II. OPTIMISATION SANS CONTRAINTE

$$\min f(x) \text{ avec } x \in \mathbb{R}^n$$

–Méthodes de recherche unidimensionnelle

–Méthodes du gradient

–Méthodes des directions conjuguées

–Méthode de Newton et méthode de Levenberg-Marquardt

–Méthodes quasi-Newton

–Méthodes sans calcul du gradient

–Résolution d'équations non linéaires

II. OPTIMISATION SANS CONTRAINTE

II.1 Méthodes de recherches unidimensionnelles (1)

II.1.1 Méthode du nombre d'or

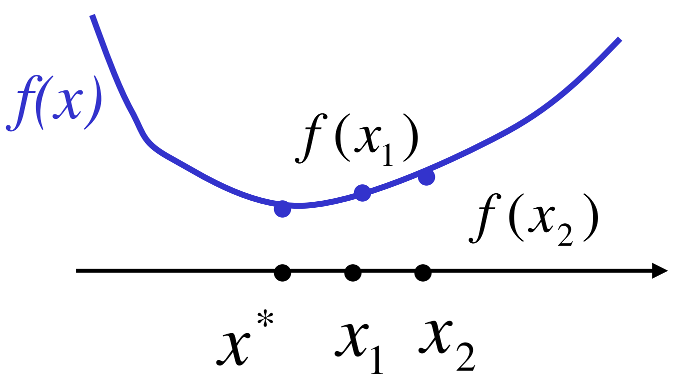
$$\Omega = [a_0, b_0]$$

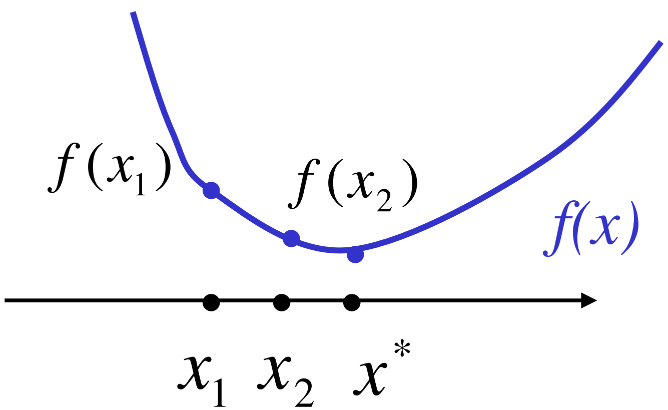
Hypothèse: f unimodulable sur $[a_0, b_0]$

<=> $\exists$ un seul $x^* \in [a_0, b_0]$ minimisant f

$\Rightarrow \forall\ x_1, x_2 \in [a_0, b_0] \qquad x_1 < x_2$

si $x_1 > x^*$ alors $f(x_1) < f(x_2)$ si $x_2 < x^*$ alors $f(x_1) > f(x_2)$

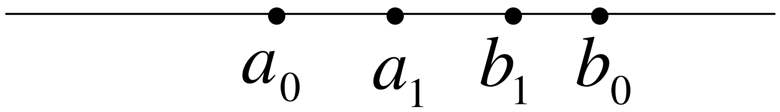




II. OPTIMISATION SANS CONTRAINTE

II.1 Méthodes de recherches unidimensionnelles (2)

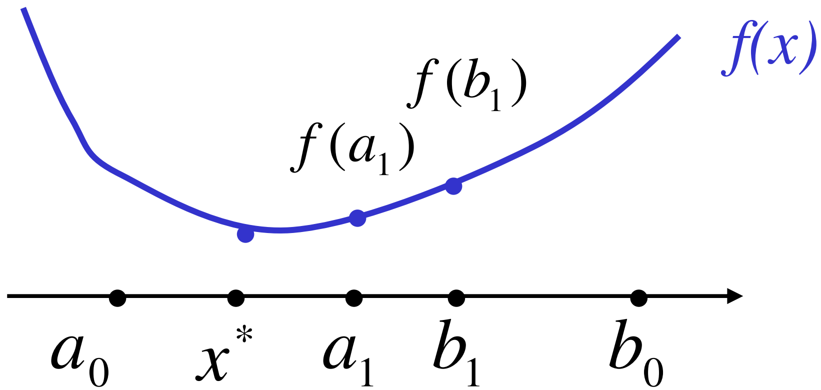
Principes de la méthode du nombre d'or



$$a_1 - a_0 = b_1 - b_0 = \rho(b_0 - a_0) \quad \rho < 1/2$$

si $f(a_1) < f(b_1)$ alors $x^* \in [a_0, b_1]$

si $f(a_1) \geq f(b_1)$ alors $x^* \in [a_1, b_0]$



II. OPTIMISATION SANS CONTRAINTE

II.1 Méthodes de recherches unidimensionnelles (3)

Principes de la méthode du nombre d'or (suite)

- Répéter $\Rightarrow a_2$ et b_2 , a_3 et b_3 , etc...
- Choisir un des points de l'étape précédente :

$$\text{si } f(a_1) < f(b_1) \quad \text{alors } x^* \in [a_0, b_1] \text{ et } b_2 = a_1$$
$$\text{et } b_1 - b_2 = a_2 - a_0 = \rho(b_1 - a_0)$$

$$\text{si } f(a_1) \geq f(b_1) \quad \text{alors } x^* \in [a_1, b_0] \text{ et } a_2 = b_1$$
$$\text{et } b_0 - b_2 = a_2 - a_1 = \rho(b_0 - a_1)$$

II. OPTIMISATION SANS CONTRAINTE

II.1 Méthodes de recherches unidimensionnelles (4)

Calcul de la valeur de ρ dans la méthode du nombre d'or:

$$b_1 - b_2 = \rho \; (b_1 - a_0)$$

$$\Leftrightarrow b_1 - a_1 = \rho \; (b_0 - a_0 + b_1 - b_0)$$

$$\Leftrightarrow (1 - 2\rho)(b_0 - a_0) = \rho(b_0 - a_0) - \rho^2 (b_0 - a_0)$$

$$\Leftrightarrow \rho^2 - 3\rho + 1 = 0 \Leftrightarrow \rho = \frac{3 - \sqrt{5}}{2} \approx 0{,}382$$

=> Facteur de réduction à chaque itération:

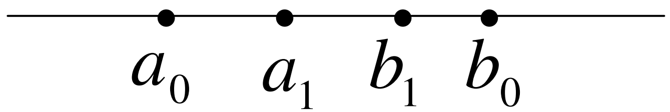
$1 - \rho \approx 0{,}61803$ = NOMBRE D'OR

II. OPTIMISATION SANS CONTRAINTE

II.1 Méthodes de recherches unidimensionnelles (5)

II.1.2 Méthode des suites de Fibonacci

ρ est variable pour optimiser la convergence



$$a_1 - a_0 = b_0 - b_1 = \rho_1 (b_0 - a_0)$$

si $f(a_1) < f(b_1)$ alors $x^* \in [a_0, b_1]$ et $b_2 = a_1$

et $b_1 - b_2 = a_2 - a_0 = \rho_2 (b_1 - a_0)$

si $f(a_1) \geq f(b_1)$ alors $x^* \in [a_1, b_0]$ et $a_2 = b_1$

et $b_0 - b_2 = a_2 - a_1 = \rho_2 (b_0 - a_1)$

$\Rightarrow b_1 - a_1 = \rho_2 (b_1 - b_0 + b_0 - a_0)$ ou $\rho_2 (b_1 - a_0 + a_0 - a_1)$

$$\Leftrightarrow (1 - 2\rho_1) = \rho_2 (1 - \rho_1) \Leftrightarrow \rho_2 = 1 - \frac{\rho_1}{1 - \rho_1} \Leftrightarrow \rho_{k+1} = 1 - \frac{\rho_k}{1 - \rho_k}$$

II. OPTIMISATION SANS CONTRAINTE

II.1 Méthodes de recherches unidimensionnelles (6)

Sélection optimale des facteurs de réduction:

Si N itérations sont permises :

$$\rho_1 = 1 - \frac{R_N}{R_{N+1}} \quad , \quad \rho_2 = 1 - \frac{R_{N-1}}{R_N} \quad , \quad \dots$$

$$\dots \quad \rho_k = 1 - \frac{R_{N-k+1}}{R_{N-k+2}}$$

$$\dots \quad \rho_N = 1 - \frac{R_1}{R_2}$$

où R_k est la suite de Fibonacci :

$$R_{k+1} = R_k + R_{k-1}$$

$$R_{-1} = 0$$

$$R_0 = 1$$

II. OPTIMISATION SANS CONTRAINTE

II.1 Méthodes de recherches unidimensionnelles (7)

II.1.3 Méthode de Newton

$$f : \mathbf{R}^{\mathrm{n}} \rightarrow \mathbf{R} \qquad f \in C^2$$

Approximation de *f(x)* par une forme quadratique autour d'un point x_k :

$$q(x) = f(x_k) + f'(x_k)(x - x_k) + f''(x_k)\frac{(x - x_k)^2}{2}$$

Minimiser *q(x)*

$$\Rightarrow q'(x) = 0 = f'(x_k) + f''(x_k)(x - x_k)$$

$$q''(x) = f''(x_k)$$

$$\Rightarrow x = \boxed{x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}}$$

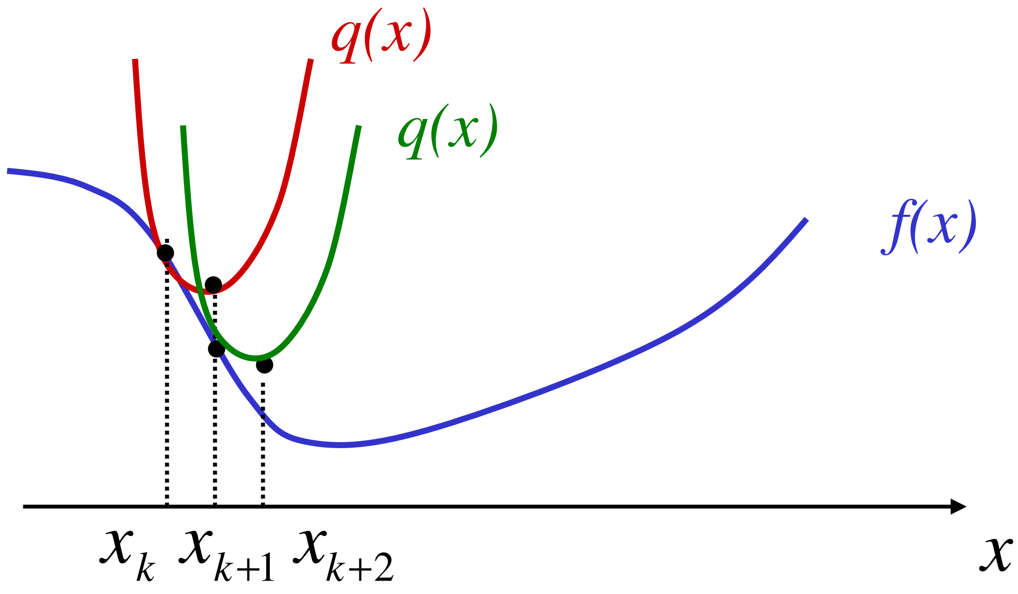
II. OPTIMISATION SANS CONTRAINTE

II.1 Méthodes de recherches unidimensionnelles (8)

II.1.3 Méthode de Newton (suite)

Méthode efficace si $f''(x) > 0 \ \forall x$

$\Rightarrow$ convergence quadratique

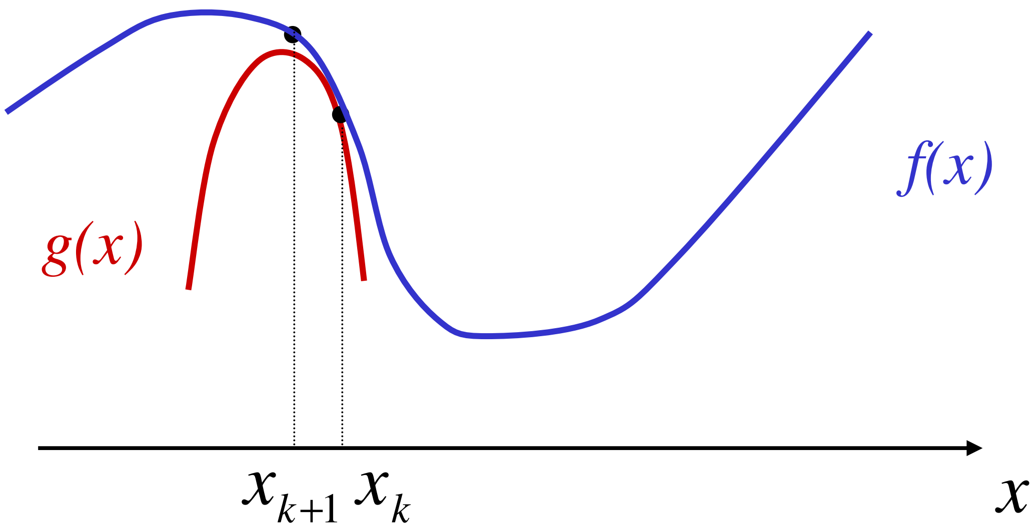


II. OPTIMISATION SANS CONTRAINTE

II.1 Méthodes de recherches unidimensionnelles (9)

II.1.3 Méthode de Newton (suite)

Par contre la méthode ne converge pas nécessairement si $f''(x) < 0$ pour certaines valeurs de x



II. OPTIMISATION SANS CONTRAINTE

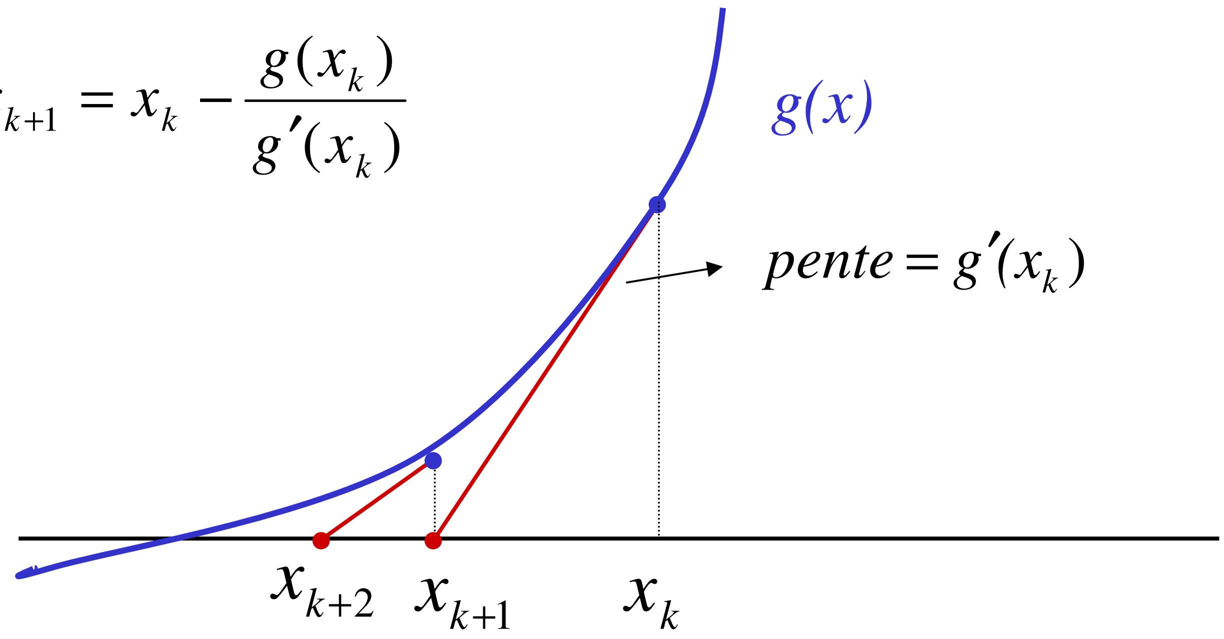
II.1 Méthodes de recherches unidimensionnelles (10)

II.1.3 Méthode de Newton (suite)

La méthode de Newton peut être vue comme une méthode de recherche de zéro d'une fonction *g(x)* si on pose

$$f'(x) = g(x) \quad f''(x) = g'(x)$$

$$\Rightarrow x_{k+1} = x_k - \frac{g(x_k)}{g'(x_k)}$$



II. OPTIMISATION SANS CONTRAINTE

II.1 Méthodes de recherches unidimensionnelles (11)

II.1.4 Méthode de la sécante

La dérivée n'est pas toujours disponible
=> approximation

$$f''(x_k) \cong \frac{f'(x_k) - f'(x_{k-1})}{x_k - x_{k-1}}$$

$$\Rightarrow x_{k+1} = x_k - \frac{x_k - x_{k-1}}{f'(x_k) - f'(x_{k-1})} f'(x_k)$$

Cette méthode amène la dérivée de *f(x)* à zéro

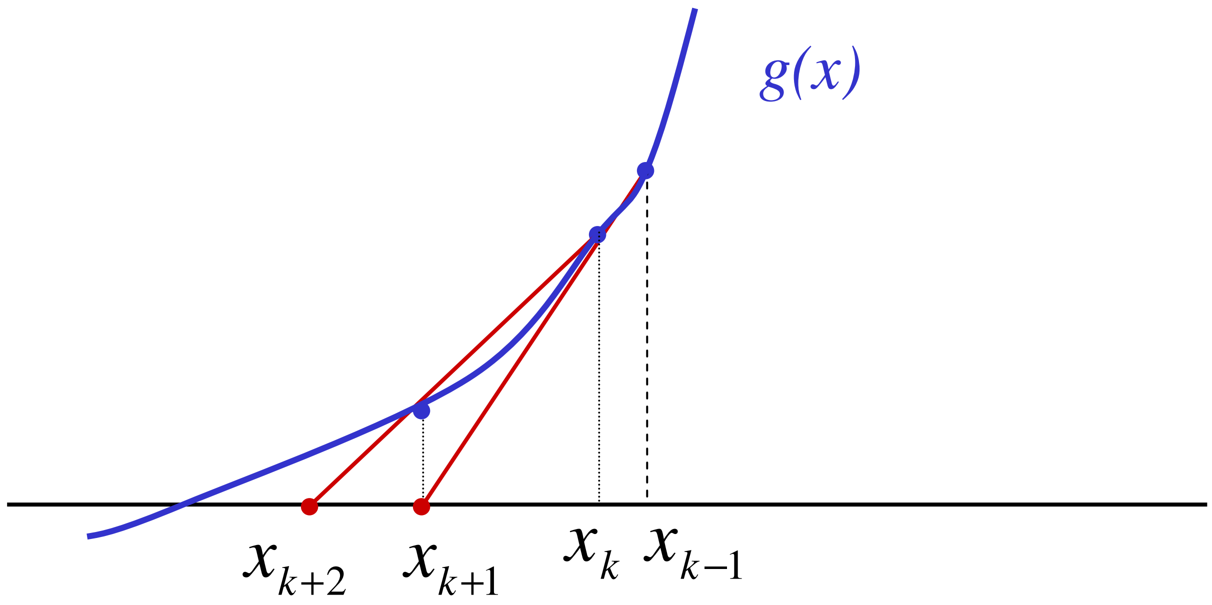
II. OPTIMISATION SANS CONTRAINTE

II.1 Méthodes de recherches unidimensionnelles (12)

II.1.4 Méthode de la sécante (suite)

Si $f'(x) = g(x)$ cette méthode trouve la racine de $g(x)$

$$x_{k+1} = x_k - \frac{x_k - x_{k-1}}{g(x_k) - g(x_{k-1})} g(x_k)$$



II. OPTIMISATION SANS CONTRAINTE

II.1 Méthodes de recherches unidimensionnelles (13)

II.1.5 Méthodes unidirectionnelles « Line search »

$$\min_{x=x_k+\alpha d_k} f(x) \quad \text{avec } \alpha \geq 0 \qquad d_k \in \mathbf{R}^{\mathbf{n}} = \text{direction de recherche}$$

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

$$\Rightarrow x_{k+1} = x_k + \alpha_k d_k \quad \text{avec} \quad \alpha_k = \alpha^* = \arg\min_{\alpha\geq 0} f(x_k+\alpha d_k)$$

$\Rightarrow \alpha_k$ = solution d'un problème unidimensionnel

Exemple: méthode de la sécante

$$\frac{d}{d\alpha} f(x_k+\alpha d_k) = d_k^T \nabla f(x_k+\alpha d_k)$$

$$\Rightarrow \alpha_{i+1} = \alpha_i - \frac{(\alpha_i-\alpha_{i-1}) d_k^T \nabla f(x_k+\alpha_i d_k)}{d_k^T(\nabla f(x_k+\alpha_i d_k) - \nabla f(x_k+\alpha_{i-1} d_k))}$$

II. OPTIMISATION SANS CONTRAINTE

II.1 Méthodes de recherches unidimensionnelles (14)

II.1.5 Méthodes unidirectionnelles (suite)

! Il n'est pas nécessaire de calculer précisément α^*
si $f(x_{k+1}) < f(x_k)$ de manière significative

- If faut distinguer 2 cas : f dérivable et f non dérivable

- La recherche de α^* se fait souvent en deux étapes:
    1. Déterminer un intervalle contenant α^*
    2. Réduire cet intervalle

II. OPTIMISATION SANS CONTRAINTE

II.1 Méthodes de recherches unidimensionnelles (15)

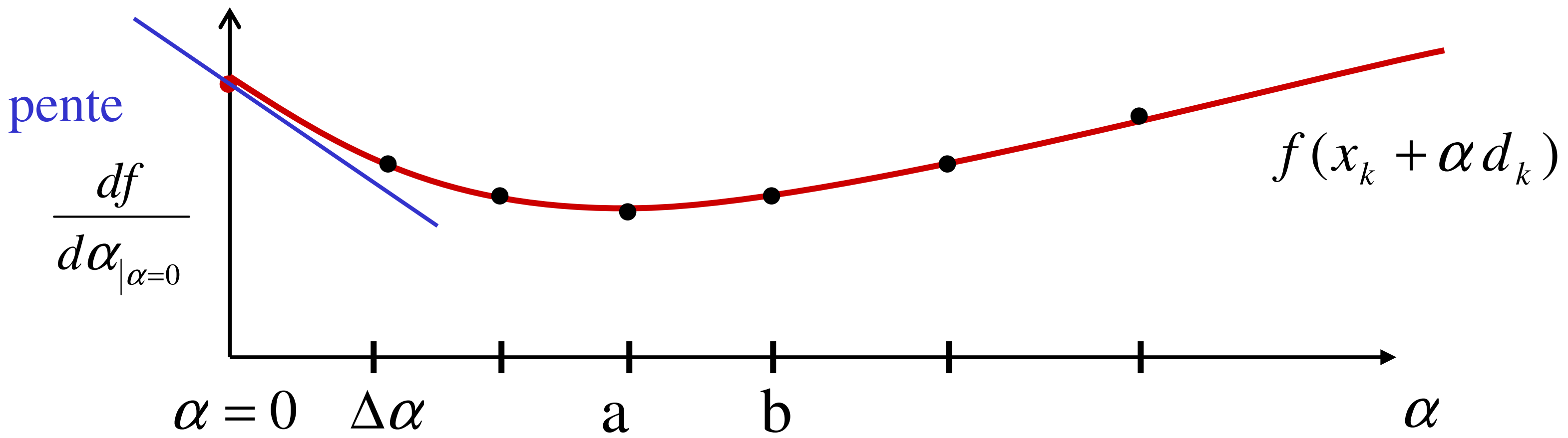
II.1.5.1 La fonction f est dérivable

Soit d_k une direction de décroissance de f

$$\frac{d}{d\alpha} f(x_k + \alpha d_k)\Big|_{\alpha=0} = d_k^T \nabla f(x_k) < 0$$

Déterminer un intervalle contenant α*

1° Trouver un intervalle [a,b] où la pente change de signe

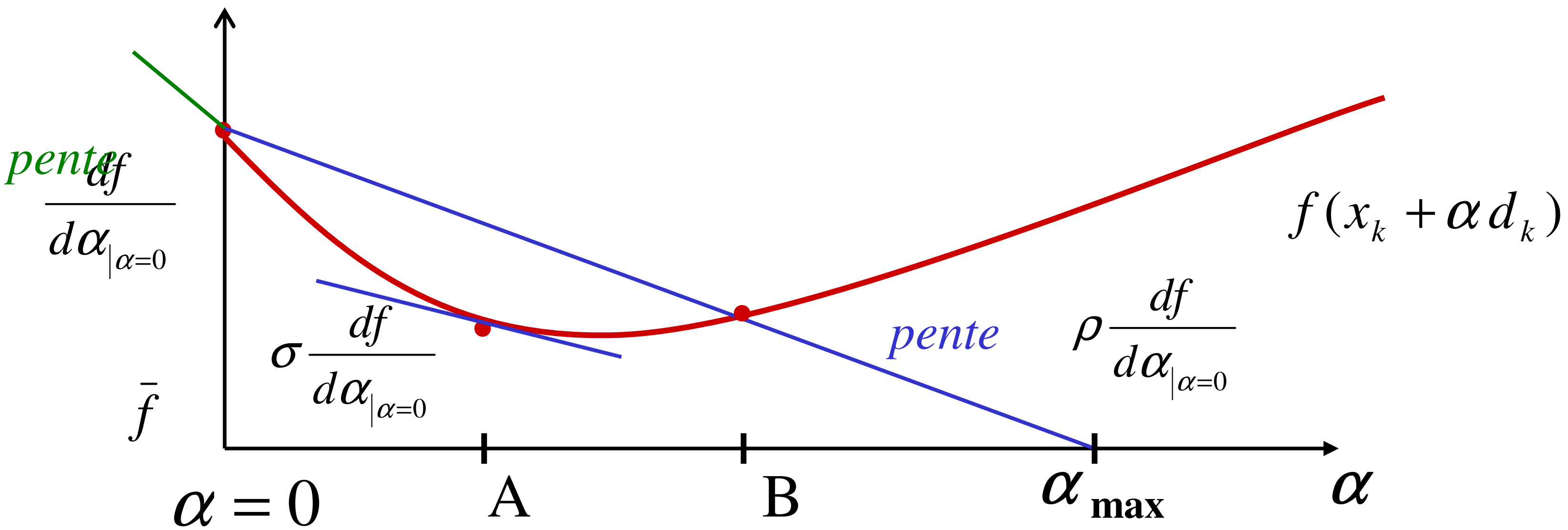


$\Delta\alpha$ mal ajusté ou $\alpha \to \infty$

II. OPTIMISATION SANS CONTRAINTE

II.1 Méthodes de recherches unidimensionnelles (16)

2° Algorithme de Fletcher



Conditions de Wolfe-Powell

$$\begin{cases} 1^\circ \quad f(x_k + \alpha d_k) < f(x_k) + \alpha\rho \dfrac{df}{d\alpha}_{|\alpha=0} & \text{avec } 0 < \rho < \tfrac{1}{2} \\ 2^\circ \quad \left| \dfrac{df}{d\alpha}(x_k + \alpha d_k) \right| < \sigma \left| \dfrac{df}{d\alpha}_{|\alpha=0} \right| & \text{avec } 1 - \rho < \sigma < 1 \end{cases}$$

II. OPTIMISATION SANS CONTRAINTE

II.1 Méthodes de recherches unidimensionnelles (17)

II.1.5.2. La fonction f est non dérivable

1. Utiliser des différences finies au lieu des dérivées.
2. Méthode du nombre d'or ou des suites de Fibonacci si un intervalle $[0, \alpha_{max}]$ a été identifié.
3. Interpolation quadratique en 3 points avec recherche de minimum: méthode souvent utilisée (e.g.. MATLAB).

II. OPTIMISATION SANS CONTRAINTE

II.2 Méthode du Gradient (1)

Soit $f : \mathbf{R}^n \rightarrow \mathbf{R} \quad f \in C^1$

$-\nabla f(x_k)$ Indique la direction avec le plus grand taux de décroissance de f au point x_k

=> Algorithme du Gradient:

$$\boxed{x_{k+1} = x_k - \alpha_k \nabla f(x_k) \quad \alpha_k \geq 0}$$

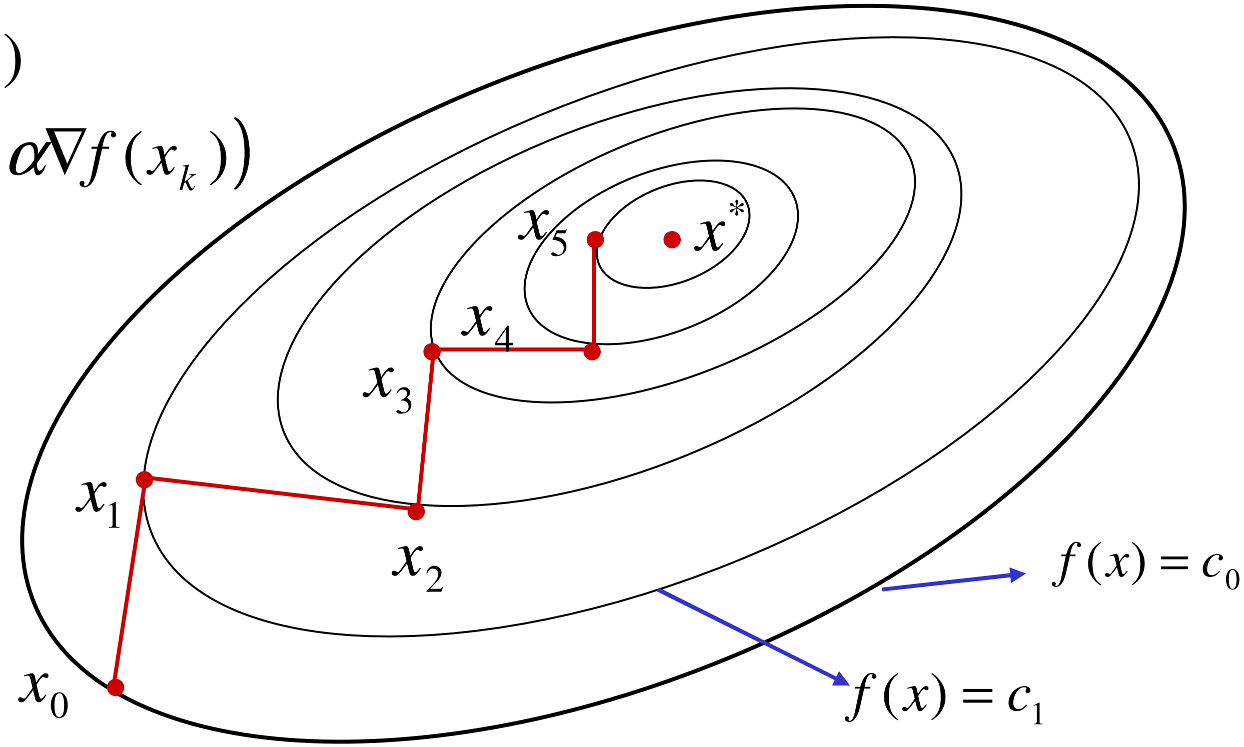
II. OPTIMISATION SANS CONTRAINTE

II.2 Méthode du Gradient (2)

II.2.1 Méthode de Cauchy ou méthode de la plus grande pente

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

$$\alpha_k = \arg\min_{\alpha \geq 0} f\left(x_k - \alpha \nabla f(x_k)\right)$$



Propriétés :

1° $(x_{k+1} - x_k) \perp (x_k - x_{k-1})$

$c_1 < c_0$

2° si $\nabla f(x_k) \neq 0$ alors $f(x_{k+1}) < f(x_k)$

II. OPTIMISATION SANS CONTRAINTE

II.2 Méthode du Gradient (3)

II.2.2 Evaluation du Gradient

Problèmes possibles: * $f \in C^1$, mais calcul du ∇f trop complexe

* ∇f connu, mais nécessite trop de calculs

* $f \notin C^1$

Si $f \in C^1$ => différences finies:

$$\frac{\partial f}{\partial x_i}(x) \approx \frac{f(x+hu_i)-f(x)}{h} \quad \text{avec} \quad u_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow i^{ème} ligne$$

Si $f \notin C^1$ => autres méthodes sans calcul du gradient

II. OPTIMISATION SANS CONTRAINTE

II.2 Méthode du Gradient (4)

II.2.3 Critères d'arrêt

En théorie si $\nabla f(x_k) = 0$

En pratique avant car convergence trop lente

Critères :

1° $\left\| \nabla f(x_k) \right\| < \varepsilon_1$

2° $\left| \dfrac{f(x_{k+1}) - f(x_k)}{f(x_k)} \right| < \varepsilon_2$

3° $\left| x_{k+1} - x_k \right| < \varepsilon_3$

II. OPTIMISATION SANS CONTRAINTE

II.2 Méthode du Gradient (5)

II.2.4 Convergence

Cas d'une fonction quadratique

$$f(x) = \tfrac{1}{2} x^T \mathrm{Q}\, x - b^T x \qquad \mathrm{Q} = \mathrm{Q}^T > 0$$

$$\Rightarrow \nabla f(x) = \mathrm{Q}\, x - b$$

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

$$\alpha_k = \arg\min_{\alpha \geq 0} f(x_k - \alpha \nabla f(x_k))$$

$$\Rightarrow \frac{d}{d\alpha} f(x_k - \alpha \nabla f(x_k))_{|\alpha=\alpha_k} = 0$$

$$\Leftrightarrow \nabla^T f(x_k - \alpha \nabla f(x_k))_{|\alpha=\alpha_k} \nabla f(x_k) = 0$$

$$\Leftrightarrow [\mathrm{Q}(x_k - \alpha_k \nabla f(x_k)) - b]^T \nabla f(x_k) = 0$$

II. OPTIMISATION SANS CONTRAINTE

II.2 Méthode du Gradient (6)

II.2.4 Convergence

Cas d'une fonction quadratique (suite)

$$\Rightarrow \alpha_k = \frac{(Q\,x_k - b)^T \nabla f(x_k)}{\nabla f(x_k)^T\, Q \nabla f(x_k)} = \frac{\nabla f(x_k)^T \nabla f(x_k)}{\nabla f(x_k)^T\, Q \nabla f(x_k)}$$

$$\Rightarrow x_{k+1} = x_k - \frac{(Q\,x_k - b)^T (Q\,x_k - b)}{(Q\,x_k - b)^T\, Q(Q\,x_k - b)}(Q\,x_k - b)$$

$$x^* = Q^{-1} b \quad \text{avec} \quad \nabla f(x^*) = 0 \quad \text{et} \quad H(x^*) = Q > 0$$

$$\Rightarrow \frac{\left\| x_{k+1} - x^* \right\|}{\left\| x_k - x^* \right\|} \leq \sqrt{\frac{\lambda_{\max}(Q)}{\lambda_{\min}(Q)}} - 1$$

Convergence linéaire qui dépend du conditionnement de Q

II. OPTIMISATION SANS CONTRAINTE

II.2 Méthode du Gradient (7)

Exemple 1° $\quad Q = qI \quad \Rightarrow x_{k+1} = x_k - \dfrac{qx_k - b}{q}$

$$\text{soit} \quad x_0 \quad x_1 = x_0 - \frac{qx_0 - b}{q} = \frac{b}{q} = x^*$$

=> *Convergence en une itération*

Exemple 2° $\quad f(x) = 5x_1^2 + \dfrac{x_2^2}{2} - 3(x_1 + x_2)$

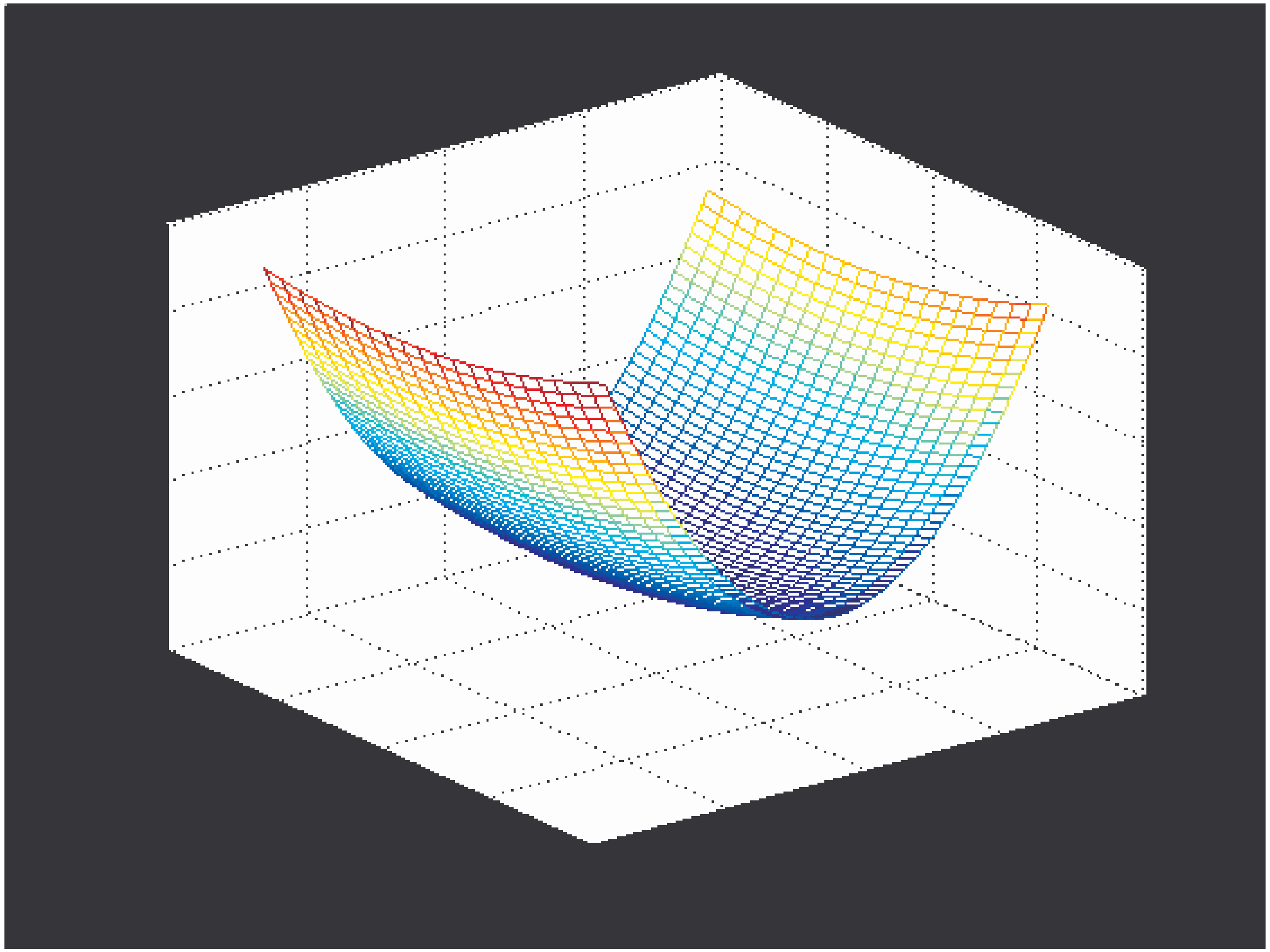
$$Q = \begin{bmatrix} 10 & 0 \\ 0 & 1 \end{bmatrix} \quad b = \begin{bmatrix} 3 \\ 3 \end{bmatrix} \quad x_0 = (-2,-7) \quad x^* = (0{,}3\,,\,3)$$

Critère d'arrêt : $\left| \dfrac{f(x_k) - f(x^*)}{f(x^*)} \right| < 10^{-3}$

=> *15 itérations*

II. OPTIMISATION SANS CONTRAINTE

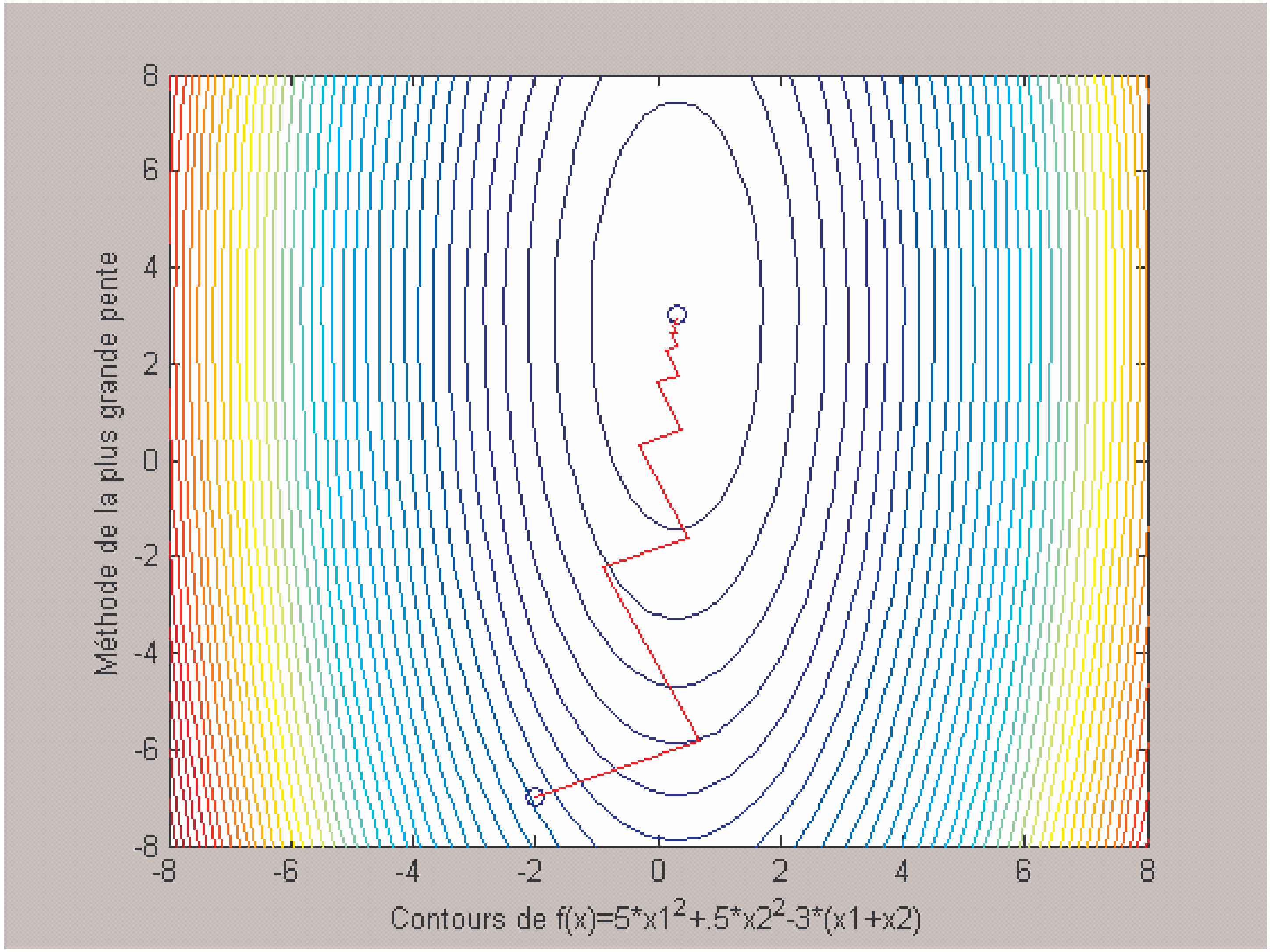
II.2 Méthode du Gradient (8)



$$f(x) = 5x_1^2 + \frac{x_2^2}{2} - 3(x_1 + x_2)$$

II. OPTIMISATION SANS CONTRAINTE

II.2 Méthode du Gradient (9)



II. OPTIMISATION SANS CONTRAINTE

II.2 Méthode du Gradient (10)

Exemple 3°

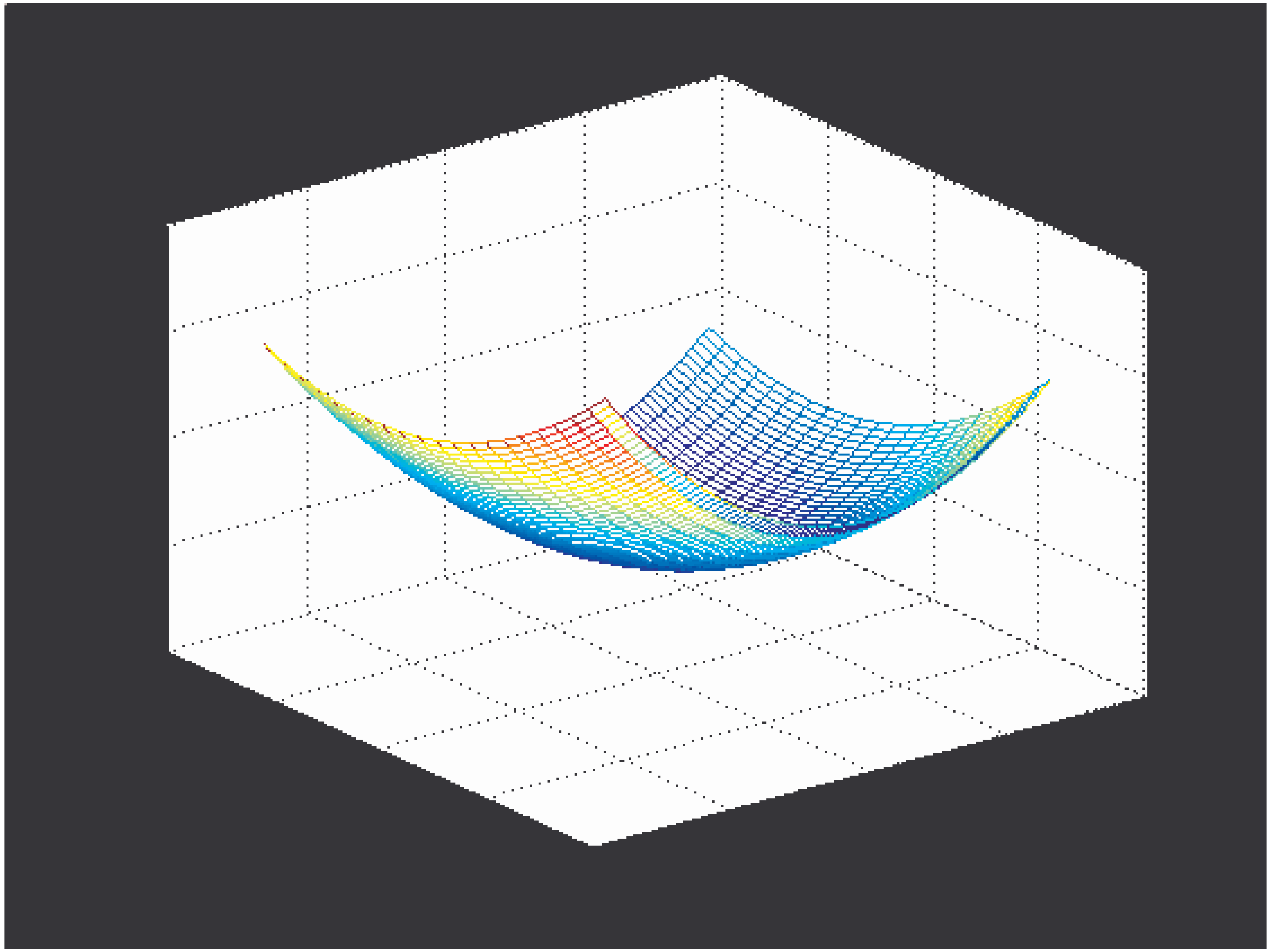
$$f(x) = x_1^2 + \frac{x_2^2}{2} - 3(x_1 + x_2) \quad x_0 = (-2,-7) \quad x^* = (1{,}5\ ,\ 3)$$

$$\mathrm{Q} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \quad b = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

=> *4 itérations* Avec le même critère d'arrêt

II. OPTIMISATION SANS CONTRAINTE

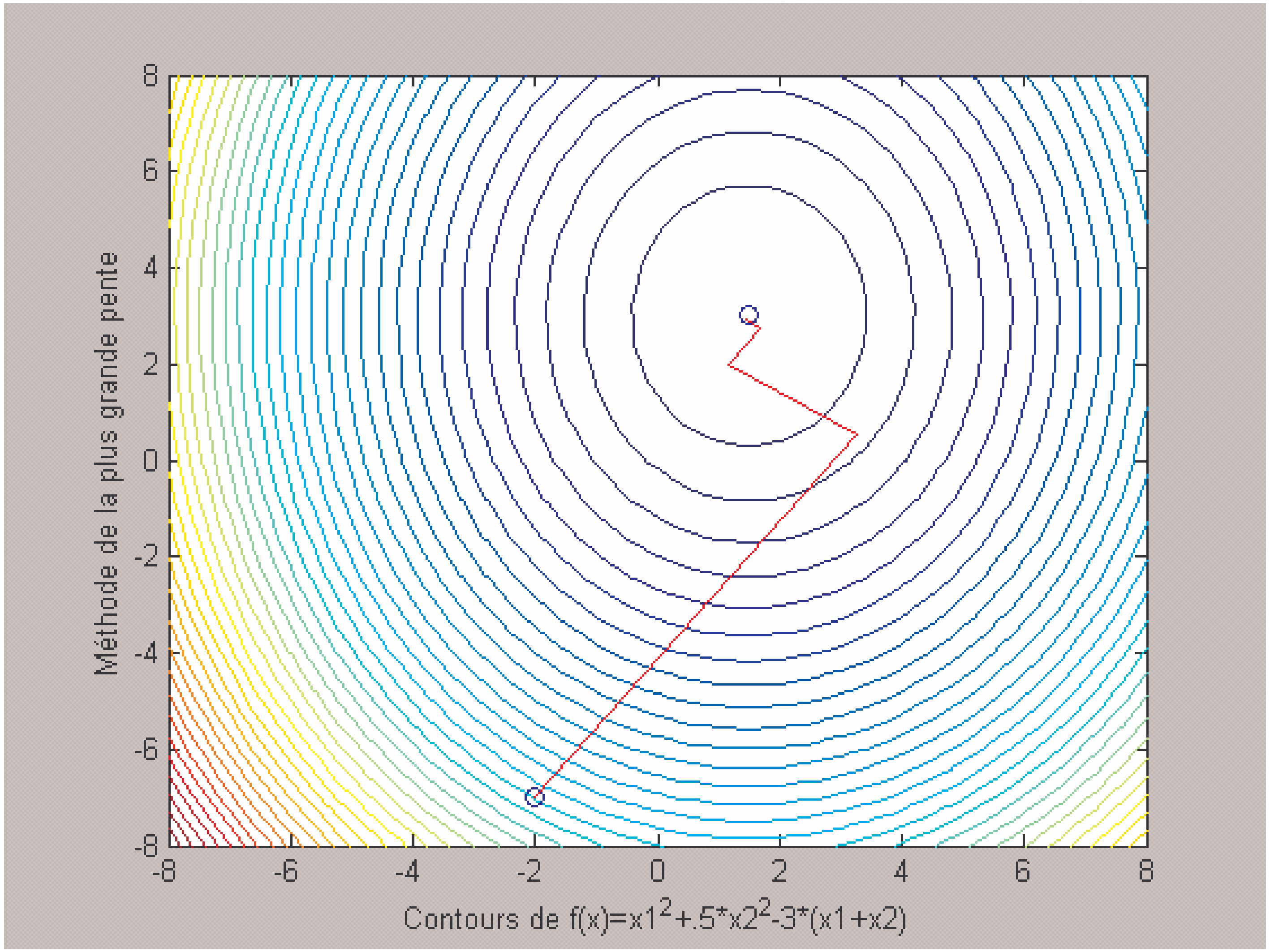
II.2 Méthode du Gradient (11)



$$f(x) = x_1^2 + \frac{x_2^2}{2} - 3(x_1 + x_2)$$

II. OPTIMISATION SANS CONTRAINTE

II.2 Méthode du Gradient (12)



II. OPTIMISATION SANS CONTRAINTE

II.3 Méthode des directions conjuguées (1)

- Résout les problèmes quadratiques en n itérations
- Convergence plus rapide que le gradient

Définition : Directions Q-conjuguées

Soit $Q = Q^T$

Les directions $d_1 \cdots d_n$ sont Q-conjuguées si

$$d_i^T Q d_j = 0 \quad \forall\, i \neq j$$

et linéairement indépendantes si $Q > 0$

II. OPTIMISATION SANS CONTRAINTE

II.3 Méthode des directions conjuguées (2)

II.3.1 Algorithme des directions conjuguées

Soit $f(x) = \frac{1}{2} x^T \mathrm{Q}\, x - b^T x \qquad \mathrm{Q} = \mathrm{Q}^T > 0$

$$\nabla f = \mathrm{Q}\, x - b \qquad x^* = \mathrm{Q}^{-1} b$$

Soient n directions conjuguées $d_0 \cdots d_{n-1}$

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

avec $\alpha_k = \arg\min_{\alpha \geq 0} f(x_k + \alpha d_k) = -\dfrac{\nabla f(x_k)^T}{d_k^T \mathrm{Q}\, d_k} d_k$

En effet :

$$\frac{d}{d\alpha} f(x_k + \alpha d_k)_{|\alpha=\alpha_k} = 0$$

$$\Leftrightarrow \nabla^T f(x_k + \alpha d_k)_{|\alpha=\alpha_k} d_k = 0$$

$$\Leftrightarrow [\mathrm{Q}(x_k + \alpha_k d_k) - b]^T d_k = 0$$

Propriétés:

$\forall x_0$, x_k converge vers x^* en n itérations

II. OPTIMISATION SANS CONTRAINTE

II.3 Méthode des directions conjuguées (3)

II.3.1 Algorithme des directions conjuguées (suite)

Démonstration :

$$\exists \quad n \quad \text{coefficients} \quad \beta_i \;\Big|\; x^* - x_0 = \sum_{i=0}^{n-1} \beta_i d_i$$

$$\Rightarrow d_k^T \, \mathrm{Q}(x^* - x_0) = \beta_k d_k^T \, \mathrm{Q} \, d_k \quad \Rightarrow \beta_k = \frac{d_k^T \, \mathrm{Q}(x^* - x_0)}{d_k^T \, \mathrm{Q} \, d_k}$$

$$x_k = x_0 + \alpha_0 d_0 + \alpha_1 d_1 + \ldots + \alpha_{k-1} d_{k-1}$$

$$x_k - x_0 = \alpha_0 d_0 + \ldots + \alpha_{k-1} d_{k-1}$$

$$x^* - x_0 = x^* - x_k + x_k - x_0$$

$$\Rightarrow d_k^T \, \mathrm{Q}(x^* - x_0) = d_k^T \, \mathrm{Q}(x^* - x_k) = -d_k^T (\mathrm{Q} \, x_k - b) = -d_k^T \nabla f(x_k)$$

$$\Rightarrow \beta_k = -\frac{d_k^T \nabla f(x_k)}{d_k^T \, \mathrm{Q} \, d_k} = \alpha_k \qquad \Rightarrow x_n = x^*$$

II. OPTIMISATION SANS CONTRAINTE

II.3 Méthode des directions conjuguées (4)

II.3.2 Algorithme du gradient conjugué

Soit $f(x) = \frac{1}{2} x^T \mathrm{Q}\, x - x^T b \quad$ avec $\quad \mathrm{Q} = \mathrm{Q}^T > 0$

Principe : Construire les directions conjuguées au fur et à mesure

$$\text{Soit } x_0 \qquad d_0 = -\nabla f(x_0)$$

$$x_1 = x_0 + \alpha_0 d_0 \qquad \alpha_0 = \arg\min_{\alpha \geq 0} f(x_0 + \alpha\, d_0)$$

$$\Rightarrow \alpha_0 = -\frac{\nabla f(x_0)^T d_0}{d_0^{\,T} \mathrm{Q}\, d_0}$$

$$d_1 = -\nabla f(x_1) + \beta_0 d_0 \quad \text{avec} \quad \beta_0 \,\Big|\, d_1^{\,T} \mathrm{Q}\, d_0 = 0$$

$$d_1^{\,T} \mathrm{Q}\, d_0 = \beta_0 d_0^{\,T} \mathrm{Q}\, d_0 - \nabla f(x_1)^T \mathrm{Q}\, d_0$$

$$\Rightarrow \beta_0 = \frac{\nabla f(x_1)^T \mathrm{Q}\, d_0}{d_0^{\,T} \mathrm{Q}\, d_0} \quad \cdots$$

II. OPTIMISATION SANS CONTRAINTE

II.3 Méthode des directions conjuguées (5)

II.3.2 Algorithme du gradient conjugué (suite)

Algorithme: choisir x_0 $\quad d_0 = -\nabla f(x_0)$

1° $\quad \alpha_k = -\dfrac{\nabla f(x_k)^T d_k}{d_k^{\,T}\,\mathrm{Q}\,d_k}$

2° $\quad x_{k+1} = x_k + \alpha_k d_k$ $\qquad$ Si $\nabla f(x_{k+1}) = 0$ => arrêt

3° $\quad \beta_k = \dfrac{\nabla f(x_{k+1})^T\,\mathrm{Q}\,d_k}{d_k^{\,T}\,\mathrm{Q}\,d_k}$

4° $\quad d_{k+1} = -\nabla f(x_{k+1}) + \beta_k d_k$

5° $\quad k = k+1$

Propriétés: $\forall x_0$, $\; x_k$ converge vers x^*
en n itérations

II. OPTIMISATION SANS CONTRAINTE

II.3 Méthode des directions conjuguées (6)

Exemple 1°

$$f(x) = \tfrac{1}{2} x^T \mathrm{Q}\, x - b^T x \quad \mathrm{Q} = \begin{bmatrix} 10 & 0 \\ 0 & 1 \end{bmatrix} \quad b = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

$$= 5x_1^2 + \frac{x_2^2}{2} - 3(x_1 + x_2) \quad \nabla f(x) = \mathrm{Q}\, x - b = \begin{bmatrix} 10x_1 - 3 \\ x_2 - 3 \end{bmatrix}$$

1° $\quad x_0 = \begin{bmatrix} -2 \\ -7 \end{bmatrix} \quad d_0 = -\nabla f(x_0) = \begin{bmatrix} 23 \\ 10 \end{bmatrix}$

2° $\quad \alpha_0 = -\dfrac{\nabla f(x_0)^T d_0}{d_0{}^T \mathrm{Q}\, d_0} = 0{,}1167$

3° $\quad x_1 = x_0 + \alpha_0 d_0 = \begin{bmatrix} 0{,}684 \\ -5{,}833 \end{bmatrix}$

4° $\quad \nabla f(x_1) = \begin{bmatrix} 3{,}8404 \\ -8{,}833 \end{bmatrix}$

II. OPTIMISATION SANS CONTRAINTE

II.3 Méthode des directions conjuguées (7)

Exemple 1° (suite)

5° $\beta_0 = \dfrac{\nabla f(x_1)^T \, \mathrm{Q} \, d_0}{d_0^{\,T} \, \mathrm{Q} \, d_0} = 0{,}1475$

6° $d_1 = -\nabla f(x_1)^T + \beta_0 d_0 = \begin{bmatrix} -0{,}4482 \\ 10{,}3079 \end{bmatrix}$

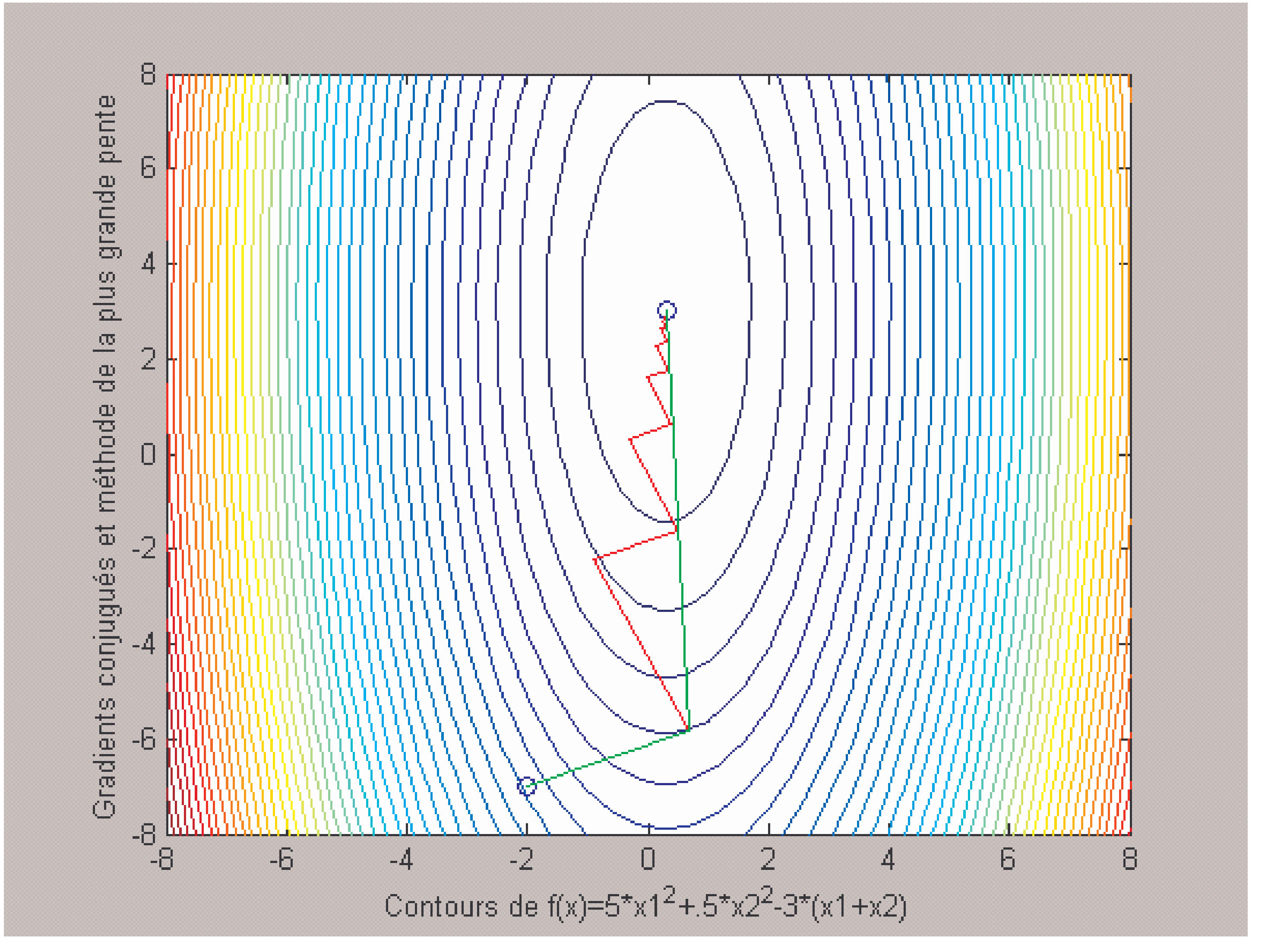
7° $\alpha_1 = -\dfrac{\nabla f(x_1)^T d_1}{d_1^{\,T} \, \mathrm{Q} \, d_1} = 0{,}8569$

8° $x_2 = x_1 + \alpha_1 d_1 = \begin{bmatrix} 0{,}3 \\ 3 \end{bmatrix} = x^*$

II. OPTIMISATION SANS CONTRAINTE

II.2 Méthode des directions conjuguées (8)

Exemple 2°



II. OPTIMISATION SANS CONTRAINTE

II.3 Méthode des directions conjuguées (9)

II.3.3 Gradients conjugués pour les fonctions non quadratiques

Le Hessien de f(x) ≠ Q => éliminer Q dans l'algorithme

$$1° \quad \alpha_k = \arg\min_{\alpha \geq 0} f(x_k + \alpha d_k)$$

Par recherche unidirectionnelle

$$2° \quad \text{modifier la formule } \beta_k = \frac{\nabla f(x_{k+1})^T \mathrm{Q} d_k}{d_k^T \mathrm{Q} d_k}$$

A. Formule de Hestenes-Stiefel

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

$$\Rightarrow \beta_k = \frac{\nabla f(x_{k+1})^T (\nabla f(x_{k+1}) - \nabla f(x_k))}{d_k^T (\nabla f(x_{k+1}) - \nabla f(x_k))}$$

NB: $\nabla f(x_{k+1}) = \mathrm{Q}(x_k + \alpha_k d_k) - b = (\mathrm{Q} x_k - b) + \alpha_k \mathrm{Q} d_k = \nabla f(x_k) + \alpha_k \mathrm{Q} d_k$

II. OPTIMISATION SANS CONTRAINTE

II.3 Méthode des directions conjuguées (10)

II.3.3 Gradients conjugués pour les fonctions non quadratiques

B. Formule de Polak-Ribière

$$d_k^T \nabla f(x_{k+1}) = d_k^T \left[\nabla f(x_k) + \alpha_k \, \mathrm{Q} \, d_k \right] \quad \text{avec} \quad \alpha_k = -\frac{d_k^T \nabla f(x_k)}{d_k^T \, \mathrm{Q} \, d_k}$$

$$= 0$$

et $\quad d_k^T \nabla f(x_k) = -\nabla f(x_k)^T \nabla f(x_k)$

$$\Rightarrow \beta_k = \frac{\nabla f(x_{k+1})^T (\nabla f(x_{k+1}) - \nabla f(x_k))}{\nabla f(x_k)^T \nabla f(x_k)}$$

C. Formule de Fletcher-Reeves

$$\nabla f(x_{k+1})^T \nabla f(x_k) = 0$$

$$\Rightarrow \beta_k = \frac{\nabla f(x_{k+1})^T \nabla f(x_{k+1})}{\nabla f(x_k)^T \nabla f(x_k)}$$

II. OPTIMISATION SANS CONTRAINTE

II.3 Méthode des directions conjuguées (11)

II.3.3 Gradients conjugués pour les fonctions non quadratiques

Algorithme de Fletcher-Reeves: choisir x_0 $\quad d_0 = -\nabla f(x_0)$

1° $\alpha_k = \arg\min_{\alpha \geq 0} f(x_k + \alpha d_k)$

2° $x_{k+1} = x_k + \alpha_k d_k$

3° $\beta_k = \dfrac{\nabla f(x_{k+1})^T \nabla f(x_{k+1})}{\nabla f(x_k)^T \nabla f(x_k)}$

4° $d_{k+1} = -\nabla f(x_{k+1}) + \beta_k d_k$

5° $k = k + 1$

Convergence après n itérations
=> re-initialiser de temps en temps $d_k = -\nabla f(x_k)$

II. OPTIMISATION SANS CONTRAINTE

II.4 Méthode de Newton (1)

Soit $f : \mathbf{R}^{\mathrm{n}} \to \mathbf{R} \quad f \in C^2$

$$f(x) \approx q(x) = f(x_k) + (x - x_k)^T \nabla f(x_k) + \tfrac{1}{2}(x - x_k)^T H(x_k)(x - x_k)$$

$$\nabla q(x) = \nabla f(x_k) + H(x_k)(x - x_k) = 0$$

$\Rightarrow$ si $\quad H(x_k) > 0$

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

Exemple:

$$f(x) = \tfrac{1}{2} x^T \mathrm{Q}\, x - b^T x$$

$$\nabla f(x) = \mathrm{Q}\, x - b \qquad x^* = \mathrm{Q}^{-1} b$$

soit $\quad x_0$

$$x_1 = x_0 - \mathrm{Q}^{-1}(\mathrm{Q}\, x_0 - b) = \mathrm{Q}^{-1} b = x^*$$

=> une seule itération !

II. OPTIMISATION SANS CONTRAINTE

II.4 Méthode de Newton (2)

⚠ fonction pas vraiment quadratique

$H(x_k)$ pas positif $\not> 0$

=> Mauvaise direction : $f(x_{k+1}) \not< f(x_k)$

=> Introduire optimisation unidirectionnelle

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

$$\text{avec} \quad \alpha_k = \arg\min_{\alpha \geq 0} f(x_k - \alpha H^{-1}(x_k)\nabla f(x_k))$$

Calcul du Hessien très lourd !

Propriété locale

II. OPTIMISATION SANS CONTRAINTE

II.5 Algorithme de Levenberg-Marquardt

Combiner Gradient et Newton

$$H_k = H(x_k) + \nu_k I$$

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

ν_k terme de régularisation tel que $H_k > 0$

Algorithme: soit $x_0 \quad \nu_0 \quad \nabla f(x_0) \quad H_0(x_0)$

1° $H_k = H(x_k) + \nu_k I$

2° si $H_k \not> 0 \Rightarrow \nu_k = c_0 \nu_k \Rightarrow 1°$

3° $x_{k+1} = x_k - H_k^{-1} \nabla f(x_k)$

4° calculer $f(x_{k+1})$

5° si $f(x_{k+1}) \geq f(x_k) \quad \nu_k = c_1 \nu_k \quad x_{k+1} = x_k \Rightarrow 1°$

si non $\nu_{k+1} = \dfrac{\nu_k}{c_2}$ et $k = k+1 \quad \Rightarrow 1°$

$c_0, \quad c_1, \quad c_2 > 1$

II. OPTIMISATION SANS CONTRAINTE

II.6 Méthodes Quasi-Newton (1)

$H^{-1}(x_k)$ est remplacé par une approximation M_k

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

avec $\alpha_k = \arg\min_{\alpha \geq 0} f(x_k - \alpha M_k \nabla f(x_k))$

Propriété soit $f \in C^1$

si $M_k = M_k^T > 0 \Rightarrow f(x_{k+1}) < f(x_k)$

Condition quasi-Newton

$$\nabla f(x_{k+1}) - \nabla f(x_k) \approx H(x_k)(x_{k+1} - x_k) \approx H(x_{k+1})(x_{k+1} - x_k)$$

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

$$M_{k+1} \Delta g_k = \Delta x_k$$

II. OPTIMISATION SANS CONTRAINTE

II.6 Méthodes Quasi-Newton (2)

II.6.1 Formule de correction de rang 1

$$M_{k+1} = M_k + cZZ^T$$

$$M_{k+1}\Delta g_k = (M_k + cZZ^T)\Delta g_k = \Delta x_k$$

$$\Leftrightarrow cZ = \frac{\Delta x_k - M_k \Delta g_k}{Z^T \Delta g_k}$$

$$\Rightarrow Z = \Delta x_k - M_k \Delta g_k \quad \text{et} \quad c = \frac{1}{Z^T \Delta g_k}$$

$$\Rightarrow M_{k+1} = M_k + \frac{[\Delta x_k - M_k \Delta g_k][\Delta x_k - M_k \Delta g_k]^T}{[\Delta x_k - M_k \Delta g_k]^T \Delta g_k}$$

M_{k+1} pas toujours > 0

=> Formule de correction de rang 2 : $M_{k+1} = M_k + c_1 Z_1 Z_1^T + c_2 Z_2 Z_2^T$

II. OPTIMISATION SANS CONTRAINTE

II.6 Méthodes Quasi-Newton (3)

II.6.2 Formule de correction de rang 2

A. Davidson-Fletcher-Powell (DFP)

$$M_{k+1} = M_k + \frac{\Delta x_k \Delta x_k^T}{\Delta x_k^T \Delta g_k} - \frac{M_k \Delta g_k \Delta g_k^T M_k}{\Delta g_k^T M_k \Delta g_k}$$

B. Broyden-Fletcher-Goldfarb-Shanno (BFGS)

$$M_{k+1} = M_k + \left(1 + \frac{\Delta g_k^T M_k \Delta g_k}{\Delta g_k^T \Delta x_k}\right) \frac{\Delta x_k \Delta x_k^T}{\Delta x_k^T \Delta g_k} - \frac{M_k \Delta g_k \Delta x_k^T + (M_k \Delta g_k \Delta x_k^T)^T}{\Delta g_k^T \Delta x_k}$$

Propriétés :

* si $f(x)$ est quadratique => convergence en n itérations

* si $f(x)$ quelconque $M_k > 0 \quad \forall k$

=> Propriété de descente garantie !

II. OPTIMISATION SANS CONTRAINTE

II.7 Méthodes sans calcul du gradient (1)

II.7.1 Méthode univariable alternée

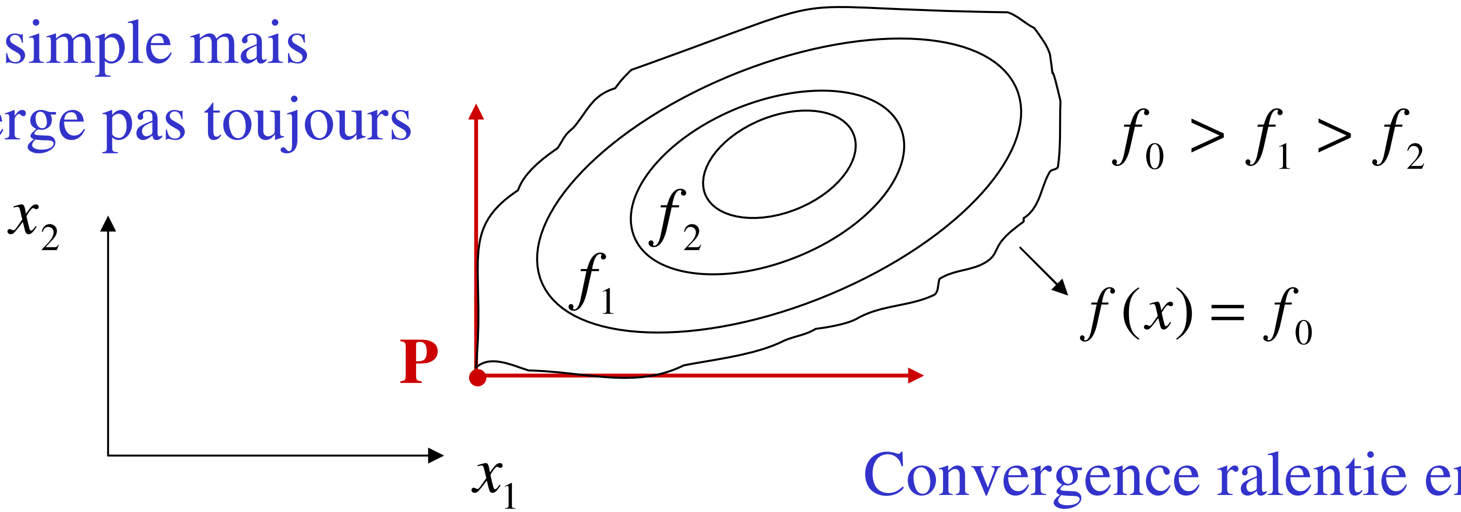
$$\min f(x) \quad \text{avec} \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

1° Minimisation par rapport à x_1 avec $x_i \quad i \neq 1$ constant

2° Minimisation par rapport à x_2 avec $x_i \quad i \neq 2$ constant

…

Méthode simple mais
ne converge pas toujours



Convergence ralentie en P

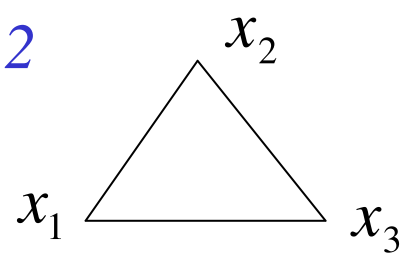
II. OPTIMISATION SANS CONTRAINTE

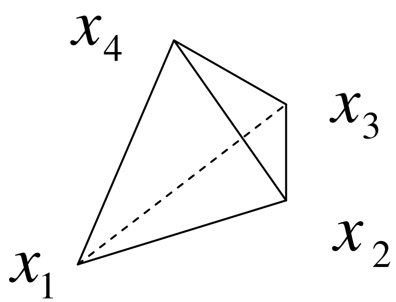
II.7 Méthode sans calcul du gradient (2)

II.7.2 Méthode du simplexe (Nelder-Mead)

Définition : Forme géométrique avec *n+1* sommets dans un espace à *n* dimensions = simplexe

n=2





n=3

Simplexe régulier si les x_i sont équidistants

Principe de l'algorithme :

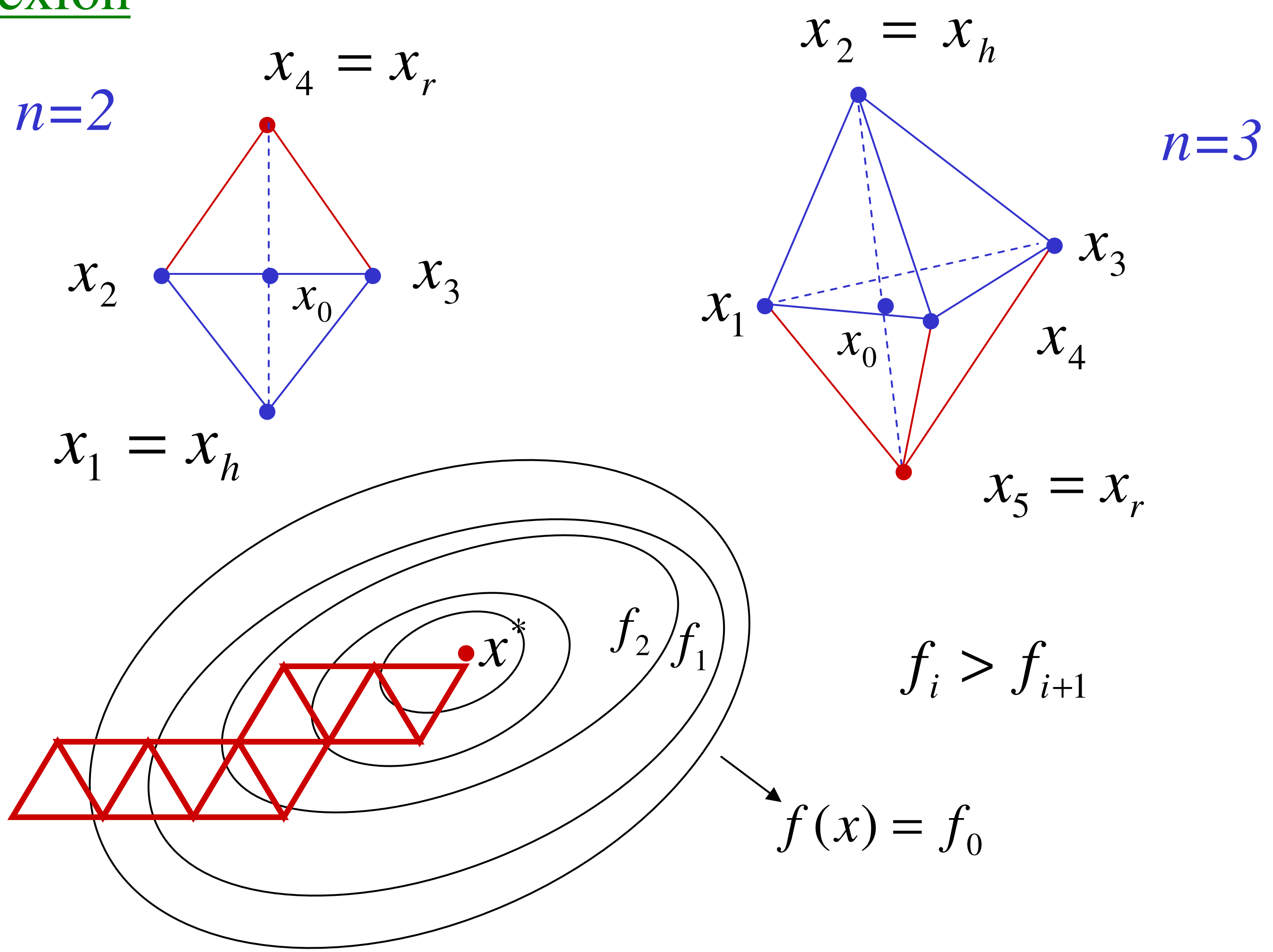
* évaluer $f(x_i)$
* chercher $x_h = \arg\max_{x_i} f(x)$
* chercher le miroir x_r de x_h
* se déplacer dans la direction de x_r

Déplacement constitué de *réflexions, expansions et contractions*

II. OPTIMISATION SANS CONTRAINTE

II.7 Méthode sans calcul du gradient (3)

A. Réflexion



II. OPTIMISATION SANS CONTRAINTE

II.7 Méthode sans calcul du gradient (4)

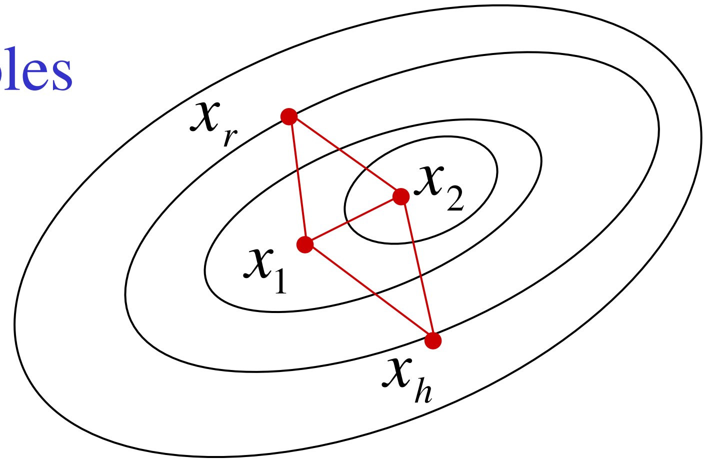
A. Réflexion

$$x_h = \arg\max_{x_i} f(x)$$

$$x_r = x_0 + \alpha(x_0 - x_h) = (1+\alpha)x_0 - \alpha x_h$$

$$x_0 = \frac{1}{n}\sum_{\substack{i=1 \\ i\neq h}}^{n} x_i \qquad \alpha > 0 \quad \alpha = \frac{\|x_r - x_0\|}{\|x_h - x_0\|}$$

⚠ Cycles limites possibles



$\Rightarrow$ *Sélectionner autre x_h

*Contraction

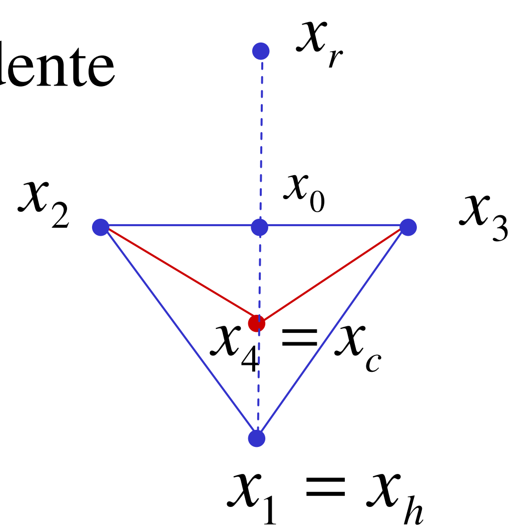
II. OPTIMISATION SANS CONTRAINTE

II.7 Méthode sans calcul du gradient (5)

B. Contraction

si $\quad f(x_r) > f(x_i) \quad \forall i \neq h$

$\Rightarrow$ retour le pas suivant à la position précédente



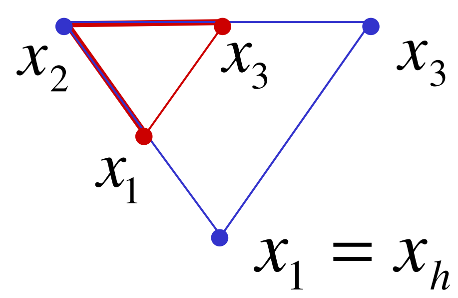
$\Rightarrow x_c = \beta x_h + (1-\beta)x_0 \quad 0 \leq \beta \leq 1$

$$\beta = \frac{\|x_c - x_0\|}{\|x_h - x_0\|}$$

si $\quad f(x_c) < \min(f(x_h), f(x_r))$

$\Rightarrow x_c$ remplace x_h

sinon $\quad x_i$ remplacés par $\dfrac{x_i + \arg\min\limits_{x_i} f(x)}{2}$



II. OPTIMISATION SANS CONTRAINTE

II.7 Méthode sans calcul du gradient (6)

C. Expansion

soit $\quad x_r \,\middle|\, f(x_r) < f(x_i) \quad \forall i$

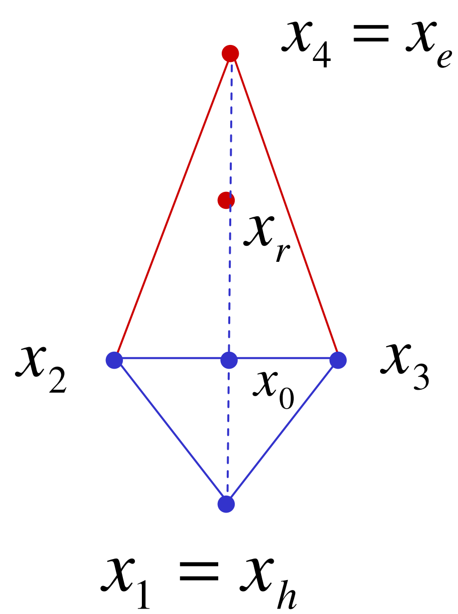
$\Rightarrow$ décroissance importante vers x_r

$$\Rightarrow x_e = \gamma x_r + (1-\gamma) x_0$$

$$\gamma = \frac{\|x_e - x_0\|}{\|x_r - x_0\|} > 1 \quad \text{coefficient d'expansion}$$

si $\quad f(x_e) < \min_{x_i} f(x_i) \Rightarrow x_e$ remplace x_h

sinon $\quad x_r$ remplace x_h



II. OPTIMISATION SANS CONTRAINTE

II.8 Sommes de carrés – Equations non linéaires (1)

II.8.1 Moindres carrés non linéaires

$$f(x) = \frac{1}{2}\sum_{i=1}^{m} f_i(x)^2 = \frac{1}{2}F(x)^T F(x) \quad \text{avec} \quad F(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix}$$

soit la matrice jacobienne $J(x) = \begin{bmatrix} \nabla f_1(x)^T \\ \vdots \\ \nabla f_m(x)^T \end{bmatrix}$

$$\Rightarrow \nabla f(x) = \sum_{i=1}^{m} f_i(x)\nabla f_i(x) = J(x)^T F(x)$$

$$H(x) = J(x)^T J(x) + \sum_{i=1}^{m} f_i(x) H_i(x)$$

II. OPTIMISATION SANS CONTRAINTE

II.8 Sommes de carrés – Equations non linéaires (2)

II.8.1.1 Méthode de Gauss-Newton

Méthode de Newton :

$$x_{k+1} = x_k - \alpha_k H(x_k)^{-1} J(x_k)^T F(x_k)$$

Gauss- Newton :

$$x_{k+1} = x_k - \alpha_k (J(x_k)^T J(x_k))^{-1} J(x_k)^T F(x_k)$$

$$\boxed{\text{si} \quad \lim_{x_k \to x^*} f(x_k) \approx 0 \quad \text{alors} \quad \lim_{x_k \to x^*} H(x_k) \to J(x_k)^T J(x_k)}$$

Convergence quadratique si $f(x^*) = 0$ et $J(x^*)$ non singulier

Problèmes si $J(x)$ singulier ou mal conditionné

II. OPTIMISATION SANS CONTRAINTE

II.8 Sommes de carrés – Equations non linéaires (3)

II.8.1.2 Algorithme de Levenberg-Marquardt

$$x_{k+1} = x_k - \alpha_k \left(J(x_k)^T J(x_k) + \upsilon_k I\right)^{-1} J(x_k)^T F(x_k)$$

$$\upsilon_k > 0 \,\middle|\, f(x_{k+1}) < f(x_k)$$

Problèmes si $f(x^*) >> 0$

II.8.1.3 Méthodes quasi-Newton

$$x_{k+1} = x_k - \alpha_k \left(J(x_k)^T J(x_k) + H_k\right)^{-1} J(x_k)^T F(x_k)$$

H_k approximation quasi-Newton de $\sum_{i=1}^{m} f_i(x_k) H_i(x_k)$

II. OPTIMISATION SANS CONTRAINTE

II.8 Sommes de carrés – Equations non linéaires (4)

II.8.2 Résolution d'équations non-linéaires

Objectif: chercher $x^* \mid F(x^*) = 0$ $\qquad F(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{bmatrix}$

Méthode de Newton-Raphson : Approximation du premier ordre

$$F(x^*) = F(x_k) + J(x_k)(x^* - x_k) + O\left(\left\| x^* - x_k \right\|\right)$$

$$\Rightarrow x_{k+1} = x_k - J(x_k)^{-1} F(x_k) \qquad J(x) = \begin{bmatrix} \nabla f_1(x) \\ \vdots \\ \nabla f_n(x) \end{bmatrix} \neq 0$$

Gauss-Newton pour minimiser $F(x)^T F(x)$

$\Rightarrow$Utilisation possible des méthodes de moindres carrés non linéaires

⚠ Méthode locale avec *J(x)* non singulier

$\exists$ Autres méthodes par intervalles

II. OPTIMISATION SANS CONTRAINTE

II.9 Exemple comparatif (1)

Fonction de Rosenbrock: $f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$

$$\nabla f(x) = \begin{bmatrix} -400(x_2 - x_1^2)x_1 - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{bmatrix}$$

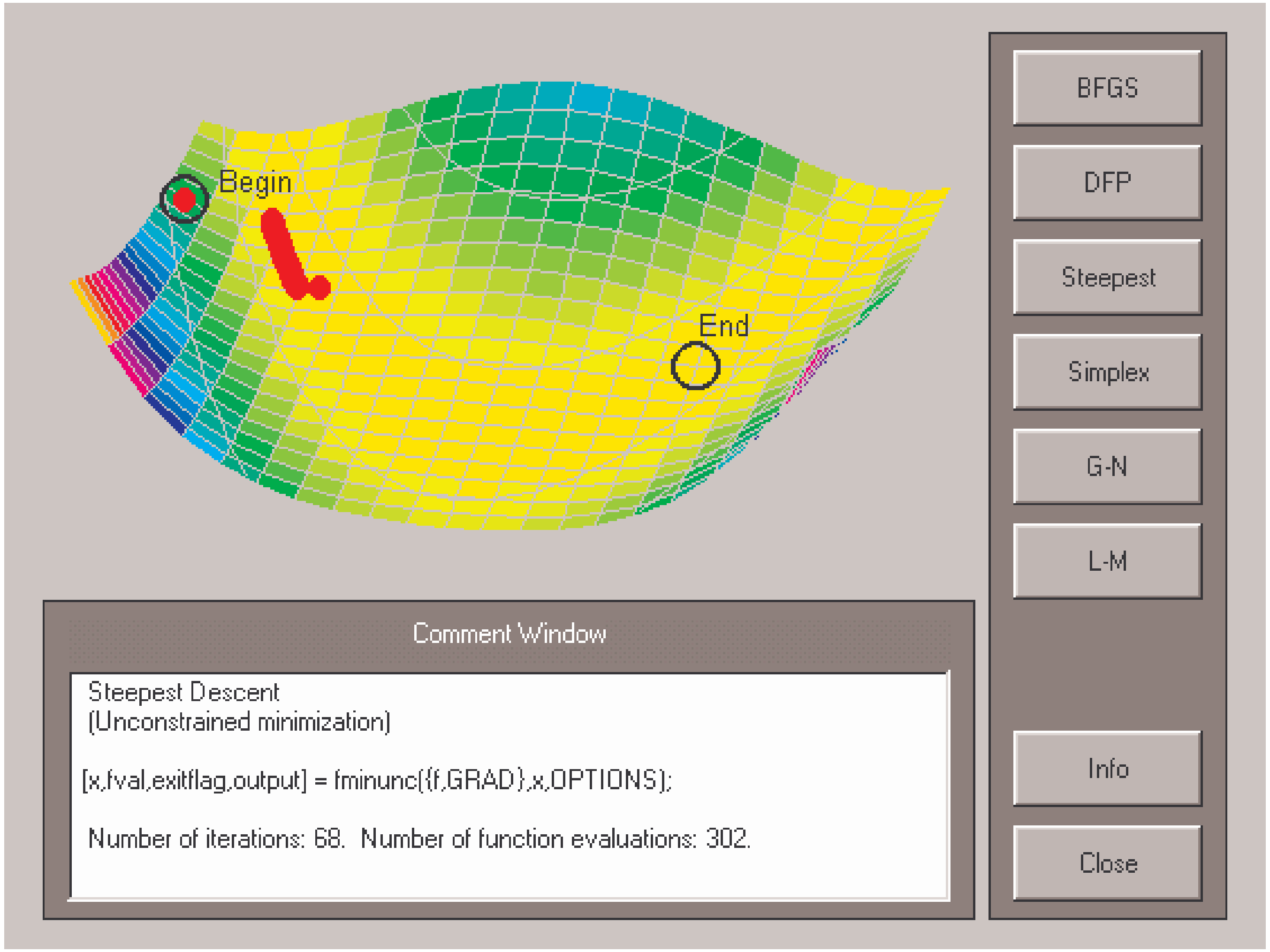
$$\nabla f(x) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{matrix} x_2 = x_1^2 \\ x_1 = 1 \end{matrix} \Leftrightarrow x^* = (1\,,1)$$

$x_0 = (-1{,}9\,,\,2)$

Méthode	# itérations	# évaluations $f(x)$
Gradient	68 + arrêt	302 + arrêt
Simplexe	109	201
Gauss-Newton	13	55
Levenberg-Marquardt	21	92
Quasi-Newton (DFP)	25	109
Quasi-Newton (BFGS)	25	105

II. OPTIMISATION SANS CONTRAINTE

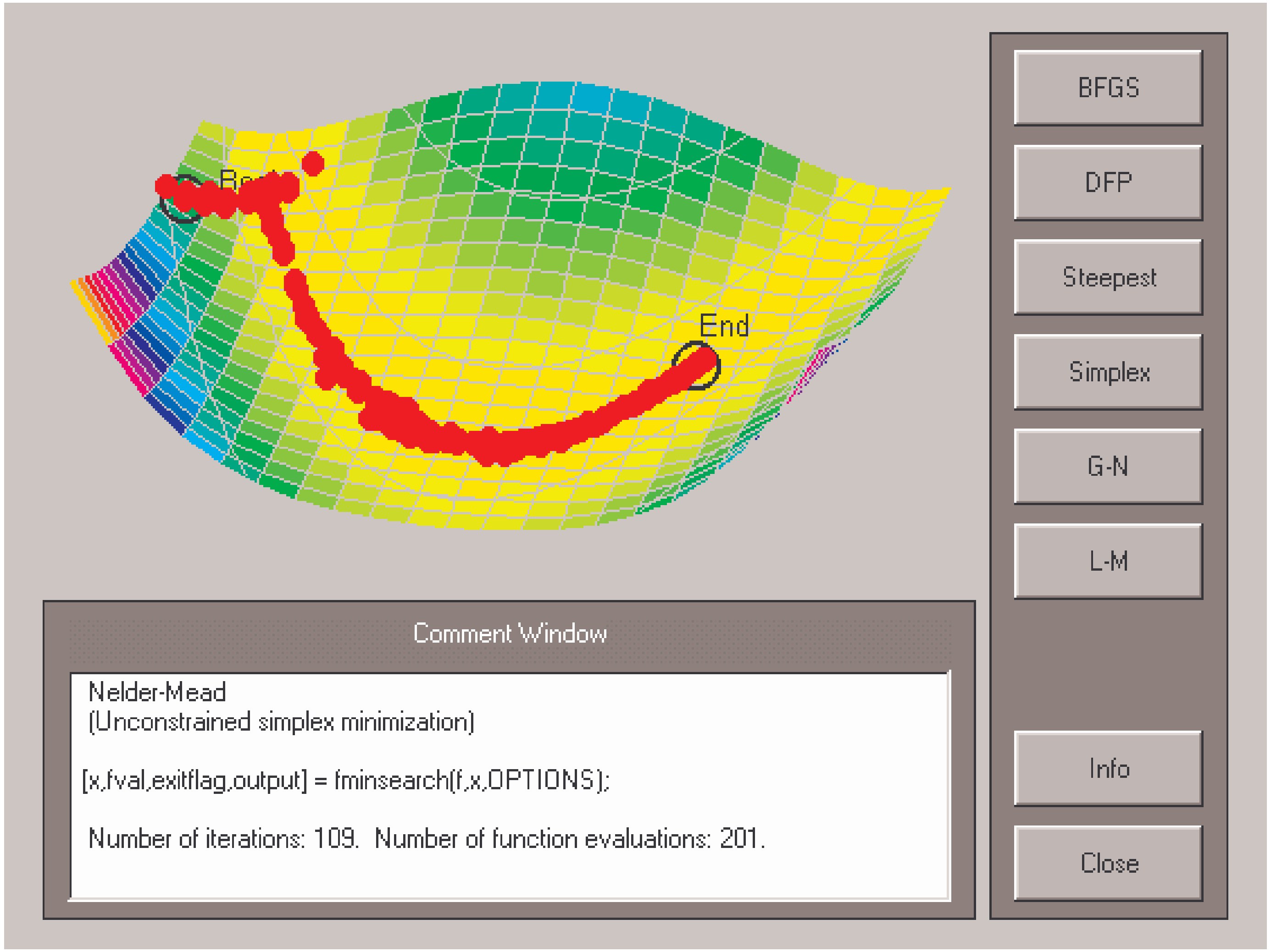
II.9 Exemple comparatif (2)



Méthode du Gradient

II. OPTIMISATION SANS CONTRAINTE

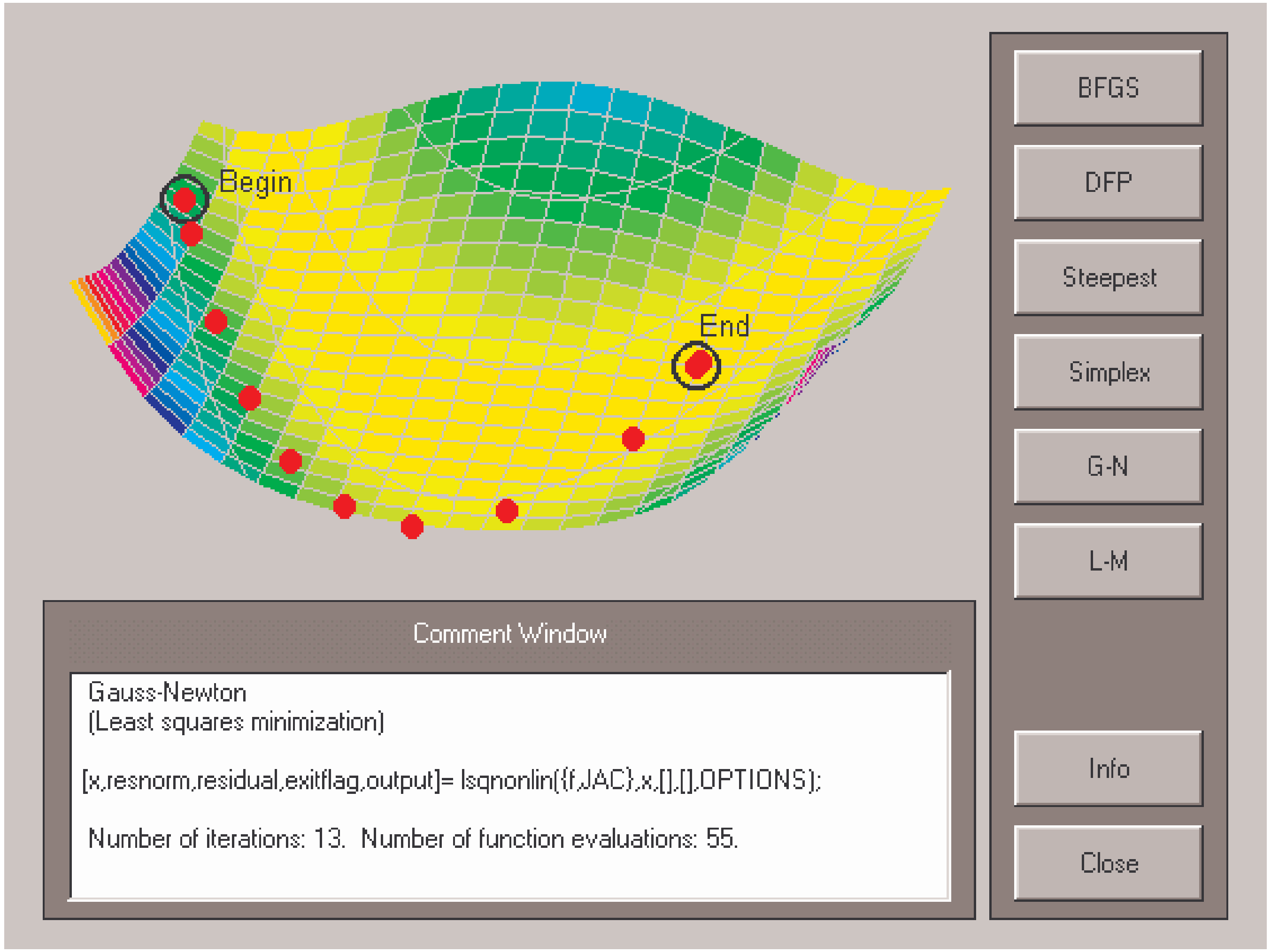
II.9 Exemple comparatif (3)



Méthode du Simplexe

II. OPTIMISATION SANS CONTRAINTE

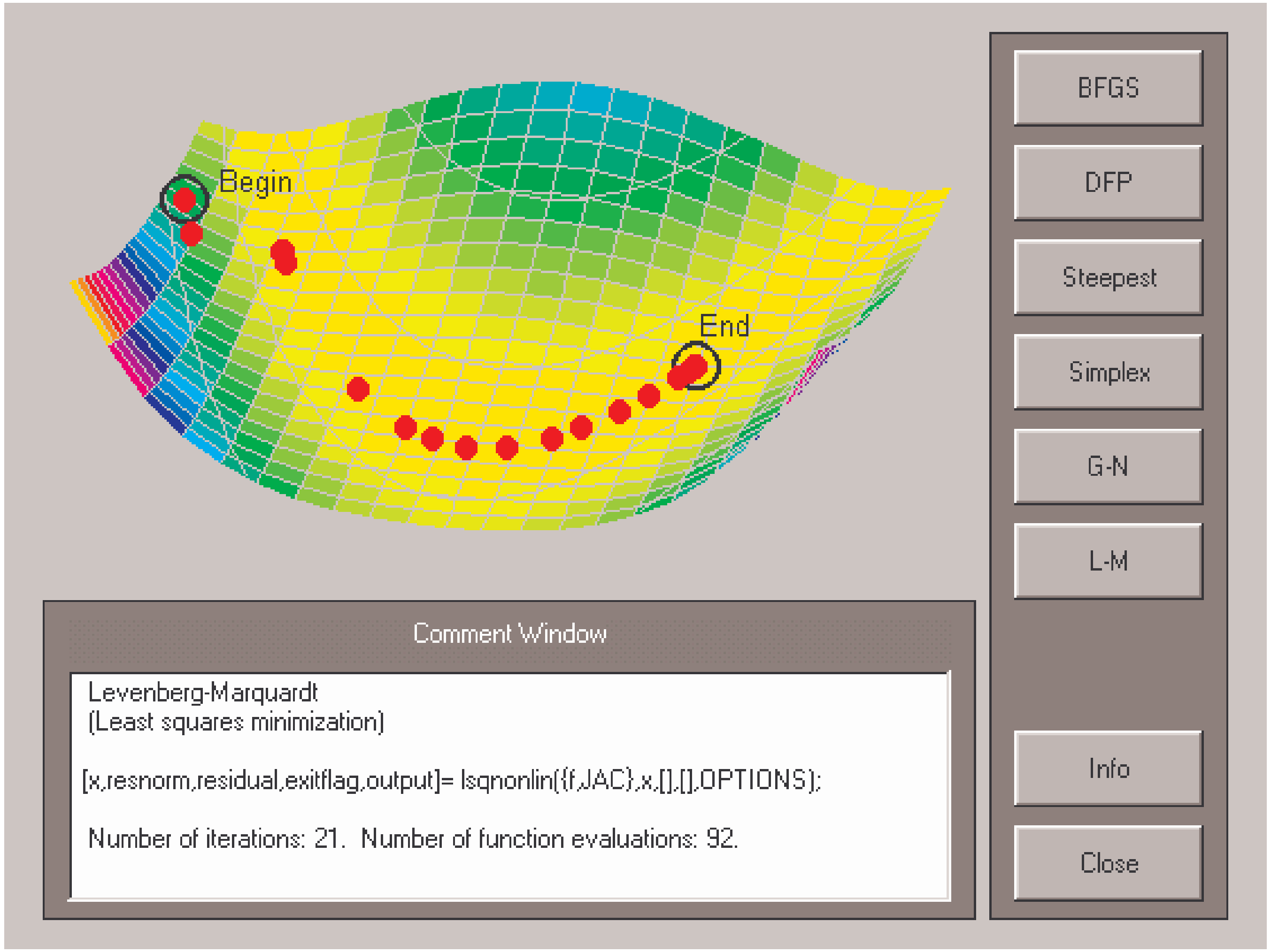
II.9 Exemple comparatif (4)



Méthode de Gauss-Newton

II. OPTIMISATION SANS CONTRAINTE

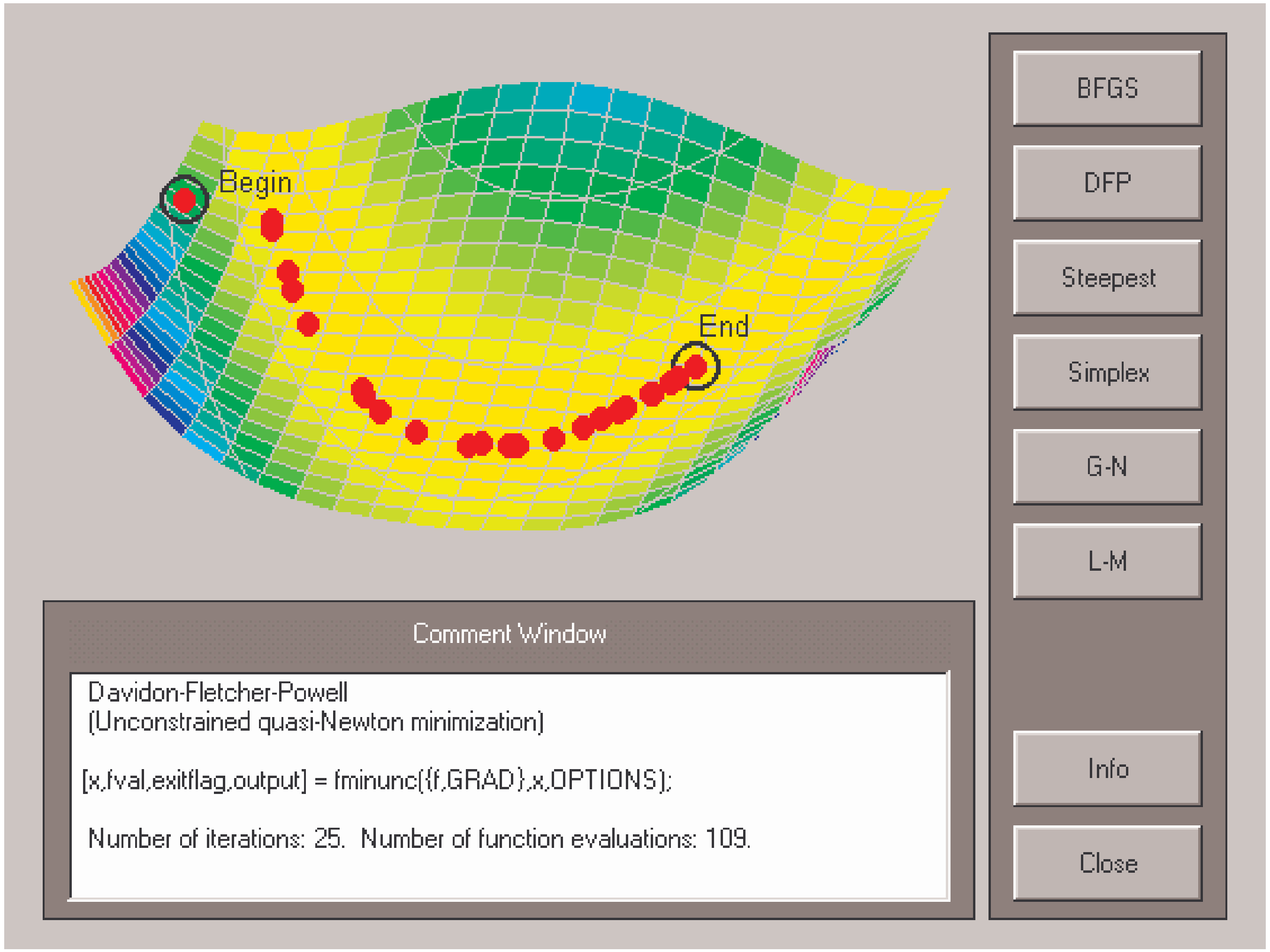
II.9 Exemple comparatif (5)



Méthode de Levenberg-Marquardt

II. OPTIMISATION SANS CONTRAINTE

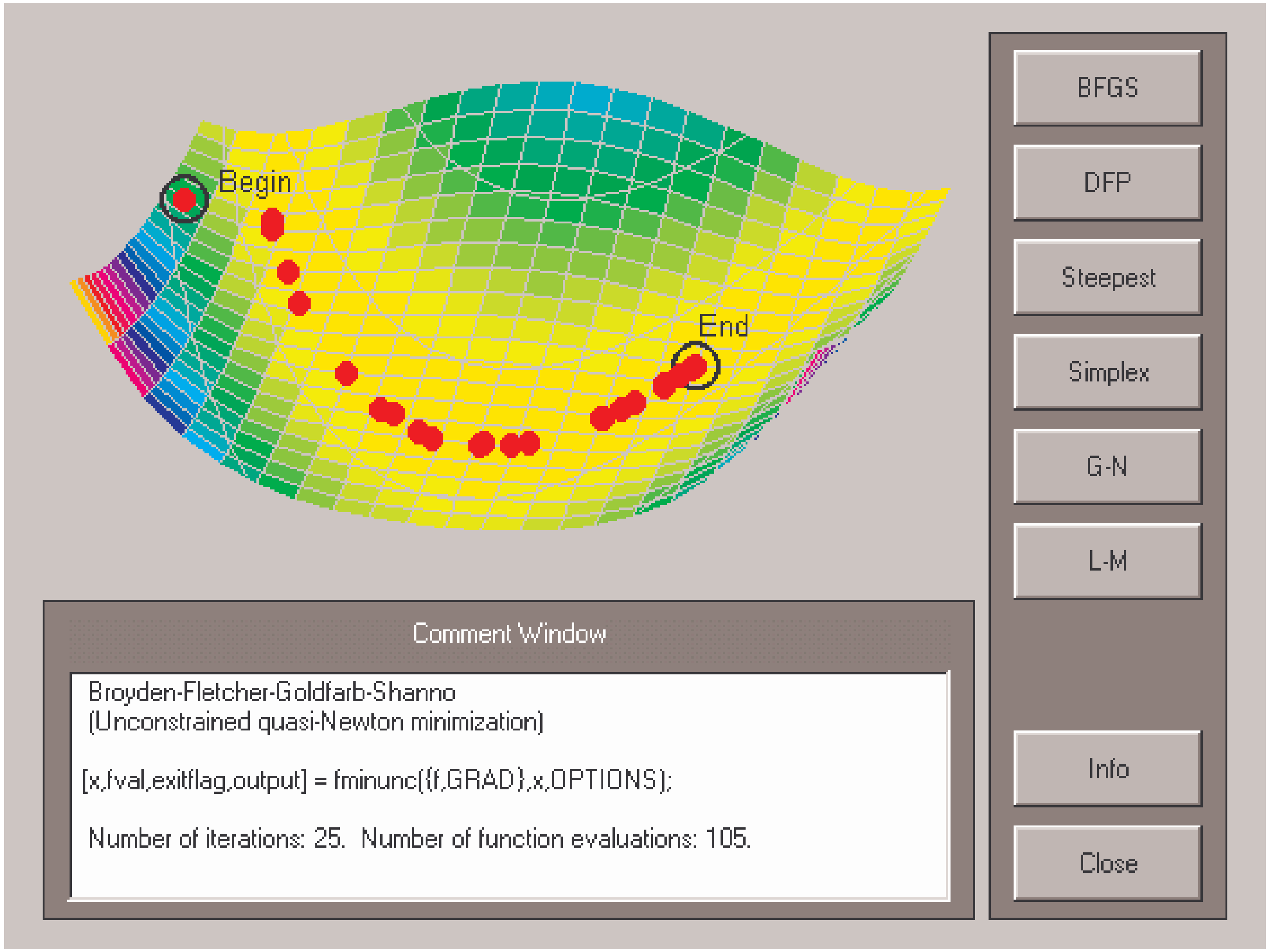
II.9 Exemple comparatif (6)



Méthode Quasi-Newton (DFP)

II. OPTIMISATION SANS CONTRAINTE

II.9 Exemple comparatif (7)



Méthode Quasi-Newton (BFGS)