ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Optimisation non linéaire sans contraintes

Recherche opérationnelle

GC-SIE

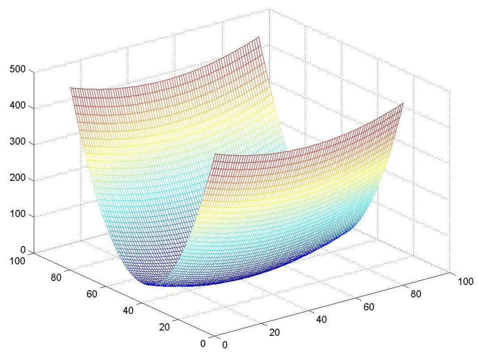
Variations sur Newton

Variations sur Newton

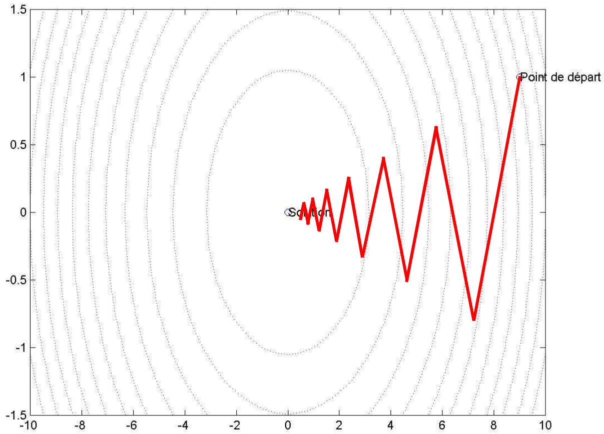
- Convergence de la méthode de la plus forte pente
- Résolution d'une équation non linéaire à une inconnue
- Résolution d'un système d'équations à plusieurs inconnues
- Convergence globale
- Méthodes sécantes ou quasi-Newton

Vitesse de convergence

- Exemple : f(x,y) = $x^2 + \frac{1}{2} y^2 \quad \frac{9}{2}$



Vitesse de convergence



Vitesse de convergence

Notes :

- La direction du gradient est perpendiculaire aux courbes de niveaux.
- La représentation précédente n'est pas à l'échelle.

Variations sur Newton

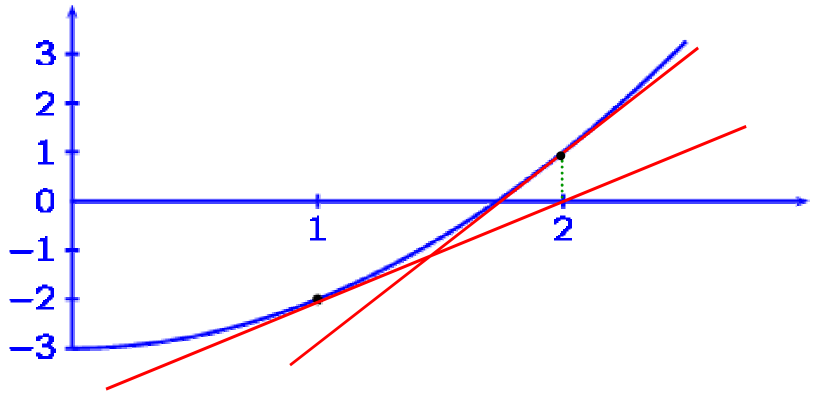
- Convergence de la méthode de la plus forte pente
- Résolution d'une équation non linéaire à une inconnue
- Résolution d'un système d'équations à plusieurs inconnues
- Convergence globale
- Méthodes sécantes ou quasi-Newton

Equation à une inconnue

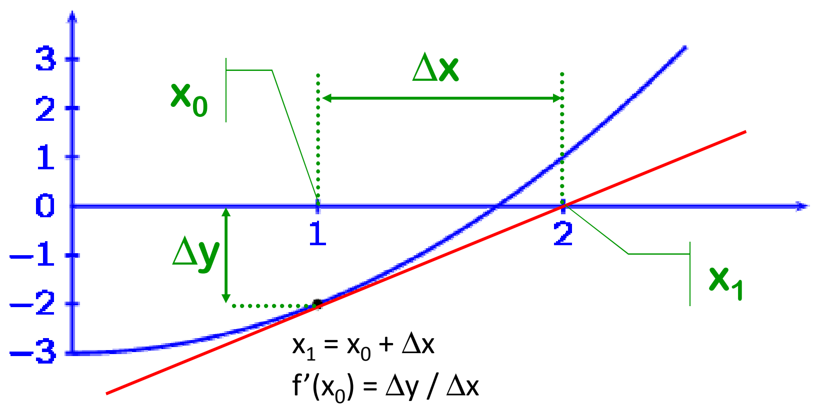
Problème :

- Soit $f:\mathbb{R}\rightarrow\mathbb{R}$ continûment dérivable. Trouver x tel que
$$f(x) = 0.$$
- Exemple :
$$f(x) = x^2 - 3$$
$$x_0 = 1$$

Equation à une inconnue



Equation à une inconnue



$x_1 = x_0 + \Delta x$

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

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

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

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
  – est tel que $M(x_0) = f(x_0)$
  – est tel que $M'(x_0) = f'(x_0)$

Equation à une inconnue

Convergence locale

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

Définition :

Une fonction g est continue au sens de Lipschitz sur X s'il existe γ >0 tel que

$$|g(x)-g(y)| \leq \gamma |x-y| \quad \forall x,y \in X$$

Equation à une inconnue

Théorème :

- Soit un intervalle ouvert D
- Soit $f:D \to IR$
- f' continue au sens de Lipschitz sur D (constante γ)
- Supposons que

$$\exists \rho > 0, |f'(x)| \geq \rho, \forall x \in D$$

Equation à une inconnue

- Si $f(x)=0$ possède une solution $x^* \in D$
- Alors $\exists\eta > 0$ tel que
  - Si $|x_0 - x^*| < \eta$
  - Alors la suite (x_k) générée par

    $$x_{k+1}=x_k - f(x_k)/f'(x_k) \quad k=0,1,2,\ldots$$

    existe et converge vers x^*.
  - De plus, pour tout k

    $$|x_{k+1}-x^*| \leq (\gamma/2\rho)\ |x_k-x^*|^2$$

Equation à une inconnue

Notes :

- Hypothèse principale :
  - $f'(x)$ a une borne inférieure strictement positive.
  - Si $f'(x^*)=0$, la méthode de Newton converge beaucoup plus lentement.

Equation à une inconnue

$x^* = 1$

$f_1'(x^*)=2$

$f_2'(x^*)=0$

$f_1(x)=x^2-1$	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

Equation à une inconnue

Note :

- Si f est linéaire, alors $\gamma = 0$.
- Donc $|x_{k+1}-x^*| \leq (\gamma/2\eta)\ |x_k-x_0|^2$
  devient
  $$|x_{k+1}-x^*| \leq 0$$
  c'est-à-dire $x^* = x_{k+1}$, pour tout k.

Ainsi, $x_1 = x^*$.

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

Variations sur Newton

- Convergence de la méthode de la plus forte pente
- Résolution d'une équation non linéaire à une inconnue
- Résolution d'un système d'équations à plusieurs inconnues
- Convergence globale
- Méthodes sécantes ou quasi-Newton

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

Théoréme :

- Soit une fonction $g:IR^n \rightarrow IR^n$.
- Soit un vecteur x^* tel que $g(x^*)=0$.
- Supposons que g est continûment différentiable dans une sphère S_α, où $S_\alpha = \{x$ t.q. $\|x-x^*\| \leq \alpha\}$, $\alpha > 0$.
- Supposons que $\nabla g(x^*)$ est inversible.

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

2) S'il existe $L > 0$, $M > 0$, $\delta > 0$ tels que

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

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

pour tout x,y dans S_δ, alors

$$\|x_{k+1}-x^*\| \leq ½ LM \|x_k-x^*\|^2, \forall k$$

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 »

Variations sur Newton

- Convergence de la méthode de la plus forte pente
- Résolution d'une équation non linéaire à une inconnue
- Résolution d'un système d'équations à plusieurs inconnues
- Convergence globale
- Méthodes sécantes ou quasi-Newton

Minimisation à une variable

- Appliquer la méthode de Newton à l'équation

$$f'(x) = 0.$$

- Cela donne

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

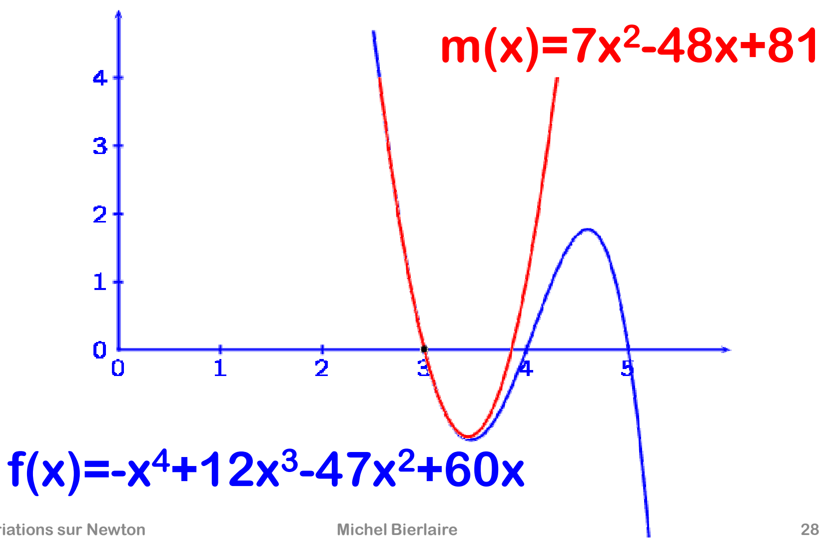
Minimisation à une variable

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

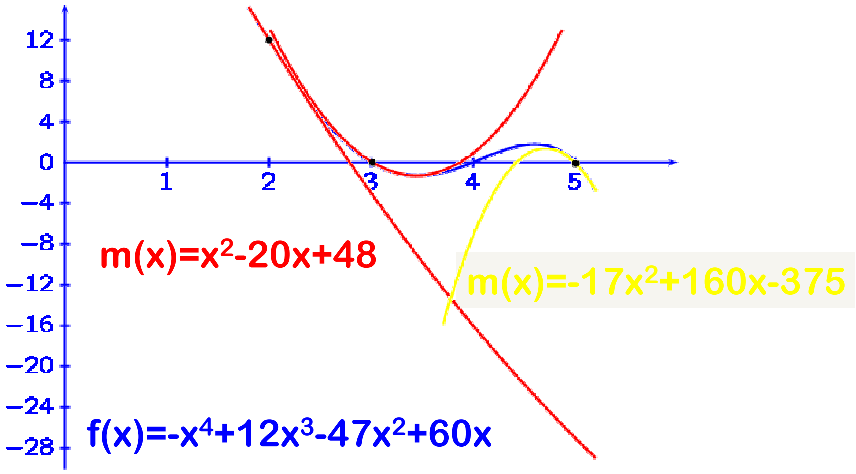
$$m_k(x) = f(x_k) + f'(x_k)(x-x_k)+½ f''(x_k)(x-x_k)^2$$

- Ce modèle
  - est quadratique
  - est tel que $m_k(x_k) = f(x_k)$
  - est tel que $m_k'(x_k) = f'(x_k)$
  - est tel que $m_k''(x_k) = f''(x_k)$

Minimisation à une variable



Minimisation à une variable



Méthode de Newton « pure »

- 1 variable :

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

- n variables :

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

- Rapide
- Mais ne fonctionne pas toujours

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.

Convergence globale

- 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.
- Attention : ne pas confondre convergence globale et minimum global.

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

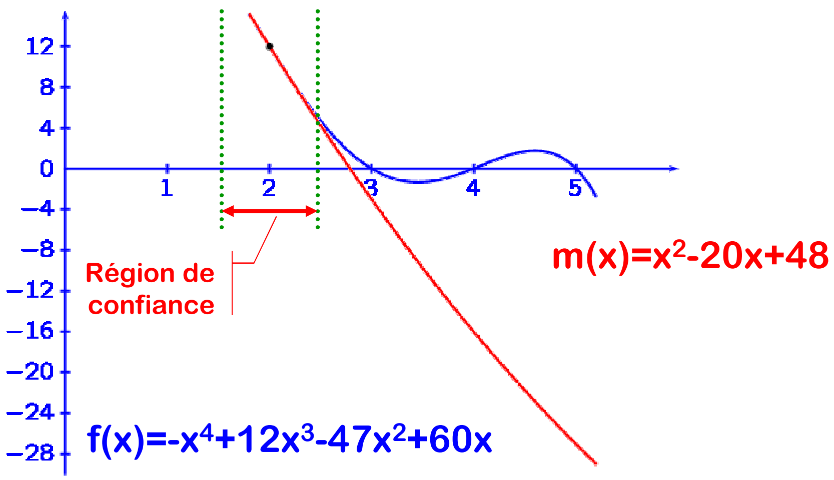
Convergence globale

Région de confiance

Idée :

- Si la direction de Newton n'est pas satisfaisante, c'est probablement parce que le modèle quadratique ne représente pas bien la fonction dans la région contenant le point de Newton.
- Au lieu de garder la même direction et de racourcir le pas, on choisit un pas plus petit et on détermine une nouvelle direction.

Convergence globale



Convergence globale

- Supposons que l'on ait une idée de la « bonne » longueur de pas: δ_k.
- Comment choisir d_k ?
- Problème de région de confiance :

$$\min_s m_k(x_k+s)$$

$$\text{sous contrainte } \|s\|_2 \leq \delta_k$$

- δ_k définit la région dans laquelle on peut faire confiance au modèle m_k

Convergence globale

Notes :

- Ce problème doit être plus simple que le problème initial.
- Fonction objectif : quadratique
- Contrainte simple
- On n'a pas nécessairement besoin de la solution exacte du problème.

Convergence globale

Théorème :

- Soit $s(\mu) = -(\nabla^2 f(x_k)+\mu I)^{-1}\nabla f(x_k)$
- La solution s^* du problème de région de confiance

$$\min_s m_k(x_k+s)$$
$$\text{sous contrainte } \|s\|_2 \leq \delta_k$$

est

- le pas de Newton $s^*=s(0)$ si $\|s(0)\|_2 \leq \delta_k$.
- Sinon, $s^*=s(\mu)$ pour le seul $\mu\geq 0$ tel que $\|s(\mu)\|_2=\delta_k$.

Convergence globale

- Résoudre le problème de région de confiance revient donc à perturber $\nabla^2 f(x_k)$ pour la rendre suffisamment définie positive.

Mise à jour du rayon :

- Soit $\rho_k = \dfrac{f(x_k)-f(x_k+s^*)}{m(x_k)-m(x_k+s^*)}$

Réduction effective

Réduction prédite

Convergence globale

- Soit $0 < \eta_1 \leq \eta_2 < 1$ ($\eta_1 = 0.01$, $\eta_2 = 0.9$)
- Soit $0 < \gamma_1 \leq \gamma_2 < 1$ ($\gamma_1 = \gamma_2 = 0.5$)
- Si $\rho_k \geq \eta_2$, s* est très satisfaisant.
  - $x_{k+1} = x_k + s^*$
  - Le rayon est augmenté
  - $\delta_{k+1} \in [\delta_k, \infty[$
  - Souvent $\delta_{k+1} = 2\ \delta_k$
- Si $\rho_k \in [\eta_1, \eta_2[$, s* est satisfaisant.
  - Le rayon reste (plus ou moins) le même
  - $\delta_{k+1} \in [\gamma_2 \delta_k, \delta_k]$
  - Souvent $\delta_{k+1} = \delta_k$

Convergence globale

- Si $\rho_k < \eta_1$, s* n'est pas satisfaisant.
  - Le rayon est diminué.
  - $x_{k+1} = x_k$
  - $\delta_{k+1} \in [\gamma_1 \delta_k, \gamma_2 \delta_k]$
  - Souvent $\delta_{k+1} = \delta_k / 2$

Variations sur Newton

- Convergence de la méthode de la plus forte pente
- Résolution d'une équation non linéaire à une inconnue
- Résolution d'un système d'équations à plusieurs inconnues
- Convergence globale
- Méthodes sécantes ou quasi-Newton

Méthodes quasi-Newton

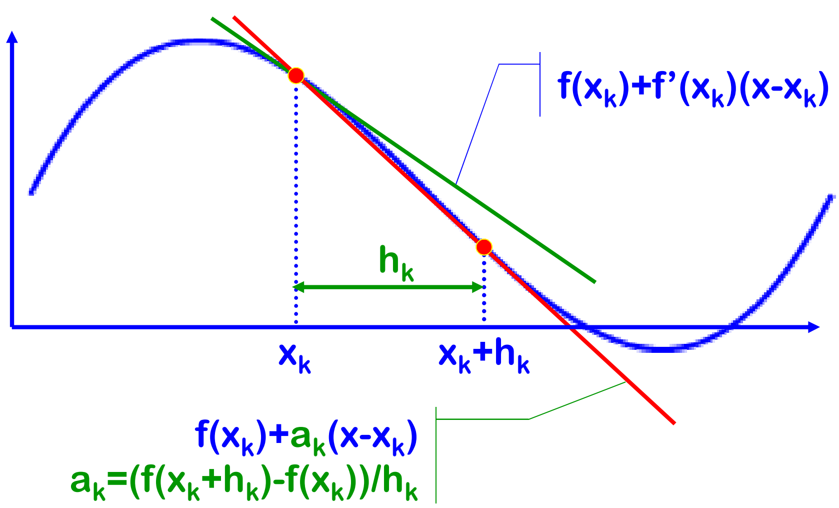
- Résolution d'équation à une inconnue
- Equations à plusieurs variables
- Minimisation à plusieurs variables

Méthodes quasi-Newton

Résolution d'équation à une inconnue

- 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



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 ?

Méthodes quasi-Newton

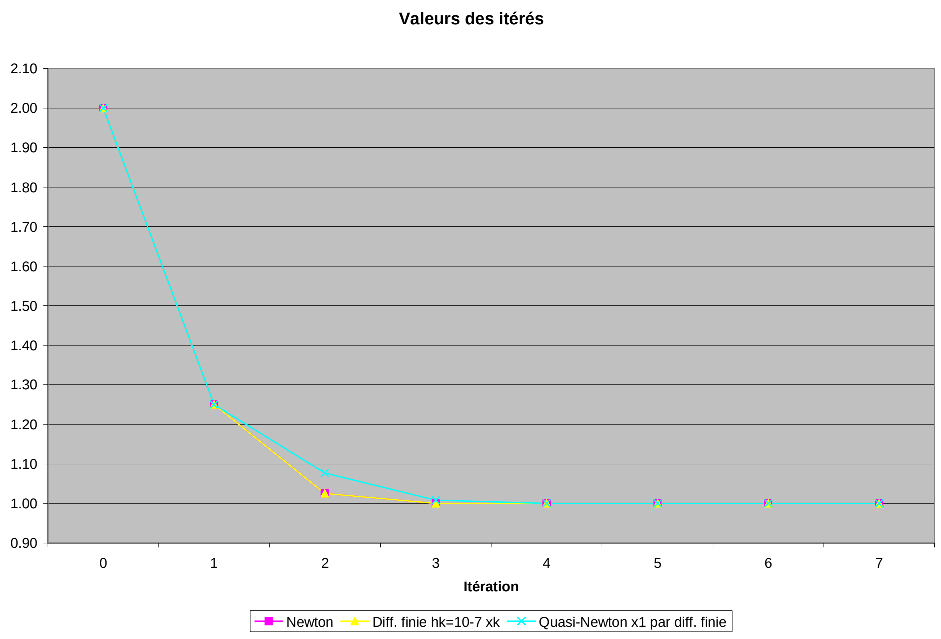
- Si $h_k \to 0$, alors $a_k \to f'(x_k)$
- Si h_k est choisi suffisamment petit, a_k est appelée une approximation de $f'(x_k)$ par différence finie.
- Problème : cela exige une évaluation supplémentaire de la fonction.
- On préfère choisir

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

Méthodes quasi-Newton

$f(x)=x^2-1$

k	Newton	Diff. finie	Quasi-Newton
		$h_k=10^{-7}\,x_k$	x_1 par diff. finie
0	2.0000000000	2.0000000000	2.0000000000
1	1.2500000000	1.2500000379	1.2500000379
2	1.0250000000	1.0250000180	1.0769230877
3	1.0003048780	1.0003048797	1.0082644650
4	1.0000000465	1.0000000465	1.0003048782
5	1.0000000000	1.0000000000	1.0000012545
6	1.0000000000	1.0000000000	1.0000000002
7	1.0000000000	1.0000000000	1.0000000000



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

$$M_k(x_{k-1})=f(x_{k-1}) \text{ et } M_k(x_k)=f(x_k).$$

Méthodes quasi-Newton

- Résolution d'équation à une inconnue
- Equations à plusieurs variables
- Minimisation à plusieurs variables

Méthodes quasi-Newton

Equations à plusieurs variables

- Même idée : utiliser un modèle linéaire

$$M_k(x) = F(x_k) + A_k(x-x_k)$$

- tel que

$$M_k(x_k) = F(x_k) \checkmark$$
$$M_k(x_{k-1})=F(x_{k-1})$$

Méthodes quasi-Newton

- $M_k(x_{k-1}) = F(x_k) + A_k(x_{k-1}-x_k)$

Donc, il faut

- $F(x_k) + A_k(x_{k-1}-x_k) = F(x_{k-1})$

c'est-à-dire

$$A_k(x_{k-1}-x_k) = F(x_{k-1}) - F(x_k)$$

- Cette équation est appelée l'équation sécante.

Méthodes quasi-Newton

Posons

- $d_k = x_k - x_{k-1}$
- $y_k = F(x_k) - F(x_{k-1})$

L'équation sécante s'écrit

$$A_k d_k = y_k.$$

- En n dimensions (n > 1), l'équation sécante ne définit pas A_k de manière unique.
- n équations pour n^2 inconnues

Méthodes quasi-Newton

- Parmi tous les A_k qui vérifient l'équation sécante, on choisira celui qui modifie le moins le modèle linéaire déjà utilisé.
- On veut donc choisir A_k qui minimise

$$M_k(x)-M_{k-1}(x)$$

Méthodes quasi-Newton

$M_k(x) - M_{k-1}(x)$

$= F(x_k) + A_k(x-x_k) - F(x_{k-1}) - A_{k-1}(x-x_{k-1})$

$= F(x_k) + A_k(x-x_k) - F(x_{k-1}) - A_{k-1}(x-x_{k-1}) + A_k x_{k-1} - A_k x_{k-1}$

$= F(x_k) - F(x_{k-1}) - A_k(x_k - x_{k-1}) + (A_k - A_{k-1})(x - x_{k-1})$

$= (A_k - A_{k-1})(x - x_{k-1})$

Equation sécante

Posons

$$x - x_{k-1} = \alpha d_k + t, \text{ avec } t \text{ tel que } t^T d_k = 0$$

$$M_k(x) - M_{k-1}(x)$$
$$=$$
$$\alpha(A_k - A_{k-1})d_k + (A_k - A_{k-1})t$$

Méthodes quasi-Newton

1. $\alpha(A_k - A_{k-1})d_k$

L'équation sécante nous impose

$$A_k d_k = y_k.$$

Donc, le premier terme est

$$\alpha(y_k - A_{k-1}d_k).$$

Il ne dépend pas de A_k. Aucune marge de manœuvre ici.

Méthodes quasi-Newton

2. $(A_k\text{-}A_{k-1})t$, avec $t^T d_k = 0$

On peut annuler ce terme en posant

$$A_k - A_{k-1} = u\, d_k^T,$$

où $u \in IR^n$.

Reprenons l'équation sécante :

$$A_k d_k = y_k$$

$$(A_k - A_{k-1}) d_k = y_k - A_{k-1} d_k$$

$$u d_k^T d_k = y_k - A_{k-1} d_k$$

Méthodes quasi-Newton

Cela donne

$$A_k - A_{k-1} = u\, d_k^T,$$

avec

$$u = \frac{y_k - A_{k-1} d_k}{d_k^T d_k}$$

et donc

$$A_k = A_{k-1} + \frac{(y_k - A_{k-1} d_k)\, d_k^T}{d_k^T d_k}$$

Méthodes quasi-Newton

C'est la mise à jour qui

1. vérifie l'équation sécante
2. en minimisant les changements du modèle linéaire.

Proposée par Broyden en 1965, on l'appelle la mise à jour de Broyden.

$F_1(x,y)=x+y-3$
$F_2(x,y)=x^2+y^2-9$

Note : dès la 1ière itération, $F_1(x,y)=0$

	F_1	F_2	Newton	
			x	y
0	3	17.0000000000	1.0000000000	5.0000000000
1	0	4.5312500000	-0.6250000000	3.6250000000
2	0	0.5683661332	-0.0919117647	3.0919117647
3	0	0.0159341321	-0.0026533419	3.0026533419
4	0	0.0000140556	-0.0000023426	3.0000023426
5	0	0.0000000000	0.0000000000	3.0000000000
6	0	0.0000000000	0.0000000000	3.0000000000
7	0	0.0000000000	0.0000000000	3.0000000000
	F_1	F_2	Broyden	
			x	y
0	3	17.0000000000	1.0000000000	5.0000000000
1	0	4.5312500000	-0.6250000000	3.6250000000
2	0	0.4660238751	-0.0757575758	3.0757575758
3	0	0.0770929956	-0.0127942682	3.0127942682
4	0	0.0018831430	-0.0003138243	3.0003138243
5	0	0.0000079954	-0.0000013326	3.0000013326
6	0	0.0000000008	-0.0000000001	3.0000000001
7	0	0.0000000000	0.0000000000	3.0000000000

Méthodes quasi-Newton

- Résolution d'équation à une inconnue
- Equations à plusieurs variables
- Minimisation à plusieurs variables

Méthodes quasi-Newton

Minimisation à plusieurs variables

- On remplace

$$F(x) \text{ par } \nabla f(x)$$
$$\nabla F(x) \text{ par } \nabla^2 f(x)$$

Attention :

le hessien doit être symétrique.

L'approximation sécante ne peut être utilisée telle quelle.

Méthodes quasi-Newton

1. Mise à jour de Powell

- On veut approximer $\nabla^2 f(x_k)$ par H_k.
- L'équation sécante est

$$H_k\, d_k = y_k$$

avec

– $d_k = x_k - x_{k-1}$

– $y_k = \nabla f(x_k) - \nabla f(x_{k-1})$.

Méthodes quasi-Newton

- Le modèle quadratique

$$m_k(x)=f(x_k)+\nabla f(x_k)^T(x-x_k)+½(x-x_k)^T H_k(x-x_k)$$

vérifie

– $m_k(x_k) = f(x_k)$

– $\nabla m_k(x_k) = \nabla f(x_k)$

– $\nabla m_k(x_{k-1}) = \nabla f(x_{k-1})$

- Utilisons la méthode de Broyden.

Méthodes quasi-Newton

- $H_k = H_{k-1} + \dfrac{(y_k - H_{k-1}d_k)\, d_k^T}{d_k^T d_k}$
- Si H_{k-1} est symétrique, H_k ne le sera que si $y_k - H_{k-1}d_k$ est multiple de d_k.
- On doit trouver une matrice H_k qui
  - vérifie l'équation sécante et
  - soit symétrique.
- Construisons une suite de matrices.

Méthodes quasi-Newton

- $H_1 = H_{k-1} + \dfrac{(y_k - H_{k-1}d_k)\, d_k^T}{d_k^T d_k}$
- H_1 vérifie l'équation sécante.
- H_1 n'est pas symétrique
- $H_2 = ½ (H_1 + H_1^T)$
- H_2 est symétrique.
- H_2 ne vérifie pas l'équation sécante.
- On lui applique donc la mise à jour de Broyden.

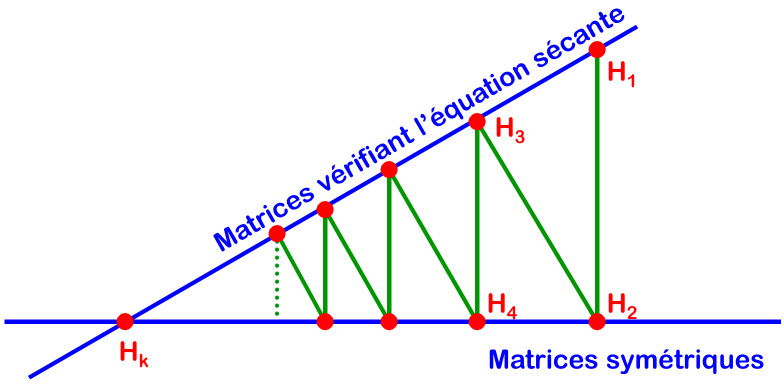
Méthodes quasi-Newton

- $H_3 = H_2 + \frac{(y_k - H_2 d_k)\, d_k^T}{d_k^T d_k}$
- H_3 vérifie l'équation sécante.
- H_3 n'est pas symétrique
- $H_4 = ½ (H_3 + H_3^T)$
- etc…

Méthodes quasi-Newton

- On génère la suite suivante :
- $H_{2i+1} = H_{2i} + \frac{(y_k - H_{2i} d_k)\, d_k^T}{d_k^T d_k}$
- $H_{2i+2} = ½ (H_{2i+1} + H_{2i+1}^T)$
- Bonne nouvelle : elle converge !

Méthodes quasi-Newton



Méthodes quasi-Newton

- Mise à jour sécante symétrique de Powell :

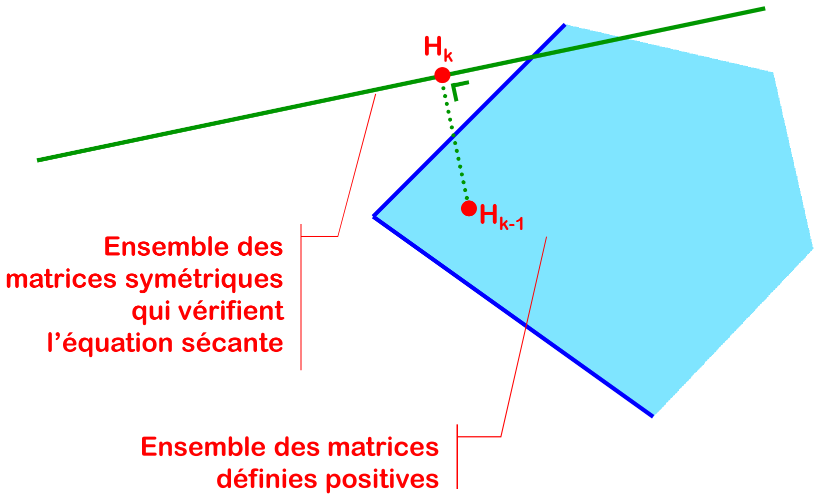
$$H_k = H_{k-1} + \frac{(y_k - H_{k-1}d_k)d_k^T + d_k(y_k - H_{k-1}d_k)^T}{d_k^T d_k} - \frac{d_k^T(y_k - H_{k-1}d_k)\, d_k d_k^T}{(d_k^T d_k)^2}$$

Méthodes quasi-Newton

Malheureusement….

- La matrice H_k obtenue par la méthode de Powell peut ne pas être définie positive.
- Et cela même si H_{k-1} est définie positive, et qu'il existe des matrices
  - définies positives
  - symétriques
  - qui vérifient l'équation sécante.

Méthodes quasi-Newton



Méthodes quasi-Newton

Truc :

- Si on veut imposer $x > 0$, on pose

$$x = y^2,\ y \in IR,\ y \neq 0$$

- Ici, on pose

$$H_k = J_k J_k^T$$

  - $J_k \in IR^{n\times n}$
  - J_k non singulier
- Ainsi, H_k sera définie positive.

Méthodes quasi-Newton

L'équation sécante devient

$$J_k J_k^T d_k = y_k$$

avec

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

Est-ce que cette équation possède une solution ?

Méthodes quasi-Newton

Lemme

- Soient d_k, $y_k \in$ IR, $d_k \neq 0$.
- Il existe une matrice non singulière $J_k \in IR^{nxn}$ telle que

$$J_k J_k^T d_k = y_k$$

- si et seulement si

$$d_k^T y_k > 0.$$

Méthodes quasi-Newton

Preuve :

Condition nécessaire

- Supposons qu'il existe une matrice non singulière $J_k \in IR^{nxn}$ telle que

$$J_k J_k^T d_k = y_k.$$

- Posons $v_k = J_k^T d_k$
- $v_k^T v_k = d_k^T J_k J_k^T d^k = d_k^T y_k > 0.$

CQFD

Méthodes quasi-Newton

Condition suffisante

- Supposons $d_k^T y_k > 0$.
- On peut montrer qu'une solution est
- $J_k = L_{k-1} + \dfrac{(y_k - L_{k-1} v_k) v_k^T}{v_k^T v_k}$
- où
  - $H_{k-1} = L_{k-1} L_{k-1}^T$ (facteurs de Cholesky)
  - $v_k = \sqrt{\dfrac{y_k^T d_k}{d_k^T H_{k-1} d^k}}\, L_{k-1}^T d_k$

Méthodes quasi-Newton

- On obtient la mise à jour suivante

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

Broyden
Fletcher
Goldfarb
Shanno

Méthodes quasi-Newton

Résumé :

- $x_{k+1} = x_k - \alpha_k D_k \nabla f(x_k)$
- Plus forte pente : $D_k = I$
  - Règle d'approximation : Armijo-Goldstein

Méthodes quasi-Newton

Résumé :

- $x_{k+1} = x_k - \alpha_k D_k \nabla f(x_k)$
- Newton : $D_k = [\nabla^2 f(x_k) + \mu I]^{-1}$
  - Newton pure :
    - $\alpha_k = 1$, $\mu=0$ -> pas globalement convergent
  - Règle d'approximation : Armijo-Goldstein

Méthodes quasi-Newton

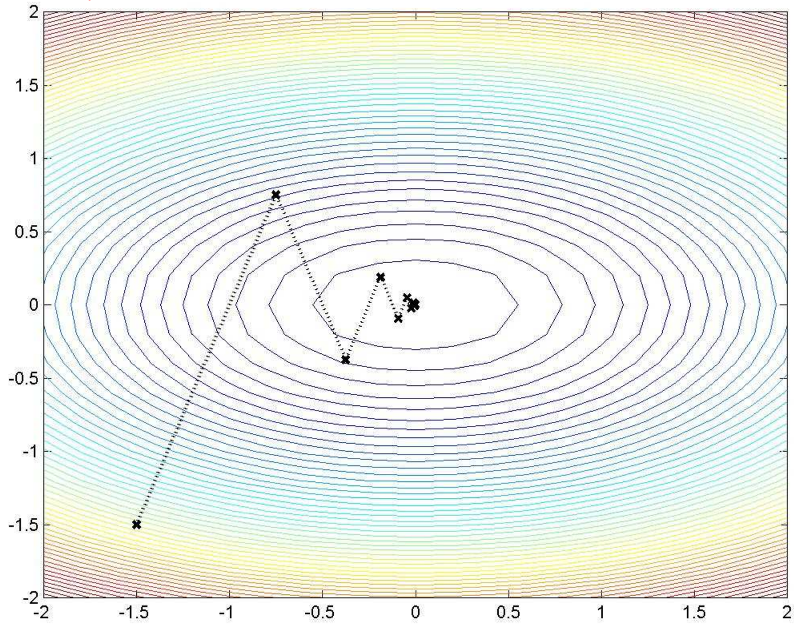
Résumé :

- $x_{k+1} = x_k - \alpha_k D_k \nabla f(x_k)$
- quasi-Newton : $D_k = H_k^{-1}$
  - 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

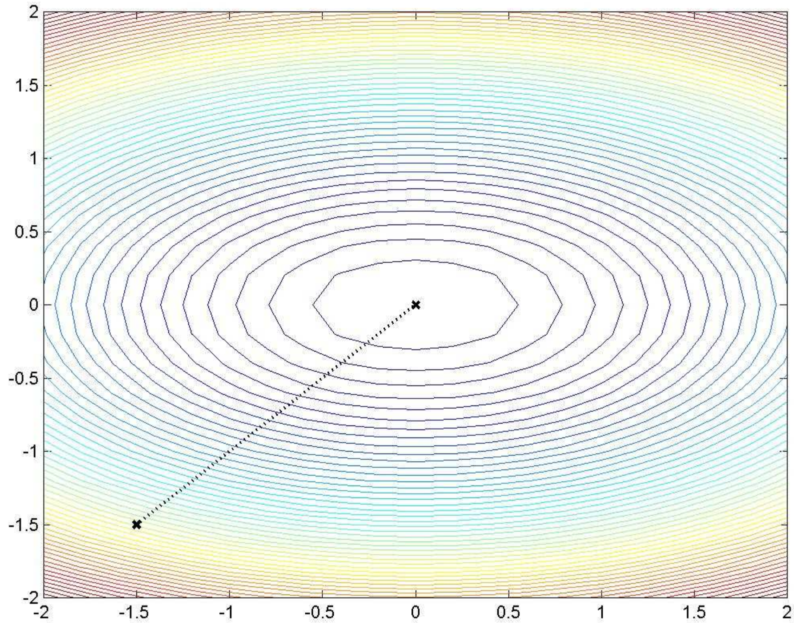
Exemples : quadratique

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

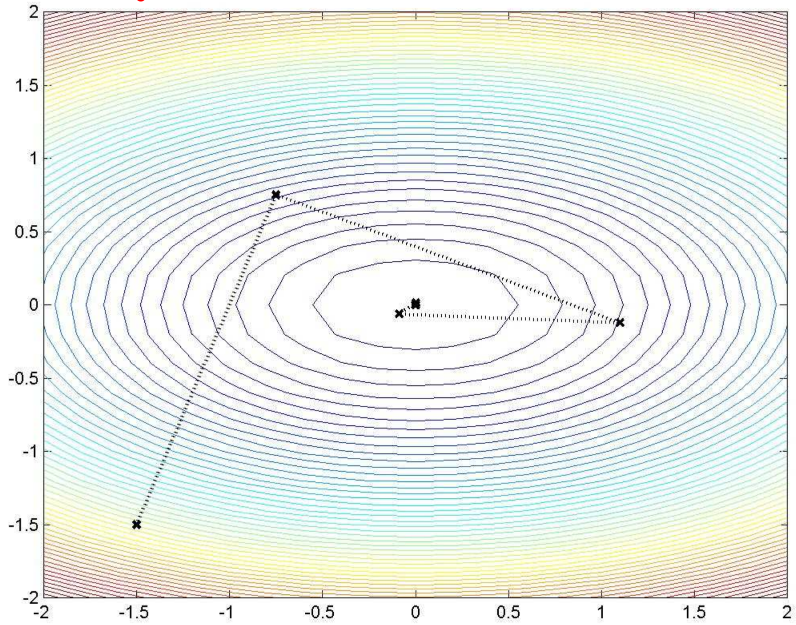
Plus forte pente



Newton

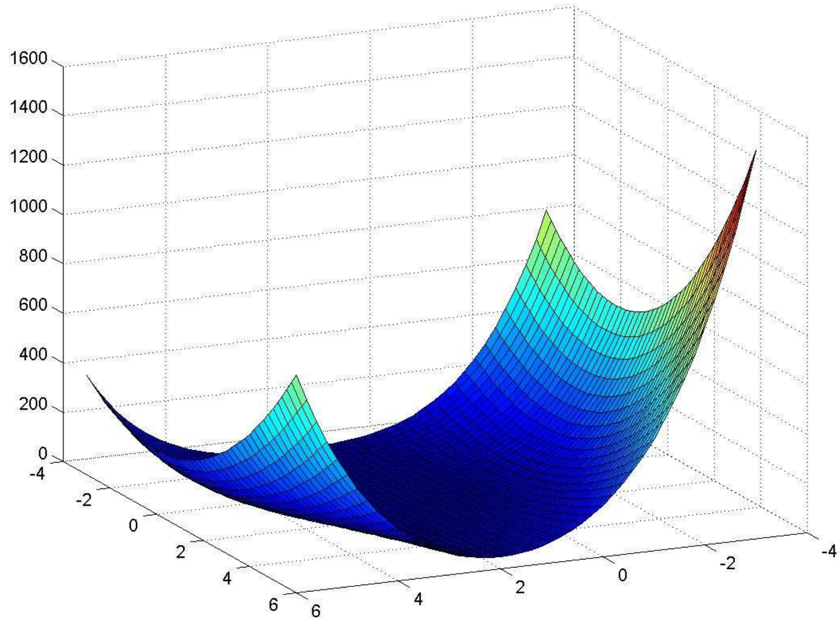


BFGS avec H_0=I

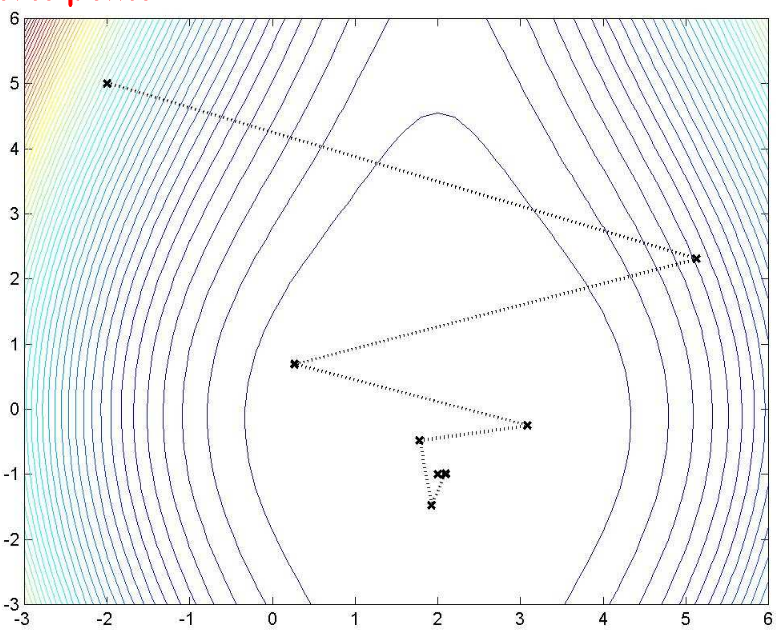


Exemples : « gentille » fonction

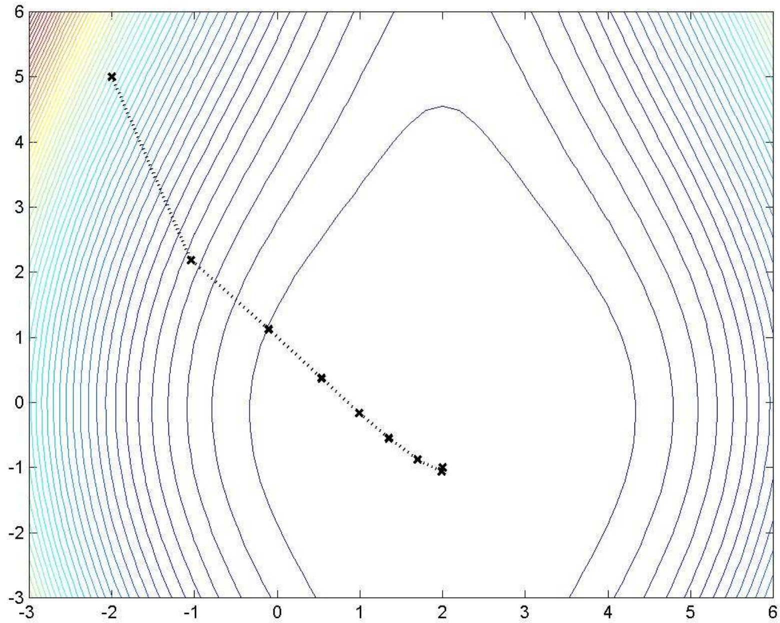
$$f(x) = (x_1 - 2)^4 + (x_1 - 2)^2 x_2^2 + (x_2 + 1)^2$$



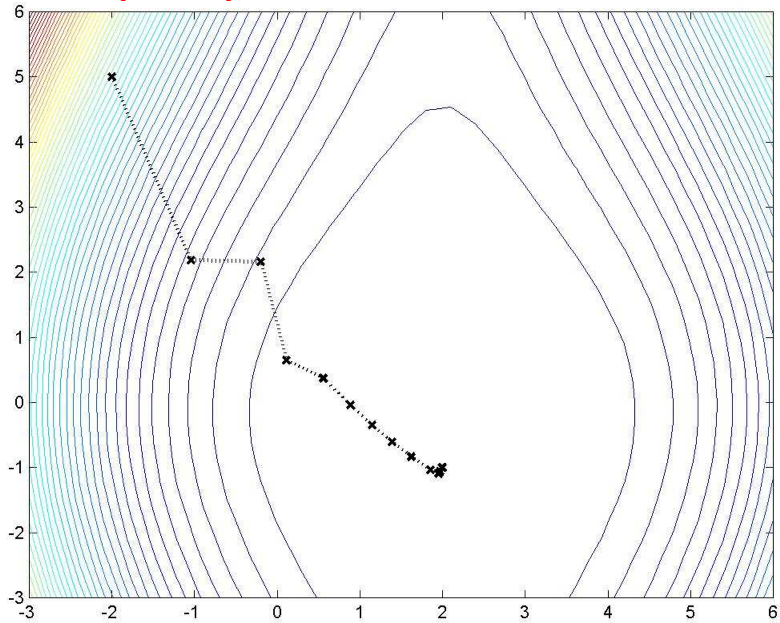
Plus forte pente



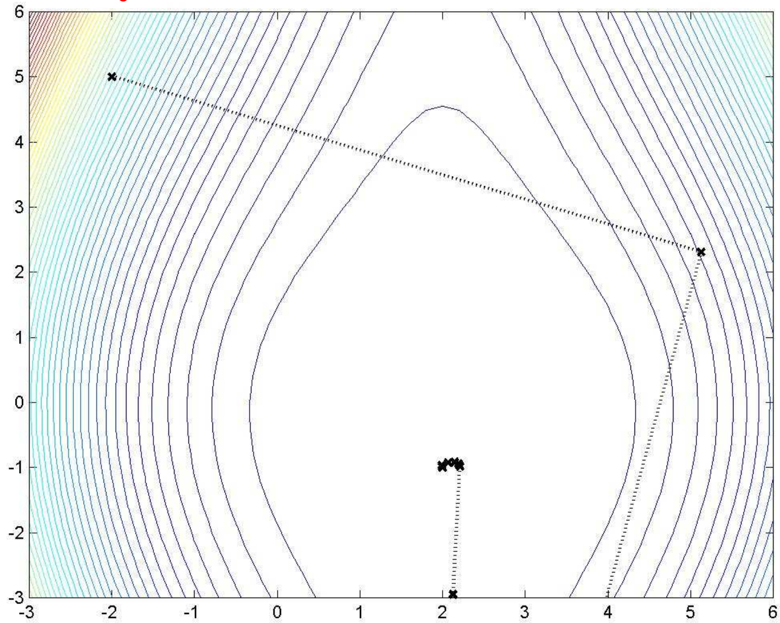
Newton



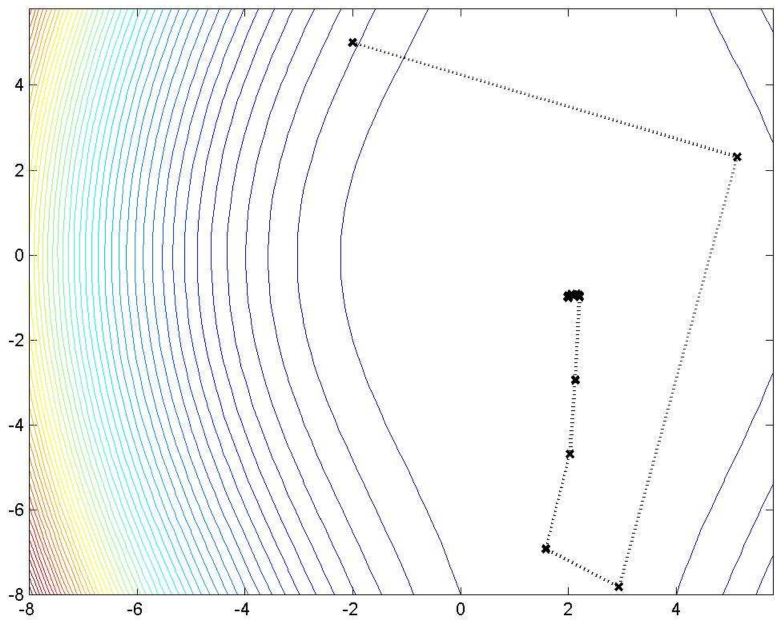
BFGS avec $H_0=\nabla^2 f(x_0)$



BFGS avec H_0=I

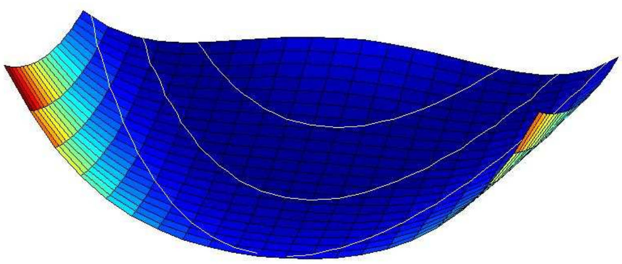


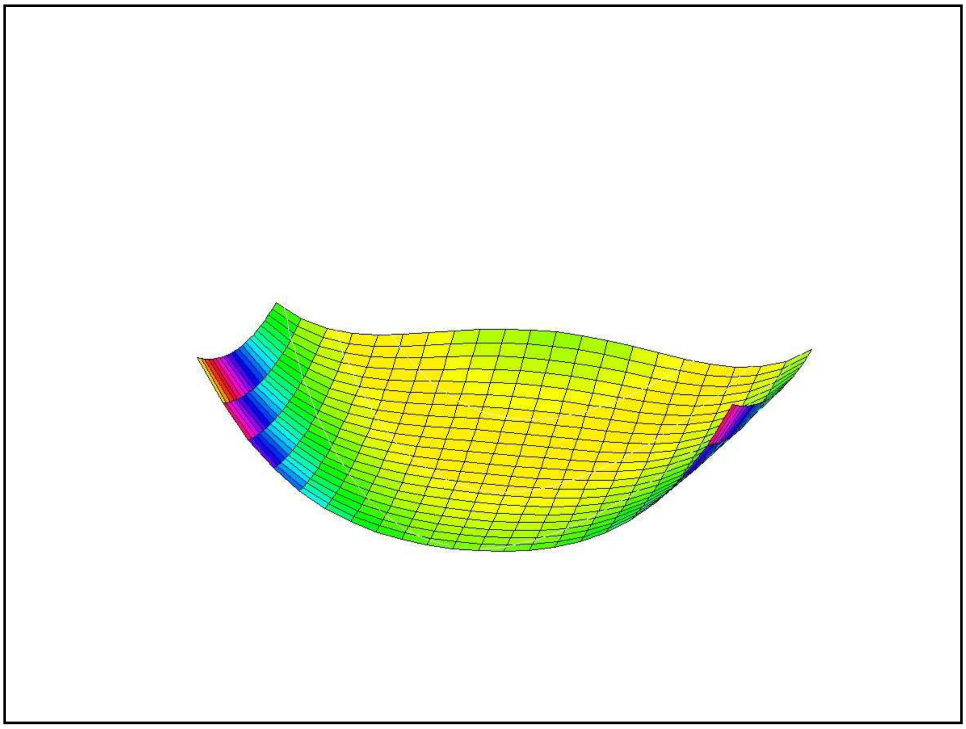
BFGS avec H_0=I (plus large)

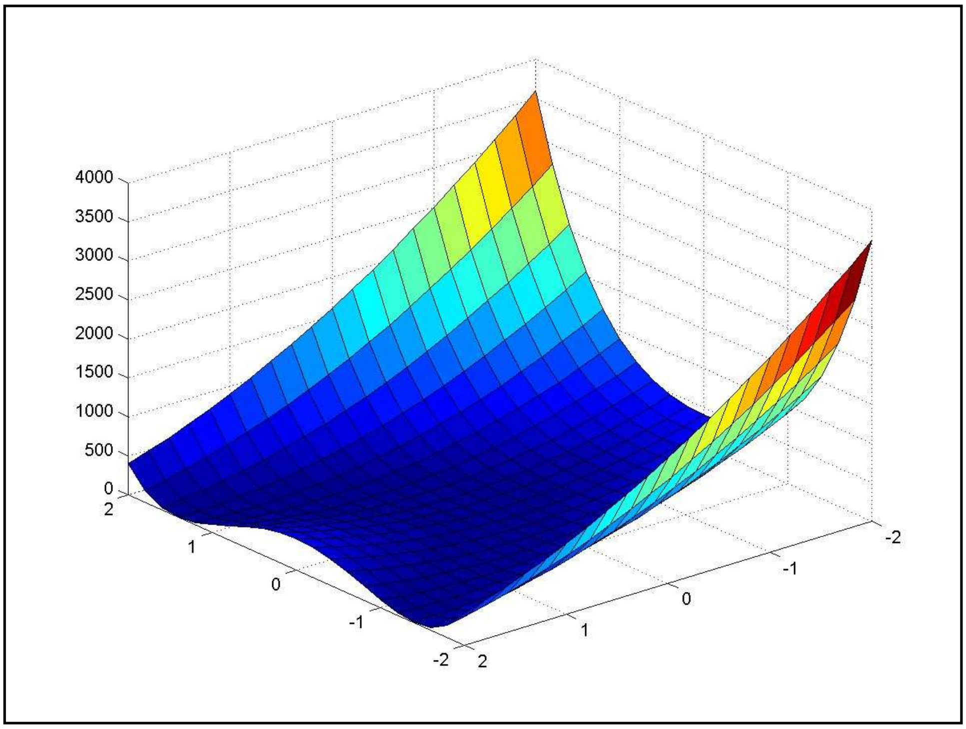


Exemples : Rosenbrock ou « fonction banane »

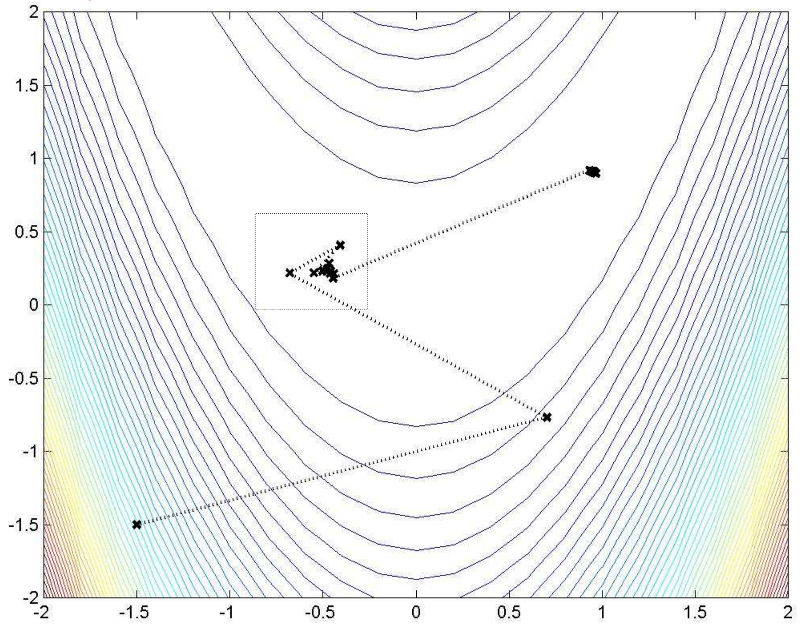
$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$



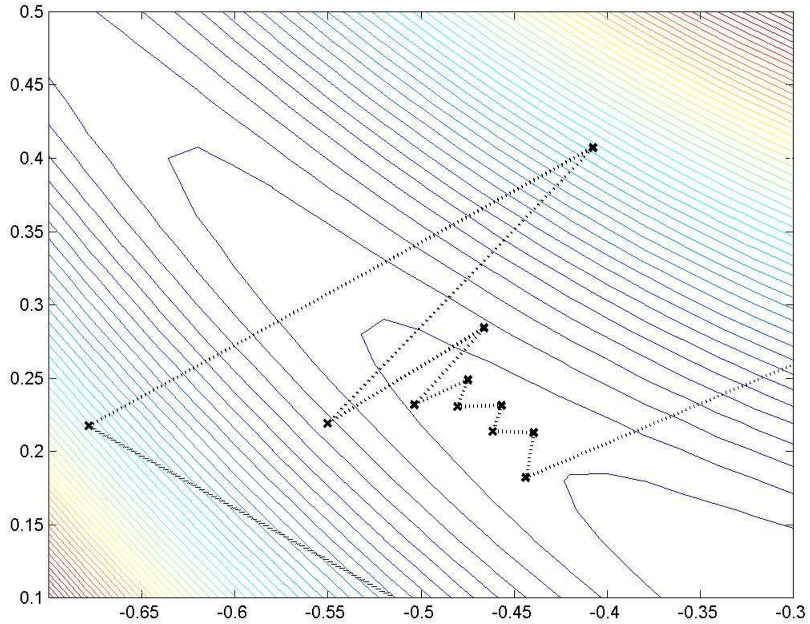



4000
3500
3000
2500
2000
1500
1000
500
0
2
1
0
-1
-2
2
1
0
-1
-2

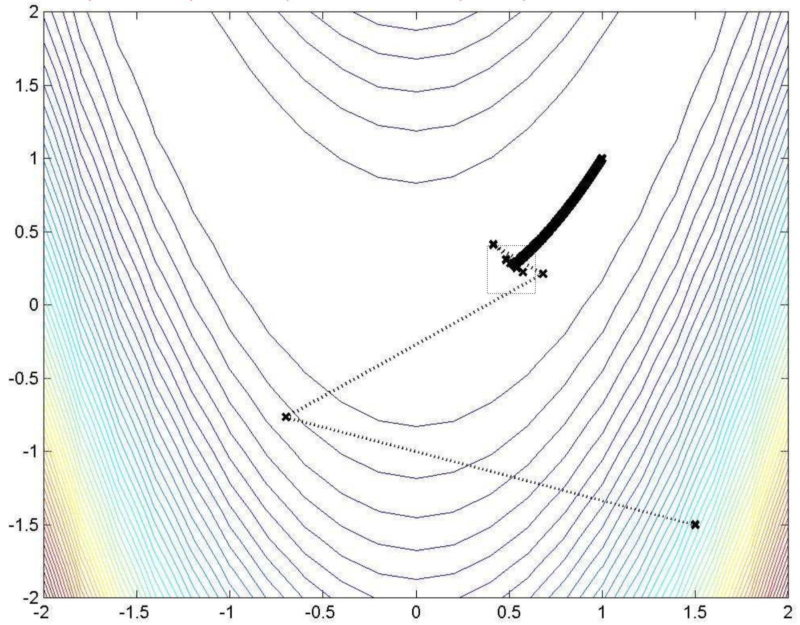
Plus forte pente



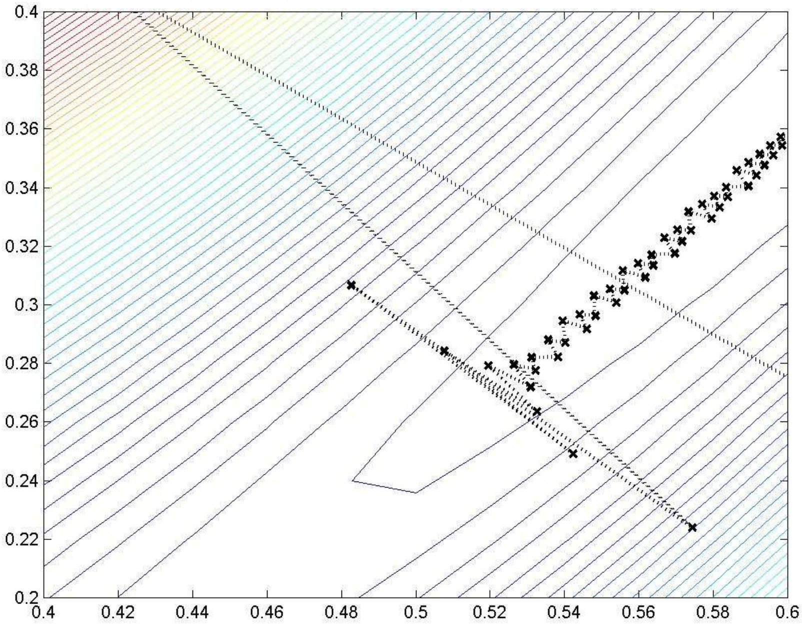
Plus forte pente (zoom)



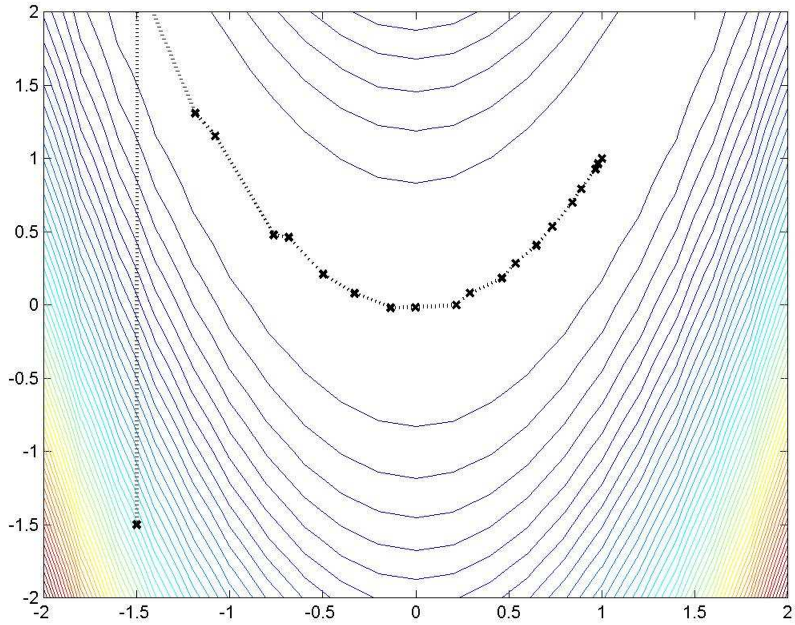
Plus forte pente (autre point de départ)



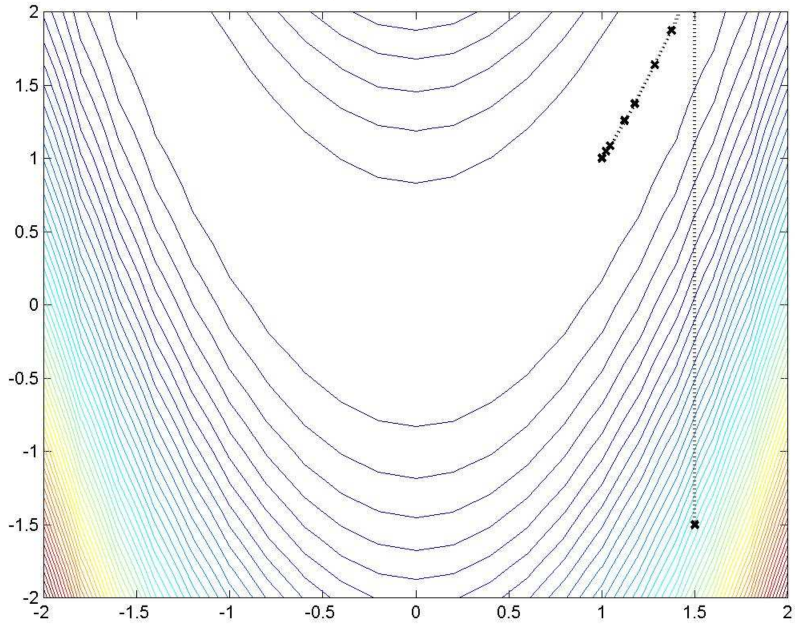
Plus forte pente (autre point de départ – zoom)



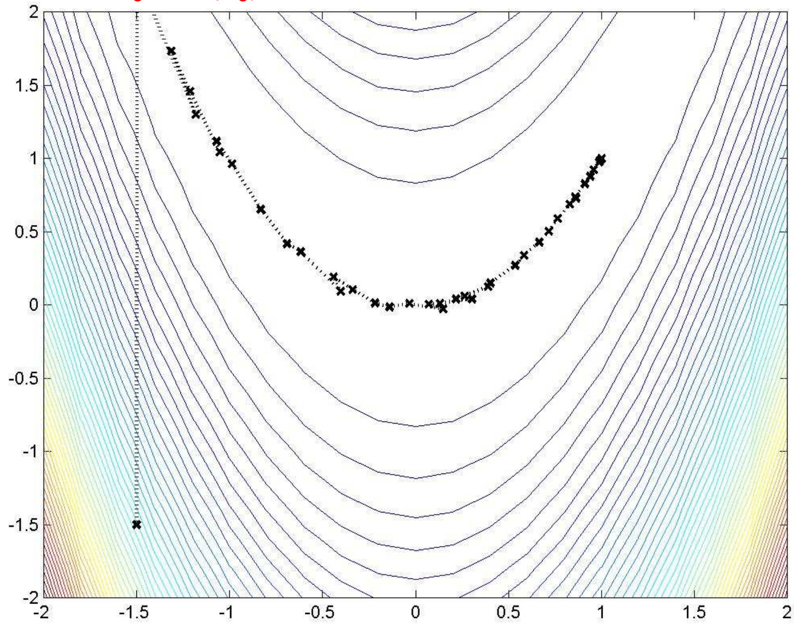
Newton



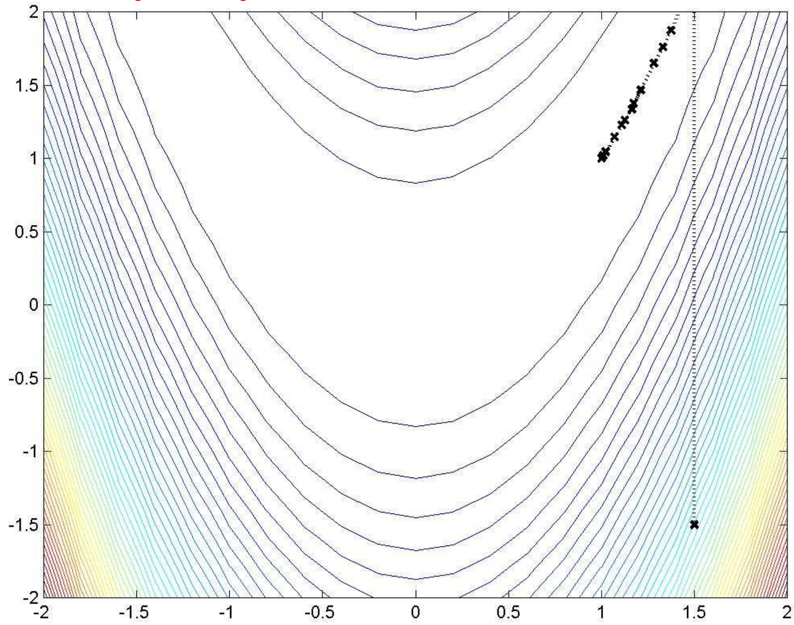
Newton (autre point de départ)



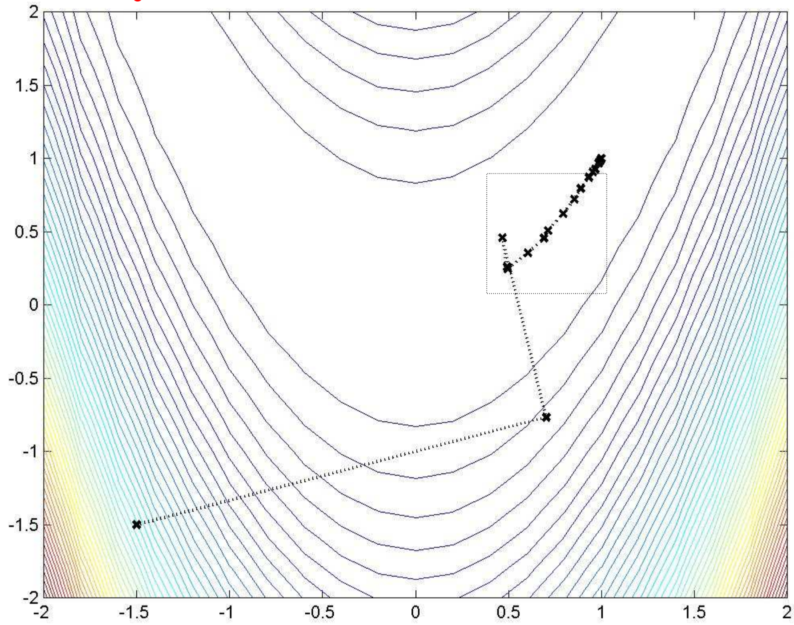
BFGS avec $H_0=\nabla^2 f(x_0)$



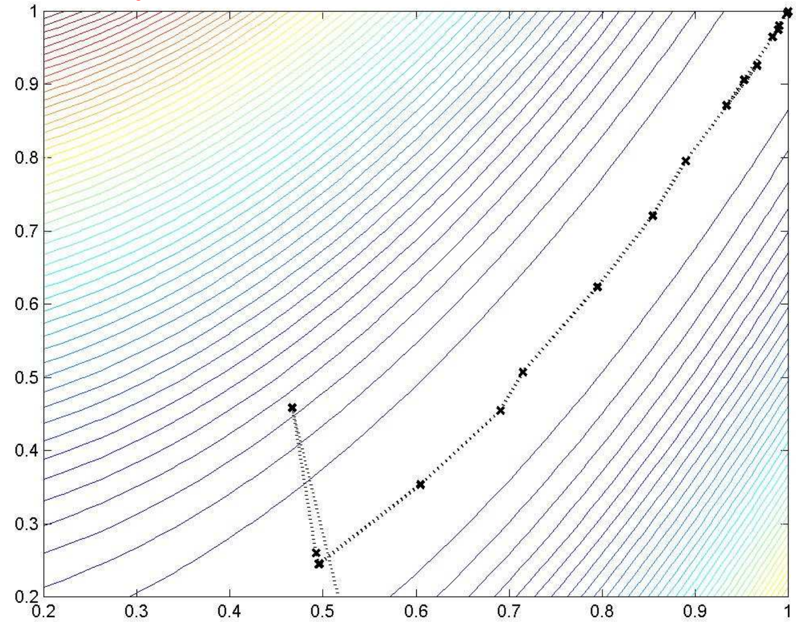
BFGS avec $H_0=\nabla^2 f(x_0)$ (autre point de départ)



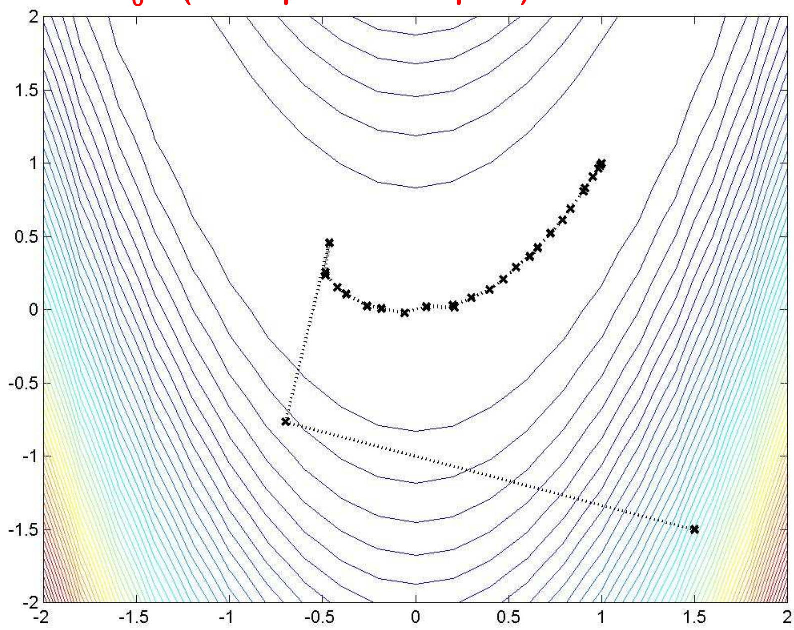
BFGS avec H_0=I



BFGS avec H_0=I (zoom)



BFGS avec H_0=I (autre point de départ)



Newton pure

- Importance du point de départ
- Soit

$$f(x,y)=\frac{x^4}{4}+\frac{y^4}{4}-\frac{3}{2}x^2y^2-x$$

- Trois points critiques : c_1,c_2,c_3
- On applique Newton pure à partir de x_0,y_0

Newton pure

- Si Newton converge vers c_1, (x0,y0) est rouge
- Si Newton converge vers c_2, (x0,y0) est bleu
- Si Newton converge vers c_3, (x0,y0) est vert