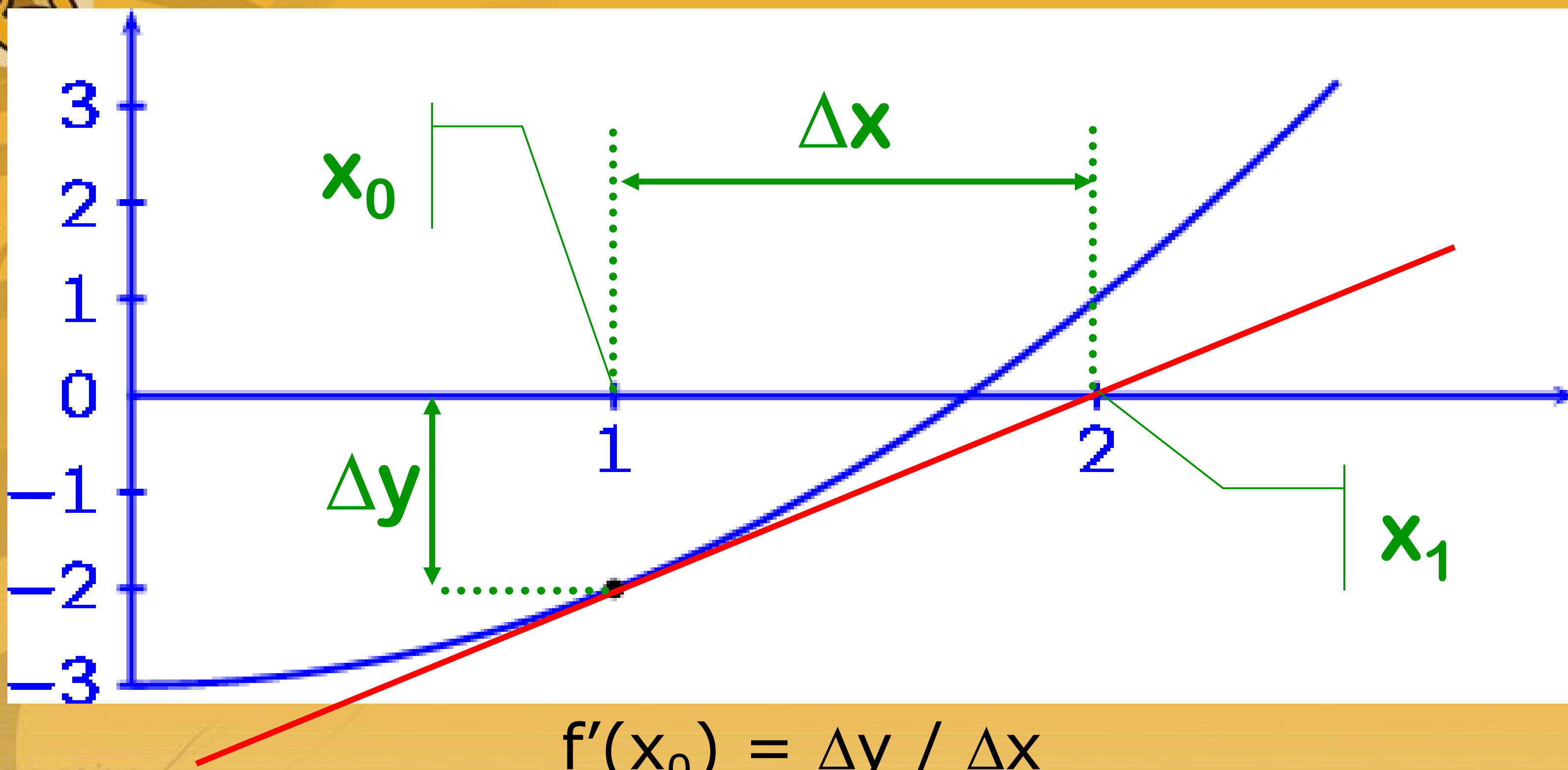
Problème :

- Soit f:IR→IR continûment dérivable. Trouver x tel que f(x) = 0.

Exemple :



Equation à une inconnue



$$f'(x_0) = \Delta y / \Delta x$$

$$\Delta x = \Delta y / f'(x_0) = - f(x_0) / f'(x_0)$$

$$x_1 = x_0 - f(x_0) / f'(x_0)$$

Equation à une inconnue

- Méthode de Newton :

$$x_{k+1} = x_k - f(x_k) / f'(x_k)$$

k	x	f(x)	f'(x)	err
0	1.00000000	-2.00000000	2.00000000	0.73205081
1	2.00000000	1.00000000	4.00000000	0.26794919
2	1.75000000	0.06250000	3.50000000	0.01794919
3	1.73214286	0.00031888	3.46428571	0.00009205
4	1.73205081	0.00000001	3.46410162	0.00000000

Equation à une inconnue

- A chaque itération, on remplace la fonction non-linéaire par un modèle local facile à calculer.

$$M(x) = f(x_0) + f'(x_0)(x-x_0)$$

- Ce modèle est : linéaire, $M(x_0) = f(x_0)$, tel que $M'(x_0) = f'(x_0)$

Convergence locale

  - Si f(x) n'est pas trop non linéaire
  - Si x_0 est suffisamment bon
  - Alors convergence rapide

Si $f'(x^*)=0$, la méthode de Newton converge beaucoup plus lentement.

Equation à une inconnue

$x^* = 1$

$f_1'(x^*)=2$

$f_2'(x^*)=0$

Note : Lorsque f est linéaire, la méthode de Newton pour la résolution d'équations converge en une seule itération.

$f_1(x)=x^2-1$	k	$f_2(x)=x^2-2x+1$
2	0	2
1.25	1	1.5
1.025	2	1.25
1.000304878	3	1.125
1.000000046	4	1.0625
1	5	1.03125
1	6	1.015625
1	7	1.0078125
1	8	1.00390625
1	9	1.001953125
1	10	1.000976563

Système d’équations à plusieurs variables

- La méthode de Newton peut être utilisée pour résoudre un système de n équations à n inconnues.

$$g(x) = 0$$

où $g:\mathbb{R}^n \to \mathbb{R}^n$ est une fonction continûment différentiable.

La méthode est alors

$$x_{k+1} = x_k - (\nabla g(x_k)^T)^{-1}\, g(x_k)$$

Système d'équations à plusieurs variables

1) Il existe δ tel que, si $x_0 \in S_\delta$, la suite (x_k) générée par

$$x_{k+1} = x_k - (\nabla g(x_k)^T)^{-1}\, g(x_k)$$

- est bien définie,
- est contenue dans S_δ,
- converge vers x^*.

Système d’équations à plusieurs variables

Notes :

- L est la constante de Lipschitz
- La condition

$$\|\nabla g(x)-\nabla g(y)\| \leq L\|x-y\|$$

  impose que la fonction « ne soit pas trop non linéaire »
- La condition

$$\|(\nabla g(x)^T)^{-1}\| \leq M$$

  impose que le problème « ne soit pas trop mal conditionné »

Plan

- Méthodes de descente
  - Choix de la direction
  - Choix du pas : minimisation / Approximation
  - Convergence / Critères d’arrêt
- Méthode de Newton
  - Résolution d’une équation non linéaire à une inconnue
  - Résolution d’un système d’équations à plusieurs inconnues
- **Optimisation par la Méthode de Newton**
  - **Minimisation**
  - **Convergence globale**
- Méthodes sécantes ou quasi-Newton
  - Minimisation à une seule variable
  - Minimisation à plusieurs variables

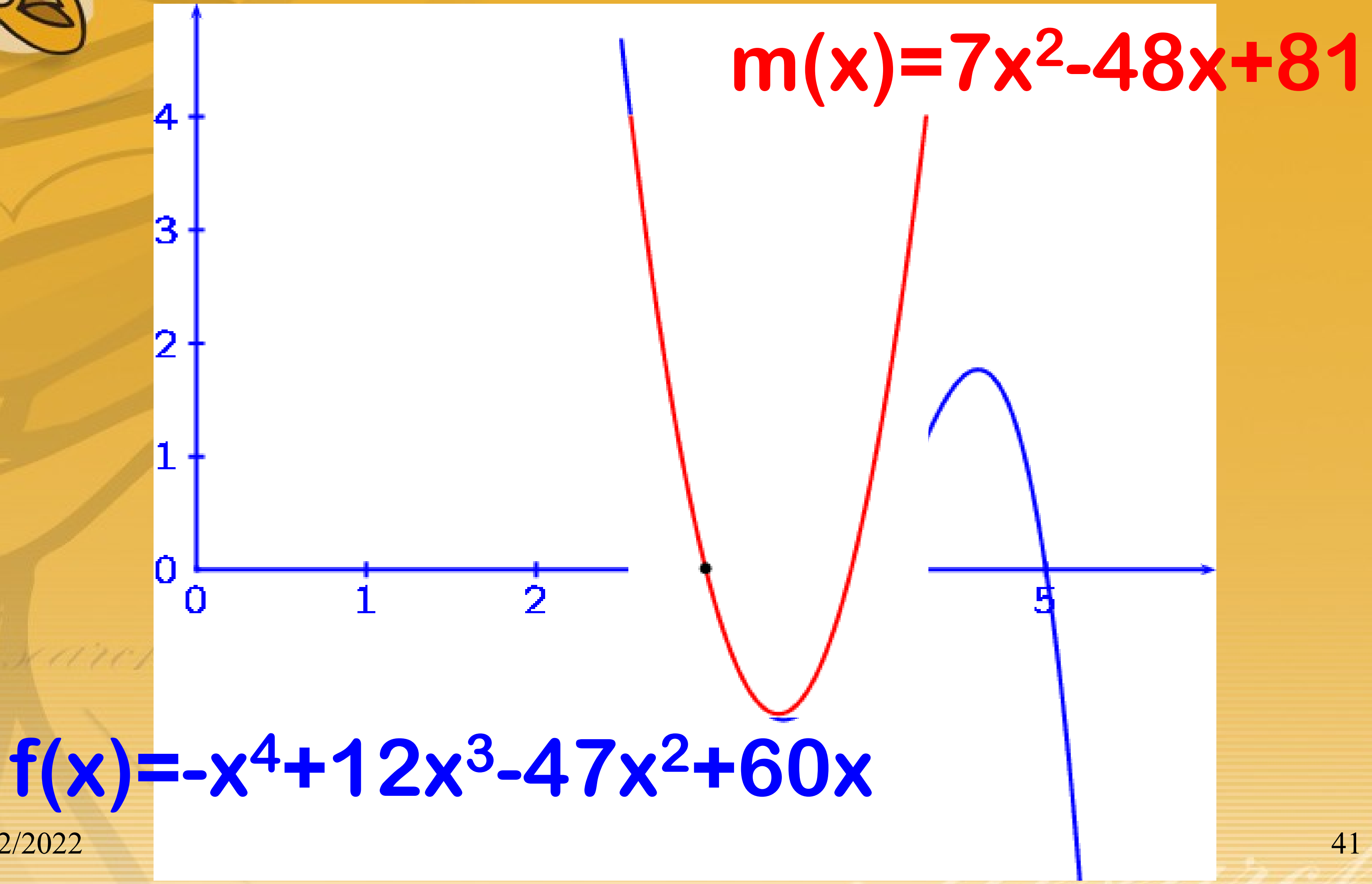
Minimisation à une variable

- Appliquer la méthode de Newton à l'équation
$$f'(x) = 0.$$
- Cela donne : $\mathbf{x_{k+1} = x_k - f'(x_k)/f''(x_k)}$
- A chaque itération, on remplace la fonction non-linéaire par un modèle local facile à calculer.

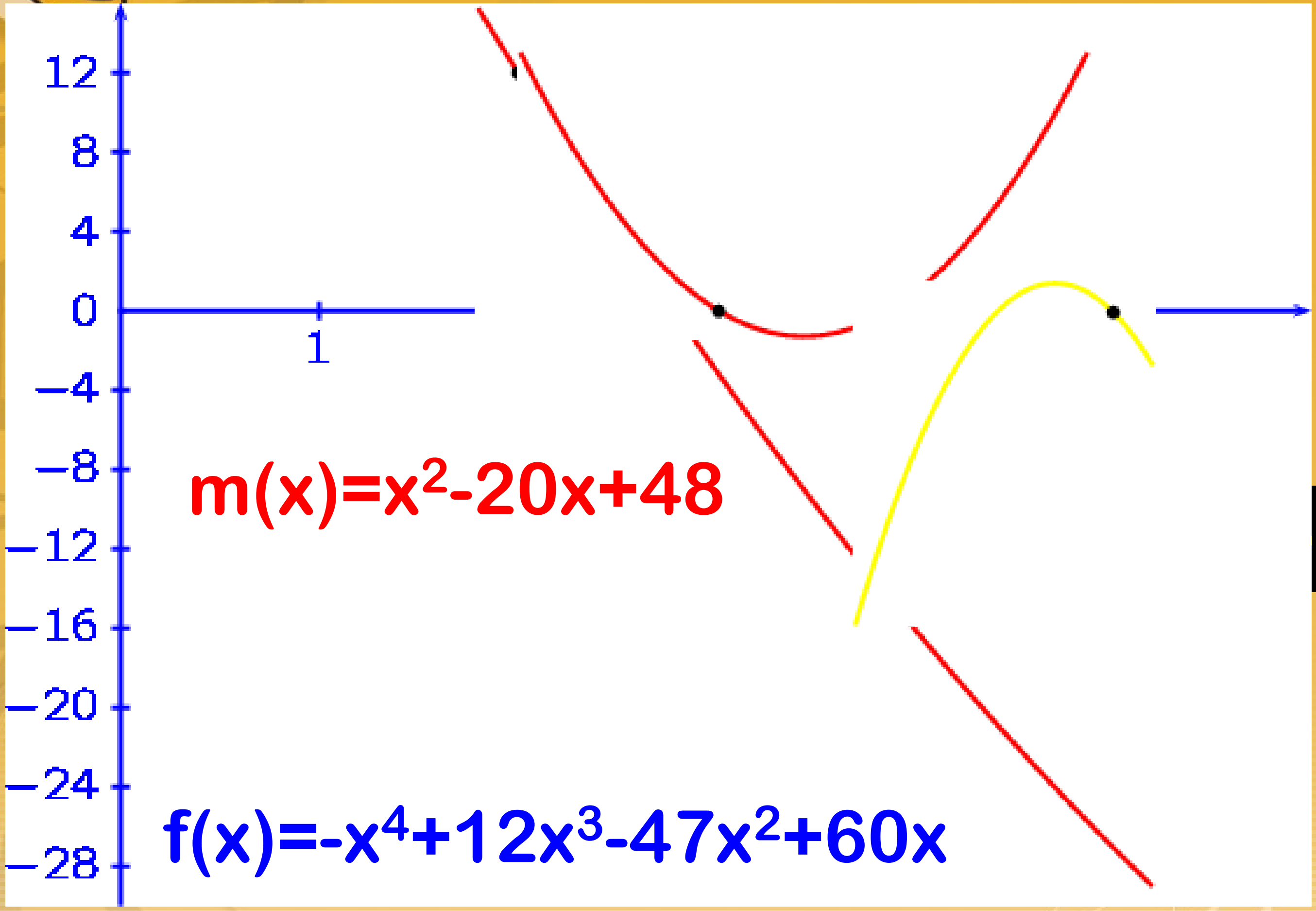
$$\mathbf{m_k(x) = f(x_k) + f'(x_k)(x-x_k)+\tfrac{1}{2}\, f''(x_k)(x-x_k)^2}$$

- Ce modèle est : quadratique, $m_k(x_k) = f(x_k)$, $m_k'(x_k) = f'(x_k)$ et $m_k''(x_k) = f''(x_k)$

Minimisation à une variable



Minimisation à une variable



160x-375

Minimisation multi- variables

- Appliquer la méthode de Newton à l'équation

$$\nabla f(x_k) = 0$$

- Cela donne : $\mathbf{x_{k+1} = x_k - \nabla^2 f(x_k)^{-1} \nabla f(x_k)}$
- A chaque itération, on remplace la fonction non-linéaire par un modèle local facile à calculer.

$$\mathbf{m_k(x) = f(x_k) + \nabla f(x_k)(x-x_k) + \tfrac{1}{2}\, \nabla^2 f(x_k)(x-x_k)^2}$$

- Ce modèle est : quadratique, $m_k(x_k) = f(x_k)$, $m_k'(x_k) = f'(x_k)$ et $m_k''(x_k) = f''(x_k)$

Convergence globale

Problèmes de la méthode de Newton pure :

- La matrice $\nabla^2 f(x_k)$ peut ne pas être définie positive.
- Elle peut ne pas être inversible.
- La méthode peut produire des itérés tels que $f(x_{k+1}) > f(x_k)$.
- La méthode se contente de résoudre $\nabla f(x)=0$. Elle peut donc converger vers des points critiques qui ne sont pas des minima.

Convergence globale

Idée :

- **modifier la méthode de Newton pour garantir la convergence globale,**
- **mais conserver sa forme pure près de la solution afin de maintenir sa rapidité.**

- **N.B. : Un algorithme globalement convergent est un algorithme qui converge vers un minimum local à partir de (presque) n’importe quel point de départ x_0.**

Convergence globale

Idée de Cauchy :

- Lorsque le pas de Newton n'est pas défini, ou ne fait pas diminuer la valeur de la fonction, préférer la direction de la plus forte pente.
- A. Cauchy (1847) *Analyse mathématique. Méthode générale pour la résolution des systèmes d'équations simultanées.* Comptes Rendus de l'Académie des Sciences de Paris.
- Désavantages de la plus forte pente.

Convergence globale

Recherche linéaire

- Newton « pur » :

$$x_{k+1} = x_k - \nabla^2 f(x_k)^{-1} \nabla f(x_k)$$

- Considérer la direction :

$$d_k = - \nabla^2 f(x_k)^{-1} \nabla f(x_k)$$

- Si $\nabla^2 f(x_k)^{-1}$ n'est pas déf. pos. :

$$d_k = - (\nabla^2 f(x_k) + \lambda I)^{-1} \nabla f(x_k)$$

Tel que d_k est direction de descente

Convergence globale

Recherche linéaire

- Algorithme de descente :

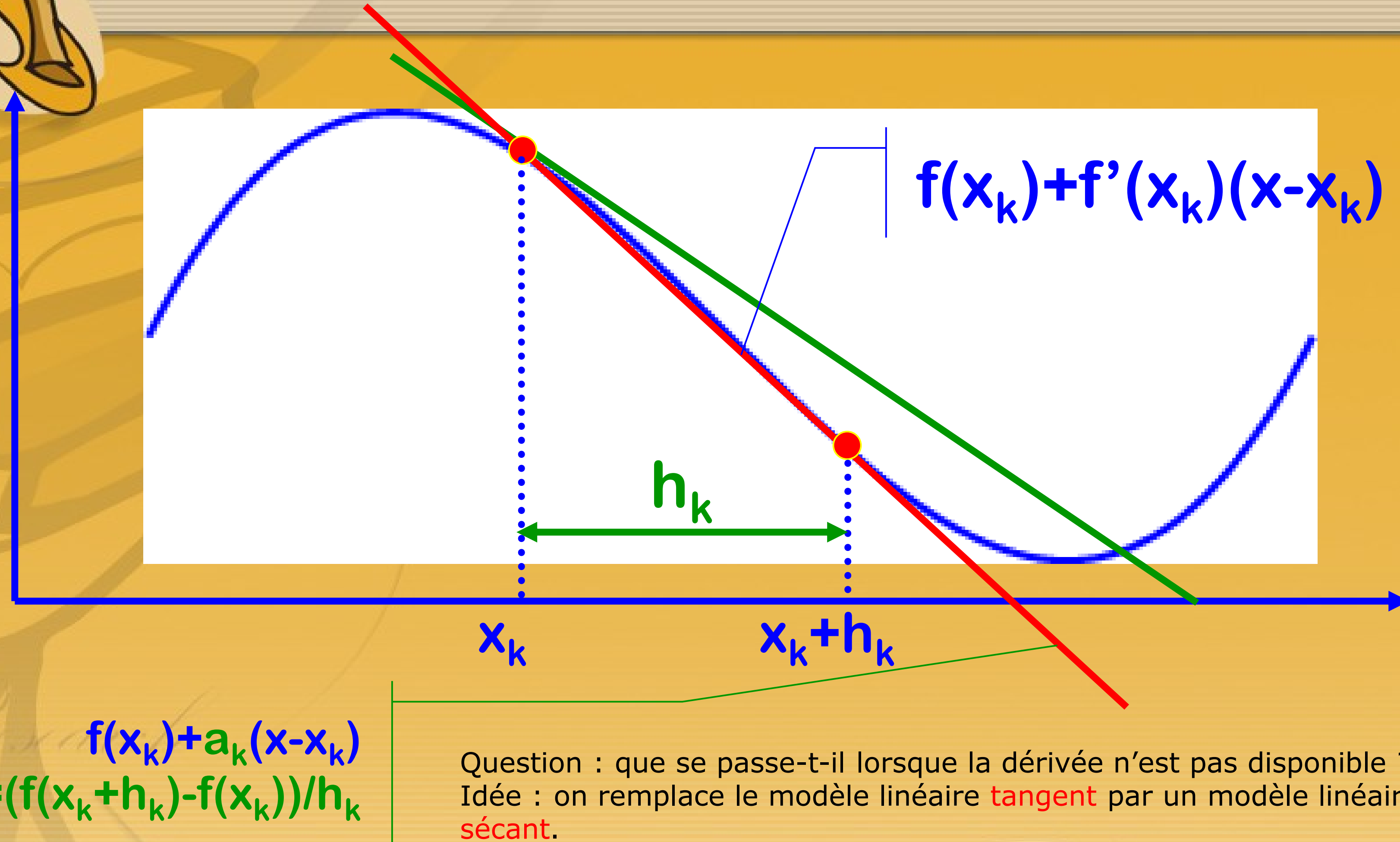
$$x_{k+1} = x_k + \alpha_k d_k$$

- On essaie d'abord $\alpha_k=1$.
- Si cela ne marche pas, on essaie des pas plus courts.
- Règle d'Armijo-Goldstein

Plan

- Méthodes de descente
  - Choix de la direction
  - Choix du pas : minimisation / Approximation
  - Convergence / Critères d’arrêt
- Méthode de Newton
  - Résolution d’une équation non linéaire à une inconnue
  - Résolution d’un système d’équations à plusieurs inconnues
- Optimisation par la Méthode de Newton
  - Minimisation
  - Convergence globale
- **Méthodes sécantes ou quasi-Newton**
  - **Minimisation à une seule variable**
  - **Minimisation à plusieurs variables**

Méthodes quasi-Newton



Question : que se passe-t-il lorsque la dérivée n'est pas disponible ?
Idée : on remplace le modèle linéaire tangent par un modèle linéaire sécant.

Méthodes quasi-Newton

- Pente de la sécante : $a_k = \dfrac{f(x_k+h_k)-f(x_k)}{h_k}$
- Le pas « quasi-Newton » est

$$x_{k+1} = x_k - f(x_k)/a_k$$

- Comment choisir h_k ?
- Si $h_k \to 0$, alors $a_k \to f'(x_k)$
- Si h_k est choisi suffisamment petit, a_k est appelée une approximation de $f'(x_k)$ par différence finie.
- cela exige une évaluation supplémentaire de la fonction. On préfère choisir :

$$h_k = x_{k-1} - x_k$$

Méthodes quasi-Newton

Notes :

- La méthode fonctionne bien. Le modèle linéaire :
- $$M_k(x)=f(x_k) + \frac{f(x_{k-1})-f(x_k)}{x_{k-1}-x_k}(x-x_k)$$

Vérifie : $M_k(x_{k-1})=f(x_{k-1})$ et $M_k(x_k)=f(x_k)$.

Minimisation à plusieurs variables

- On remplace : $F(x)$ par $\nabla f(x)$ & $\nabla F(x)$ par $\nabla^2 f(x)$

Attention : le hessien doit être symétrique défini positif.

$$x_{k+1} = x_k - \alpha_k H_k \nabla f(x_k)$$

Méthodes quasi-Newton

Mise à jour de Powell

- $d_k = x_k - x_{k-1}$ $\quad y_k = \nabla f(x_k) - \nabla f(x_{k-1})$
- On génère la suite suivante :
- $H_{2i+1} = H_{2i} + \dfrac{y_k - H_{2i} d_k}{d_k^T d_k} d_k^T$
- $H_{2i+2} = ½ \, (H_{2i+1} + H_{2i+1}^T)$

Méthodes quasi-Newton

- Mise à jour suivante :

$$H_k = H_{k-1} + \frac{y_k y_k^T}{y_k^T d_k} - \frac{H_{k-1} d_k d_k^T H_{k-1}}{d_k^T H_{k-1} d_k}$$

$$d_k = x_k - x_{k-1} \qquad y_k = \nabla f(x_k) - \nabla f(x_{k-1})$$

Broyden- Fletcher-Goldfarb-Shanno

Méthodes quasi-Newton

Résumé : $x_{k+1} = x_k - \alpha_k D_k \nabla f(x_k)$

- Plus forte pente : $D_k = I$

- Newton : $D_k = \nabla^2 f(x_k)^{-1}$
  - Newton pure :

    $\alpha_k = 1$ -> pas globalement convergent

  - Règle d’approximation : Armijo-Goldstein

Méthodes quasi-Newton

Résumé :

- $x_{k+1} = x_k - \alpha_k D_k \nabla f(x_k)$
- Newton : $D_k = - \left(\nabla^2 f(x_k) + \lambda I\right)^{-1}$
  - Newton pure :

    $\alpha_k = 1$ -> pas globalement convergent
  - Règle d'approximation : Armijo-Goldstein

Méthodes quasi-Newton

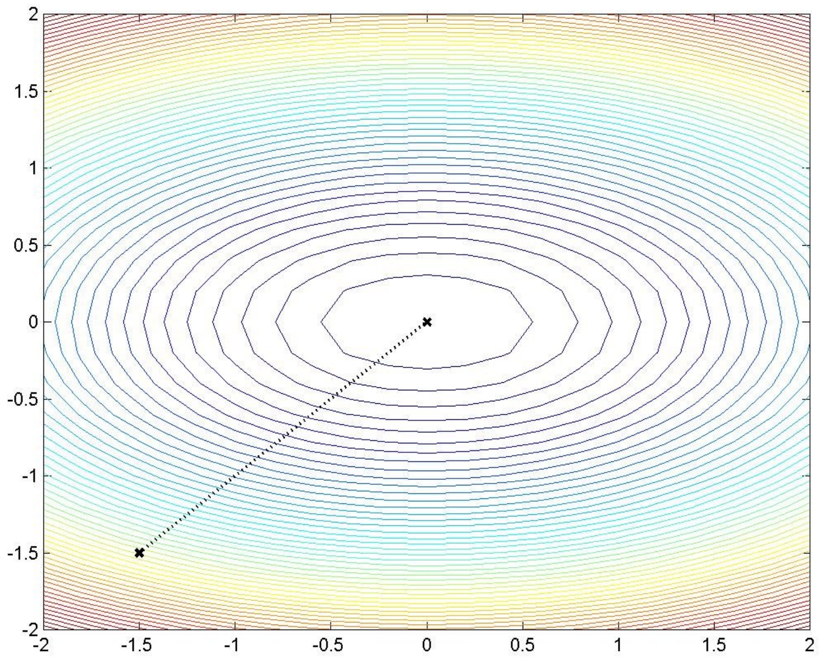
Résumé :

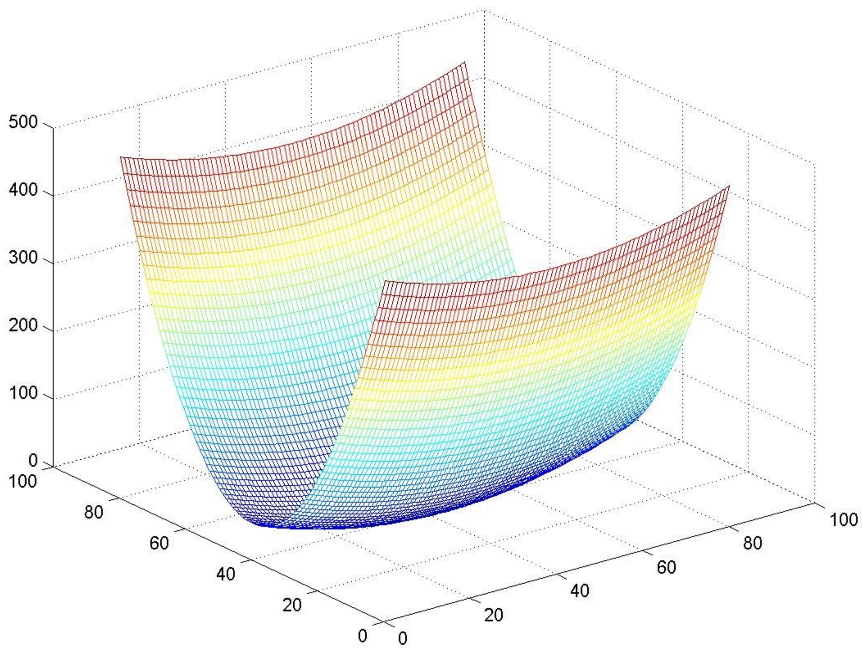
- $x_{k+1} = x_k - \alpha_k D_k \nabla f(x_k)$
- quasi-Newton : $D_k = H_k$
  - H_0 arbitraire, symétrique définie positive
  - $H_k = H_{k-1} + \dfrac{y_k y_k^T}{y_k^T d_k} - \dfrac{H_{k-1} d_k d_k^T H_{k-1}}{d_k^T H_{k-1} d_k}$
  - Règle d'approximation : Armijo-Goldstein

Exemples : quadratique

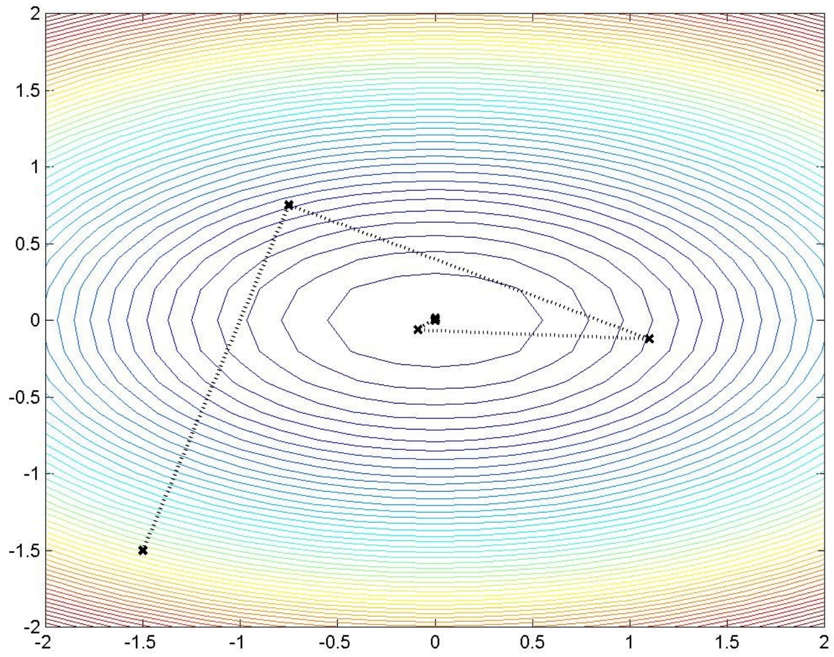
Newton

$$f(x) = \frac{1}{2}x_1^2 + \frac{9}{2}x_2^2$$

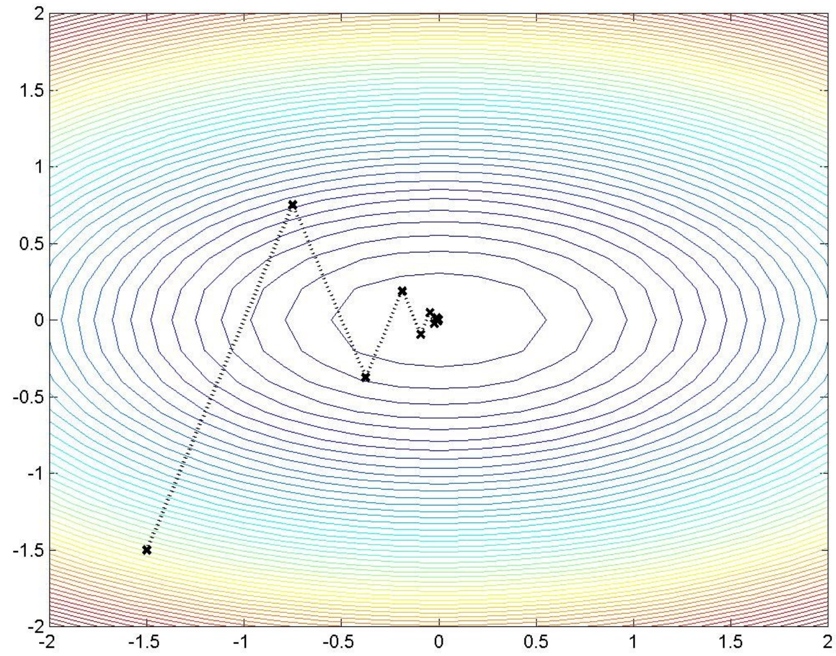




BFGS avec H_0=I

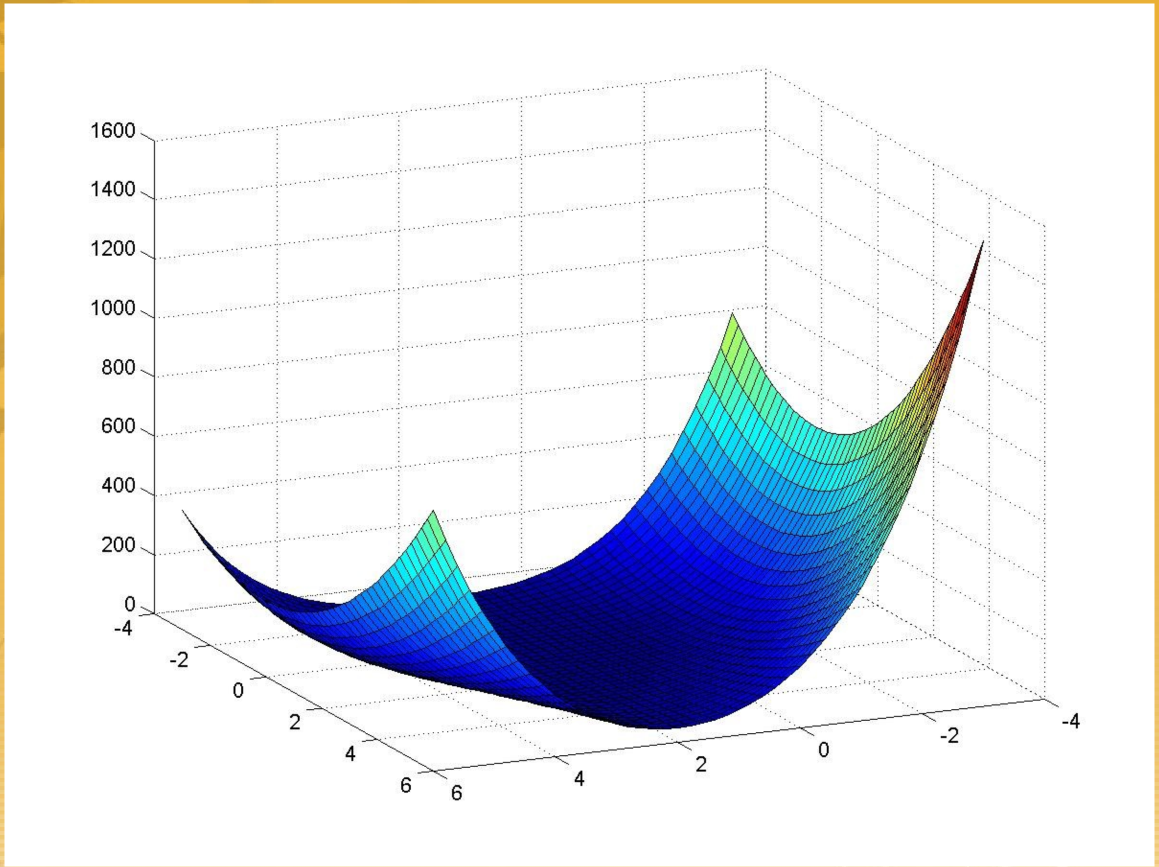


Plus forte pente

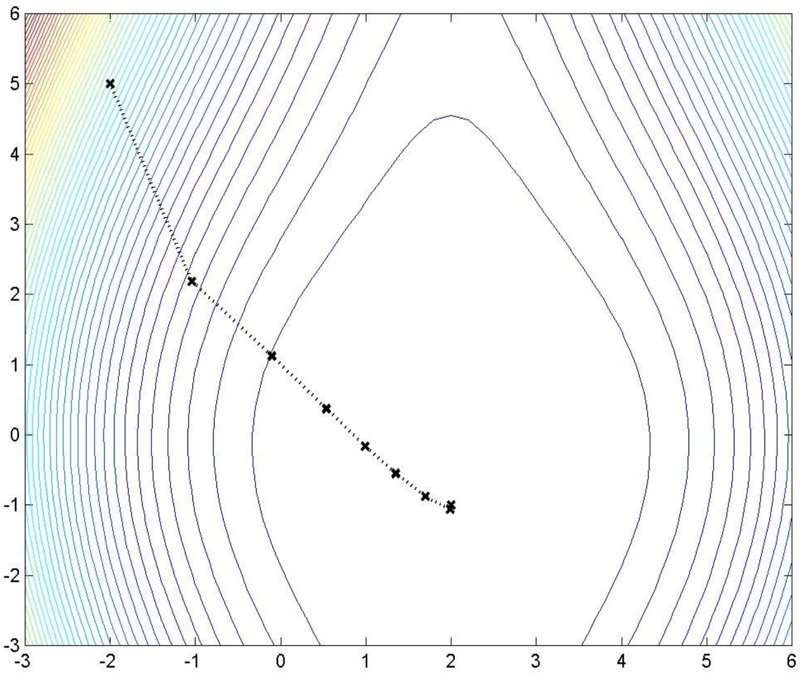


Exemples : fonction «gentille»

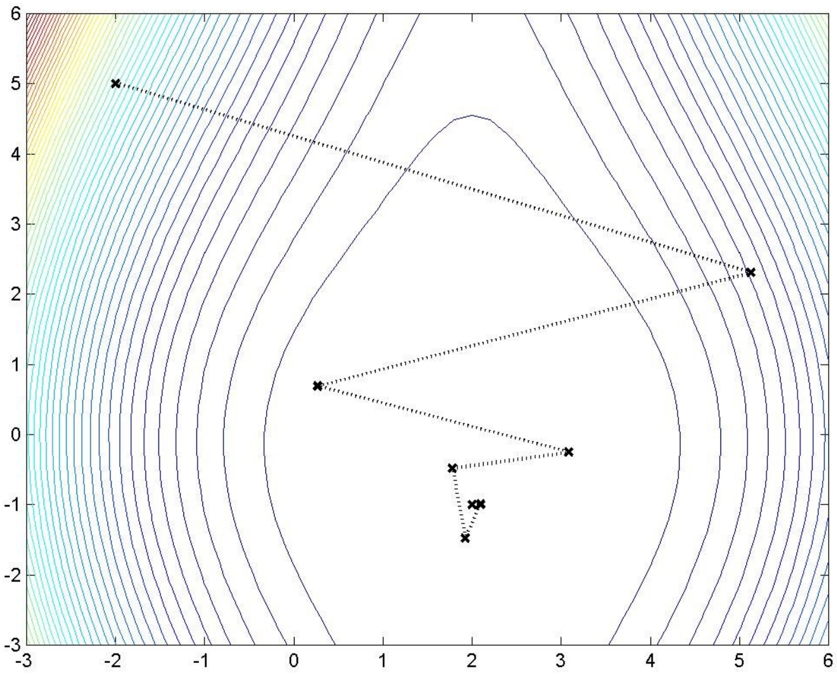
$$f(x) = (x_1 - 2)^4 + (x_1 - 2)^2 x_2^2 + (x_2 + 1)^2$$



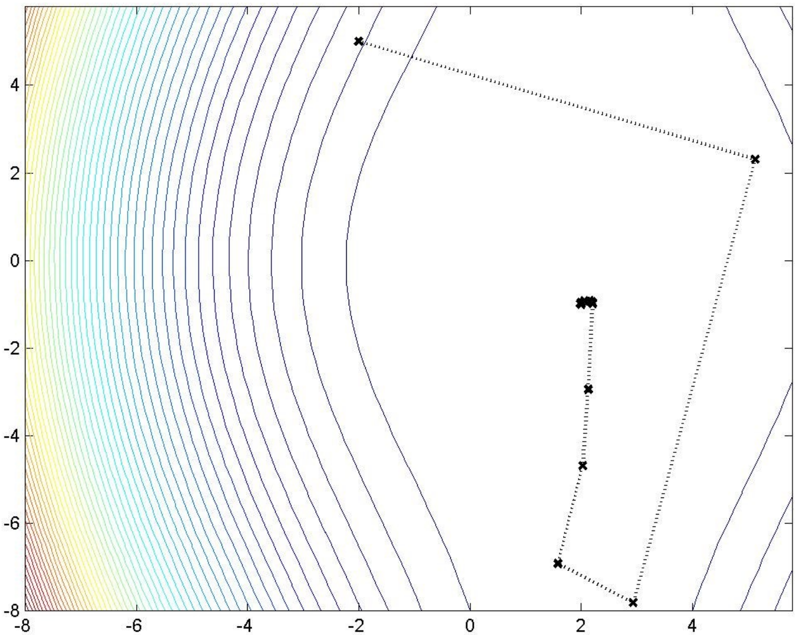
Newton



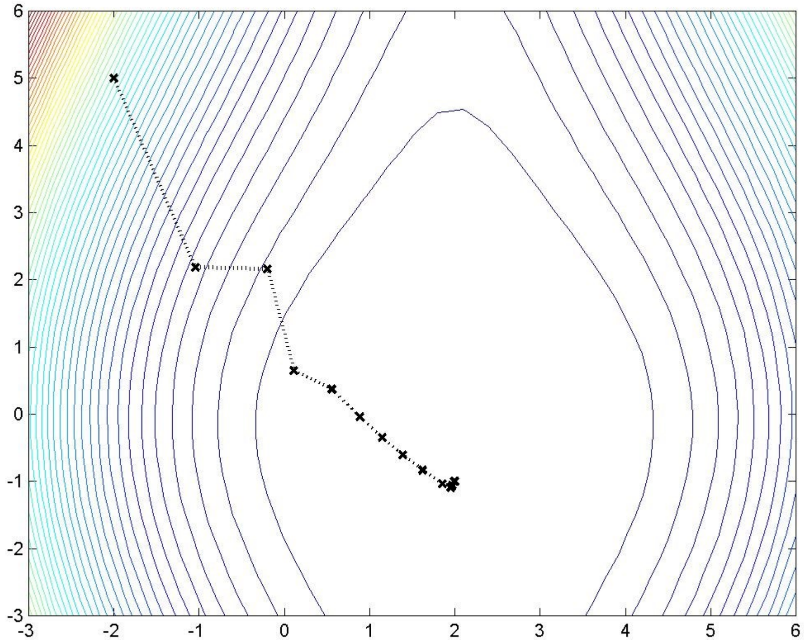
Plus forte pente



BFGS avec $H_0=I$ (plus large)

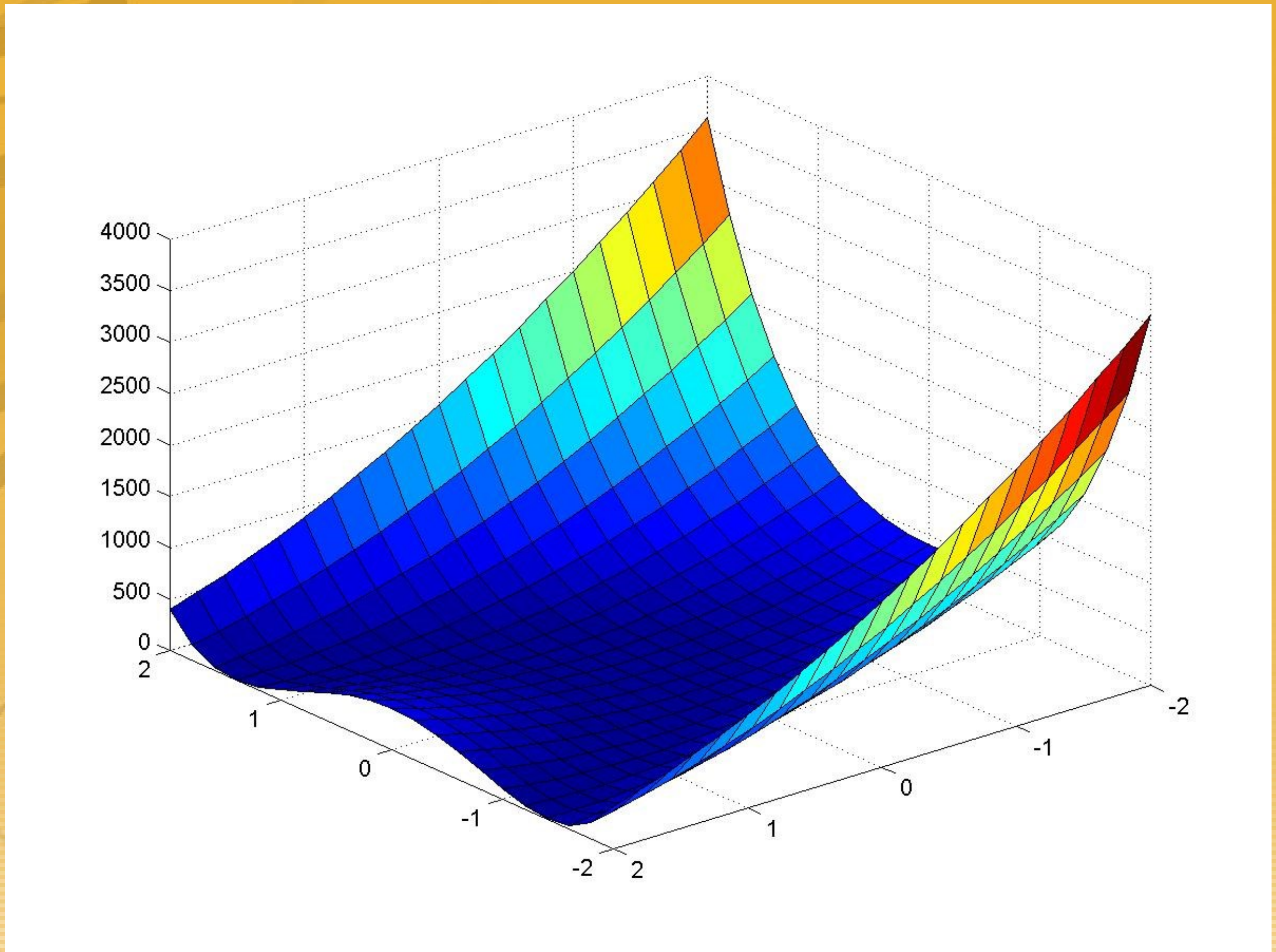


BFGS avec $H_0=\nabla^2 f(x_0)$

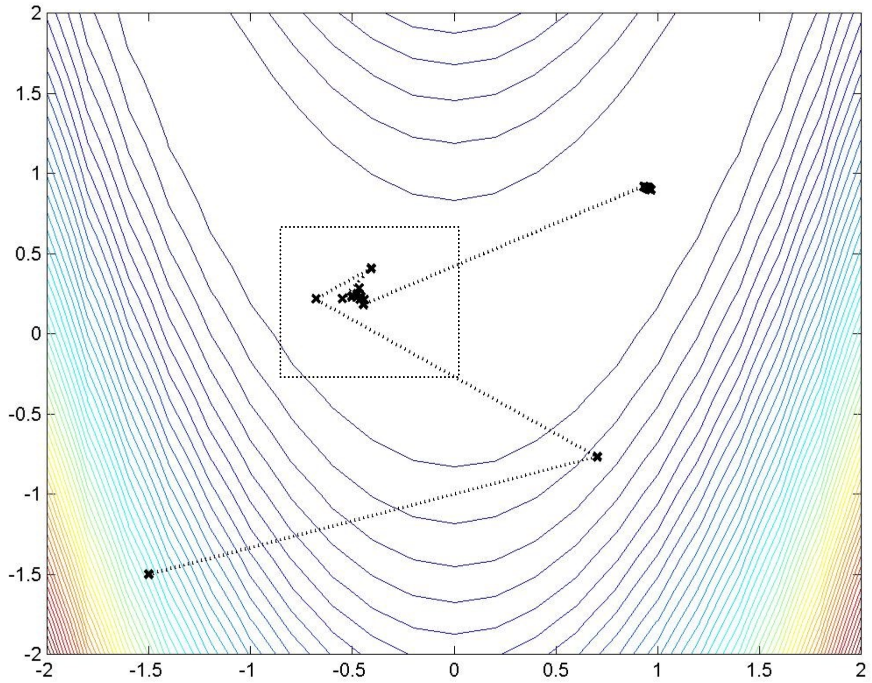


Exemples : Rosenbrock ou « fonction banane »

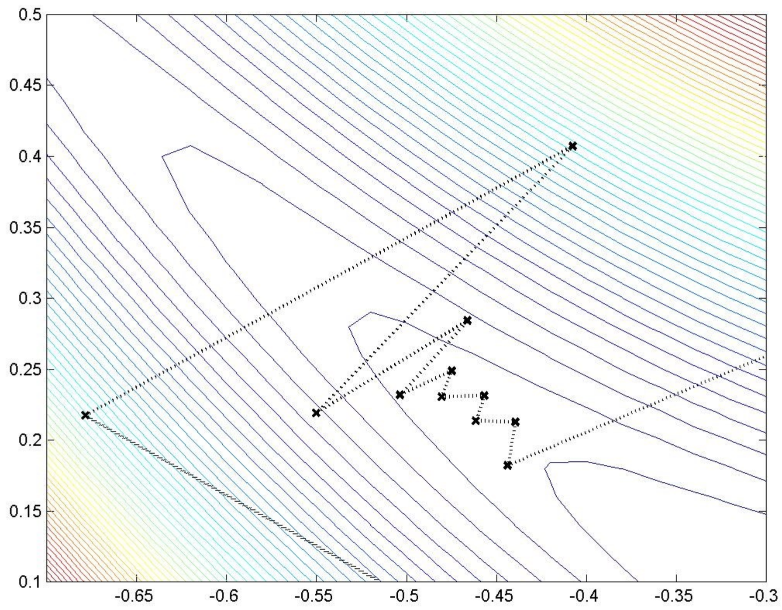
$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$



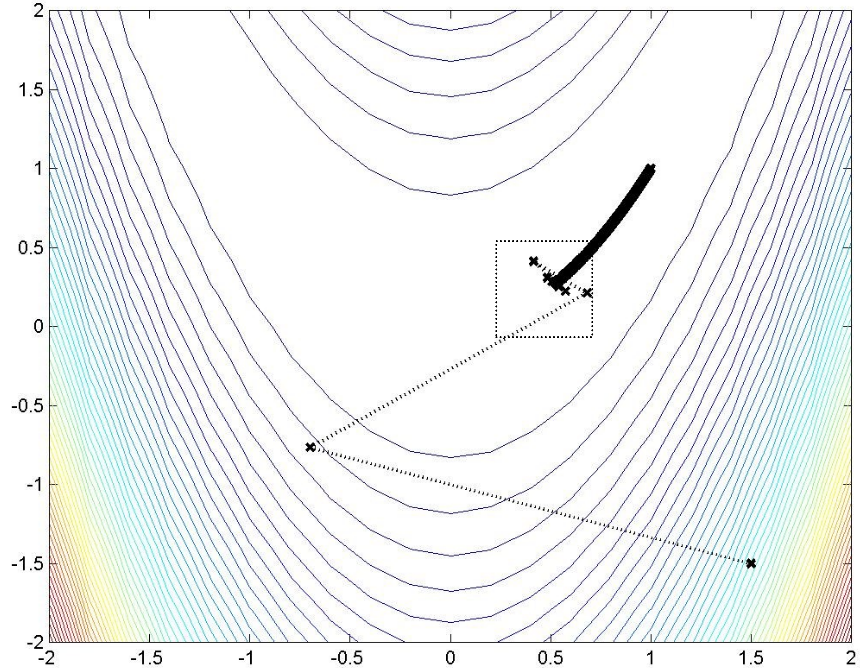
Plus forte pente



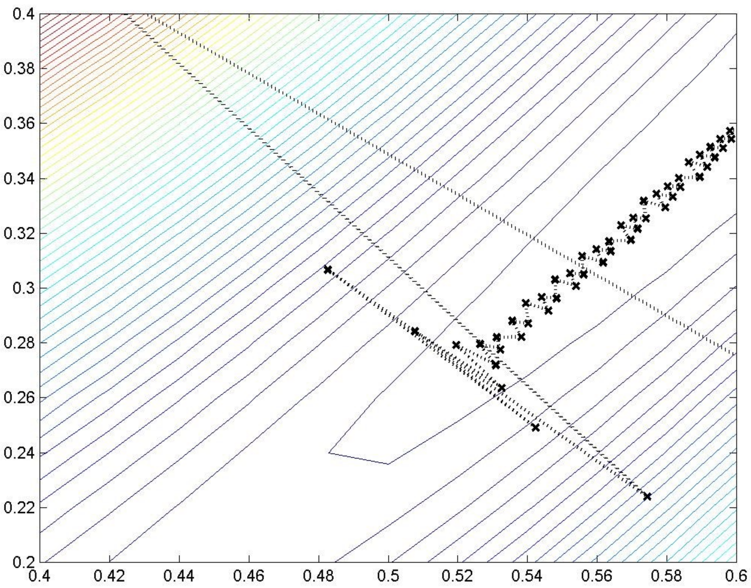
Plus forte pente (zoom)



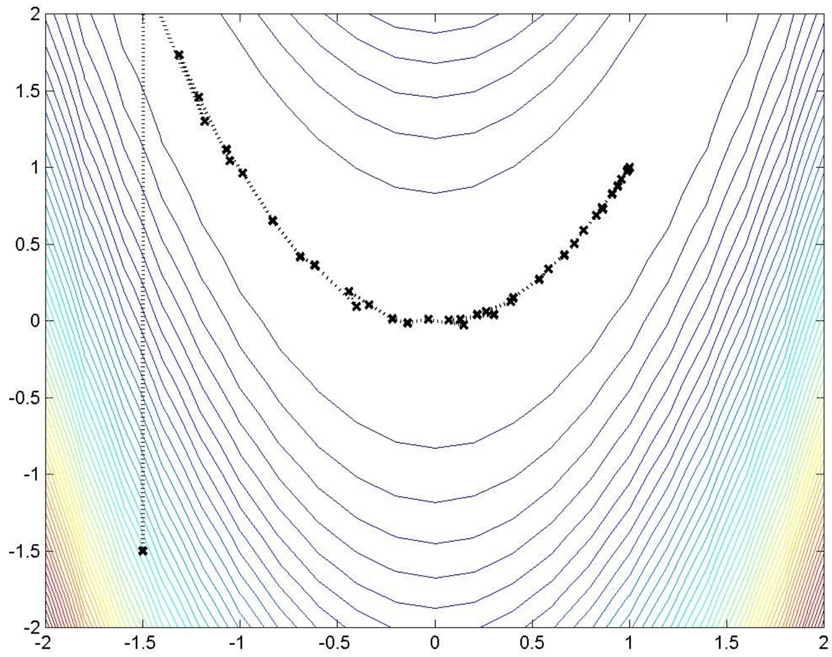
Plus forte pente (autre point de départ)

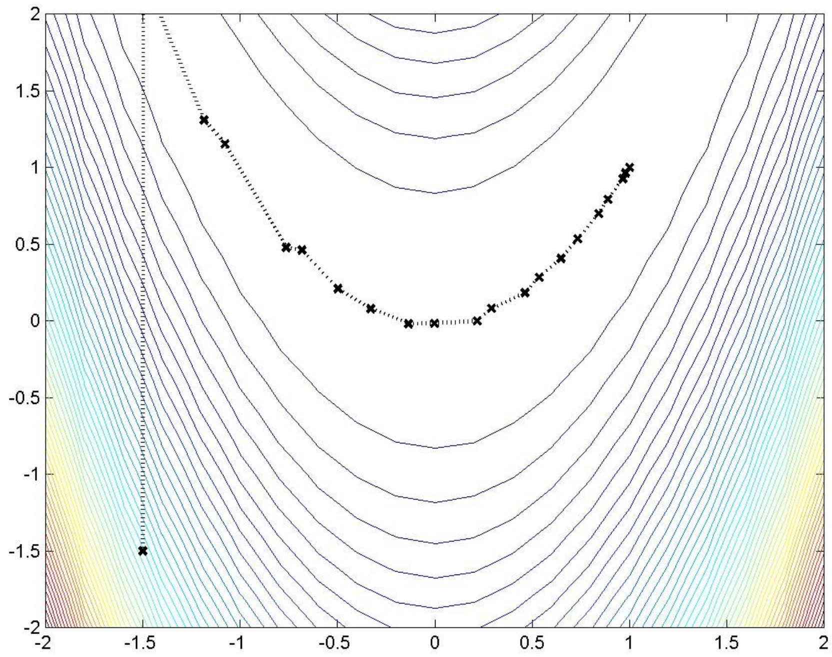


Plus forte pente (zoom)

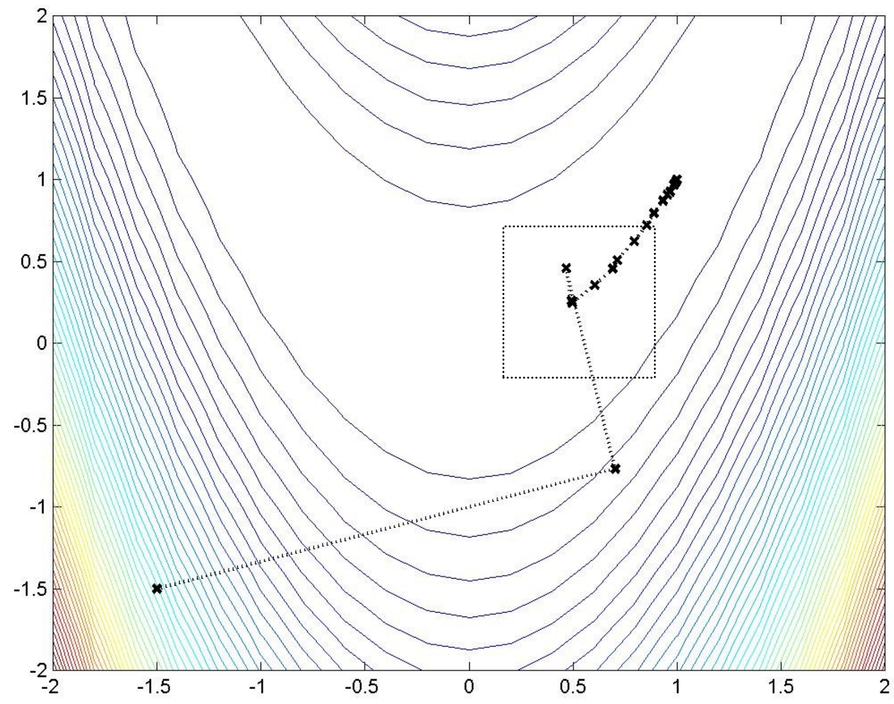


BFGS avec $H_0=\nabla^2 f(x_0)$ Newton





BFGS avec H_0=I



BFGS avec H_0=I (zoom)