Méthodes quasi-Newton : BFGS

- Matrice hessienne coûteuse
- Adapter les méthodes quasi-Newton pour les équations à l'optimisation
- Approximer $\nabla^2 f(\widehat{x})$ en utilisant Broyden

$$H_k = H_{k-1} + \frac{(y_{k-1} - H_{k-1} d_{k-1}) d_{k-1}^T}{d_{k-1}^T d_{k-1}},$$

avec

$$\begin{array}{rcl} d_{k-1} & = & x_k - x_{k-1} \\ y_{k-1} & = & \nabla f(x_k) - \nabla f(x_{k-1}). \end{array}$$

Méthodes quasi-Newton : BFGS

- H_k vérifie l'équation sécante
- H_k n'est pas nécessairement symétrique
- H_k n'est pas nécessairement définie positive

On désire forcer H_k à être symétrique et définie positive

$$H_k = L_k L_K^T$$

Idée : travailler sur L_k plutôt que sur H_k.
Mise à jour de L_k en A_k.

Méthodes quasi-Newton : BFGS

Equation sécante (oublions les indices k) :

$$AA^T d = y$$

ou encore

$$\begin{aligned} Ax &= y \\ A^T d &= x \end{aligned}$$

Considérons $Ax = y$ comme équation sécante
Mise à jour de Broyden de L

$$A = L + \frac{(y - Lx)x^T}{x^T x}.$$

(p. 312)

Méthodes quasi-Newton : BFGS

Mise à jour BFGS

Soient une fonction $f : \mathbb{R}^n \to \mathbb{R}$ *continument différentiable, et deux itérés* x_{k-1} *et* x_k*, tels que* $d_{k-1}^T y_{k-1} > 0$*, avec* $d_{k-1} = x_k - x_{k-1}$ *et* $y_{k-1} = \nabla f(x_k) - \nabla f(x_{k-1})$.

Soit une matrice symétrique définie positive $H_{k-1} \in \mathbb{R}^{n \times n}$.

La mise à jour BFGS est définie par

$$H_k = H_{k-1} + \frac{y_{k-1} y_{k-1}^T}{y_{k-1}^T d_{k-1}} - \frac{H_{k-1} d_{k-1} d_{k-1}^T H_{k-1}}{d_{k-1}^T H_{k-1} d_{k-1}}.$$

Méthodes quasi-Newton : BFGS

Elle tient son nom des initiales des mathématiciens C. G. Broyden, R. Fletcher, D. Goldfarb et D. F. Shanno, qui l'ont découverte indépendamment à la fin des années 60.

Méthodes quasi-Newton : BFGS

Lemme 1 *Soit* $d, y \in \mathbb{R}^n$, $d \neq 0$. *Alors il existe une matrice* $A \in \mathbb{R}^{n\times n}$ *non singulière telle que*

$$AA^T d = y$$

si et seulement si

$$d^T y > 0.$$

(p. 314)

Cette condition est toujours vérifiée si la seconde condition de Wolfe est utilisée (p. 315).

Méthodes quasi-Newton : BFGS

- Méthode de Newton avec recherche linéaire
- Remplaçons $\nabla^2 f(x_k)$ par H_k
- Garantie que H_k est définie positive
- Nécessité de résoudre $H_k d_k = -\nabla f(x_k)$
- On peut calculer H_k^{-1} analytiquement

$$H_k^{-1} =$$
$$\left(I - \frac{d_{k-1} y_{k-1}^T}{d_{k-1}^T y_{k-1}}\right) H_{k-1}^{-1} \left(I - \frac{d_{k-1} y_{k-1}^T}{d_{k-1}^T y_{k-1}}\right) + \frac{d_{k-1} d_{k-1}^T}{d_{k-1}^T y_{k-1}}$$

Algorithme : quasi-Newton : BFGS

Objectif

Trouver une approximation d'un minimum local du problème

$$\min_{x \in \mathbb{R}^n} f(x). \tag{1}$$

Input

- La fonction $f : \mathbb{R}^n \to \mathbb{R}$ différentiable;
- Le gradient de la fonction $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$;
- Une première approximation de la solution $x_0 \in \mathbb{R}^n$;

Algorithme : quasi-Newton : BFGS

Input (suite)

- Une première approximation de l'inverse du hessien $H_0^{-1} \in \mathbb{R}^{n \times n}$ symétrique définie positive. Par défaut, $H_0^{-1} = I$.
- La précision demandée $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$.

Output

Une approximation de la solution $x^* \in \mathbb{R}$

Algorithme : quasi-Newton : BFGS

Initialisation

$k = 0$

Itérations

- Calculer $d_k = -H_k^{-1} \nabla f(x_k)$
- Déterminer α_k en appliquant la recherche linéaire avec $\alpha_0 = 1$.
- $x_{k+1} = x_k + \alpha_k d_k$.
- $k = k + 1$.

Algorithme : quasi-Newton : BFGS

Itérations (suite)

- Mettre à jour H_k^{-1}

$$H_k^{-1} =$$
$$\left(I - \frac{\bar{d}_{k-1} y_{k-1}^T}{\bar{d}_{k-1}^T y_{k-1}}\right) H_{k-1}^{-1} \left(I - \frac{\bar{d}_{k-1} y_{k-1}^T}{\bar{d}_{k-1}^T y_{k-1}}\right) + \frac{\bar{d}_{k-1} \bar{d}_{k-1}^T}{\bar{d}_{k-1}^T y_{k-1}}$$

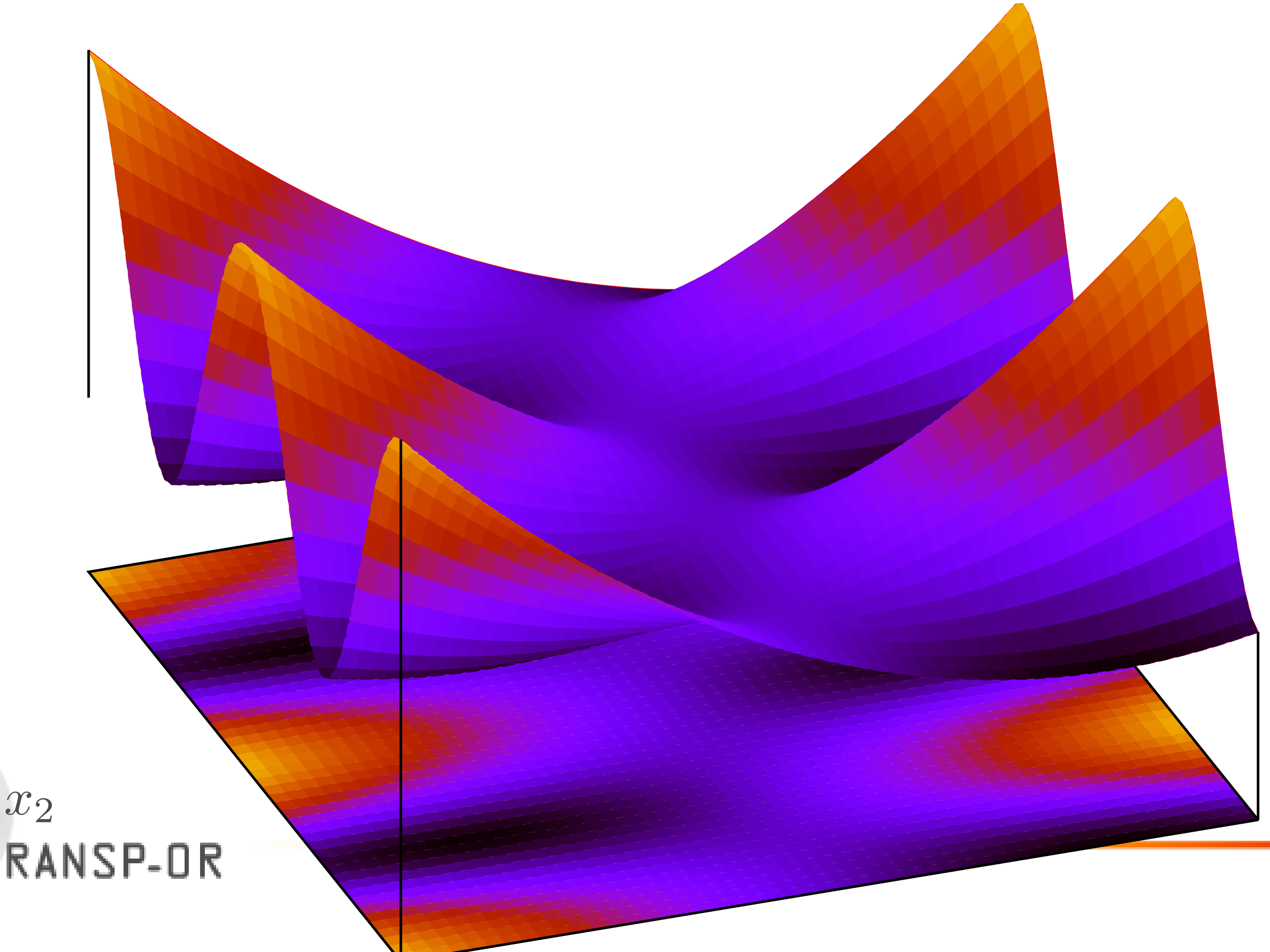
avec $\bar{d}_{k-1} = \alpha_{k-1} d_{k-1} = x_k - x_{k-1}$ et $y_{k-1} = \nabla f(x_k) - \nabla f(x_{k-1})$.

Critère d'arrêt

Si $\|\nabla f(x_k)\| \leq \varepsilon$, alors $x^* = x_k$.

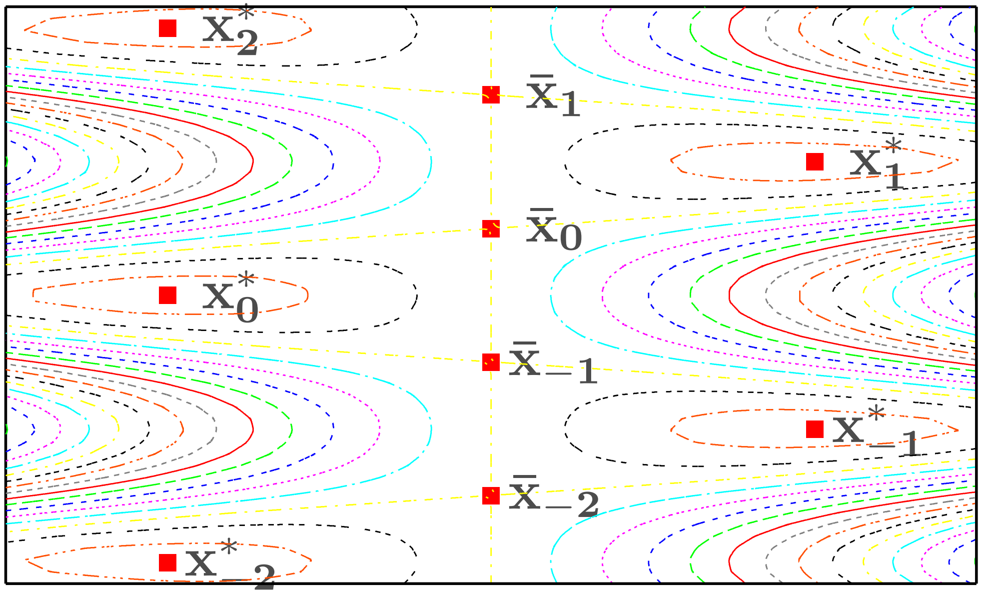
BFGS : exemple

$$\min f(x_1, x_2) = \frac{1}{2}x_1^2 + x_1 \cos x_2,$$



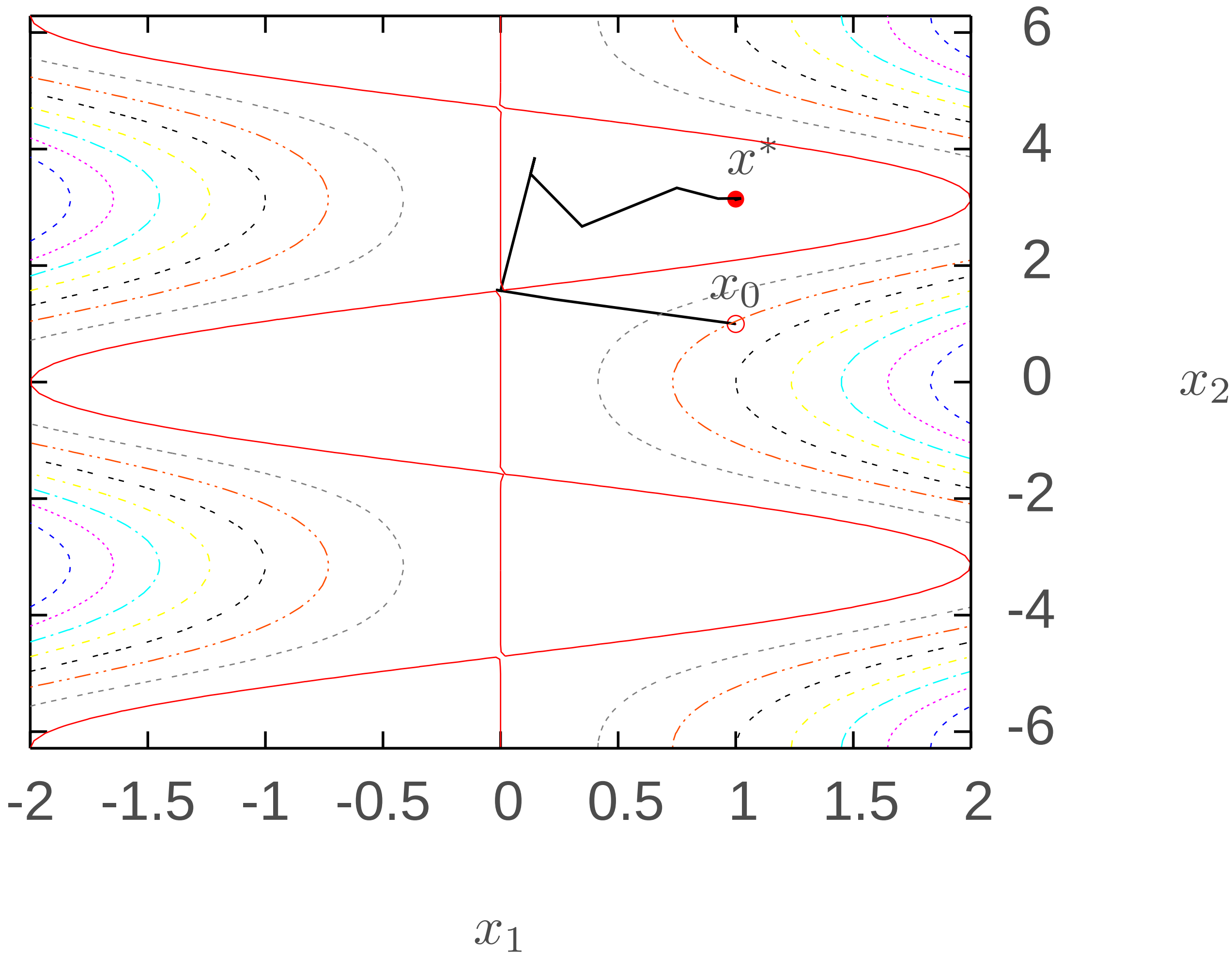
BFGS : exemple

$$\min f(x_1, x_2) = \frac{1}{2}x_1^2 + x_1 \cos x_2,$$

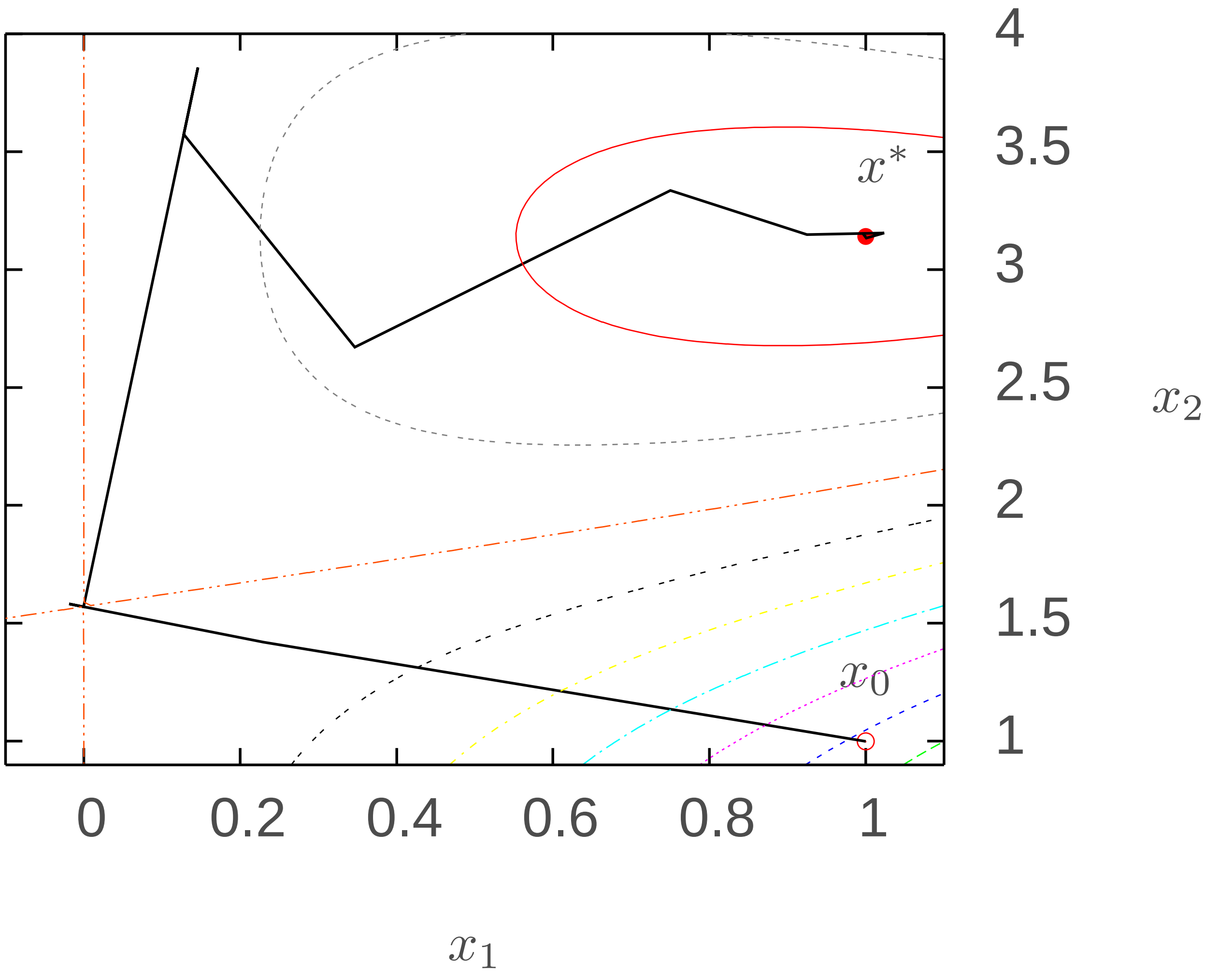


Point de départ $x_0 = (1\ 1)^T$.

BFGS : exemple



BFGS : exemple



Méthodes quasi-Newton : SR1

- BFGS : mise à jour de rang 2 (rang de $H_k - H_{k-1}$)
- SR1 : mise à jour de rang 1

Symmetric Rank 1

$$H_k = H_{k-1} + \beta v v^T,$$

avec $\beta = 1$ ou -1.

(p. 317)

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Méthodes quasi-Newton : SR1

Mise à jour SR1

Soient une fonction $f : \mathbb{R}^n \to \mathbb{R}$ *continument différentiable, et deux itérés* x_{k-1} *et* x_k*, tels que* $d_{k-1}^T y_{k-1} > 0$*, avec* $d_{k-1} = x_k - x_{k-1}$ *et* $y_{k-1} = \nabla f(x_k) - \nabla f(x_{k-1})$.

Soit une matrice symétrique définie positive $H_{k-1} \in \mathbb{R}^{n \times n}$.

La mise à jour SR1 est définie par

$$H_k = H_{k-1} + \frac{(y_{k-1} - H_{k-1}d_{k-1})(y_{k-1} - H_{k-1}d_{k-1})^T}{d_{k-1}^T(y_{k-1} - H_{k-1}d_{k-1})}.$$

Il s'agit d'une mise à jour symétrique de rang 1, ou Symmetric Rank 1 *(SR1) en anglais.*

Méthodes quasi-Newton : SR1

- Bien définie que si

$$d_{k-1}^T(y_{k-1} - H_{k-1}d_{k-1}) \neq 0$$

- Pas nécessairement définie positive
- Préférable de l'utiliser avec la région de confiance

Algorithme : quasi-Newton : SR1

Objectif

Trouver une approximation d'un minimum local du problème

$$\min_{x \in \mathbb{R}^n} f(x). \qquad (2)$$

Input

- La fonction $f : \mathbb{R}^n \to \mathbb{R}$ différentiable;
- Le gradient de la fonction $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$;
- Une première approximation de la solution $x_0 \in \mathbb{R}^n$;
- $H_0 \in \mathbb{R}^{n \times n}$ symétrique. ($H_0 = I$)

Algorithme : quasi-Newton : SR1

Input (suite)

- Le rayon de la première région de confiance Δ_0 (par défaut, $\Delta_0 = 10$).
- La précision demandée $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$.

Output

Une approximation de la solution $x^* \in \mathbb{R}$

Algorithme : quasi-Newton : SR1

Initialisation

$k = 0$, $\eta_1 = 0.01$, $\eta_2 = 0.9$.

Itérations

- Calculer d_k en résolvant (approximativement) le sous-problème de région de confiance en utilisant Steihaug-Toint
- Calculer

$$\rho = \frac{f(x_k) - f(x_k + d_k)}{m_{x_k}(x_k) - m_{x_k}(x_k + d_k)}$$

Algorithme : quasi-Newton : SR1

Itérations (suite)

- Si $\rho < \eta_1$, alors
  - $x_{k+1} = x_k$,
  - $\Delta_{k+1} = \frac{1}{2}\|d_k\|$.
- Si $\rho \geq \eta_1$, alors
  - $x_{k+1} = x_k + d_k$,
  - Si $\rho \geq \eta_2$, alors
    - $\Delta_{k+1} = 2\Delta_k$

    sinon
    - $\Delta_{k+1} = \Delta_k$
- $k = k + 1$.

Algorithme : quasi-Newton : SR1

Itérations (suite)

- Poser $\bar{d}_{k-1} = x_k - x_{k-1}$ et $y_{k-1} = \nabla f(x_k) - \nabla f(x_{k-1})$.
- Si $|\bar{d}_{k-1}(y_{k-1} - H_{k-1}d_{k-1})| \geq 10^{-8}\|\bar{d}_{k-1}\|\|y_{k-1} - H_{k-1}d_{k-1}\|$, alors

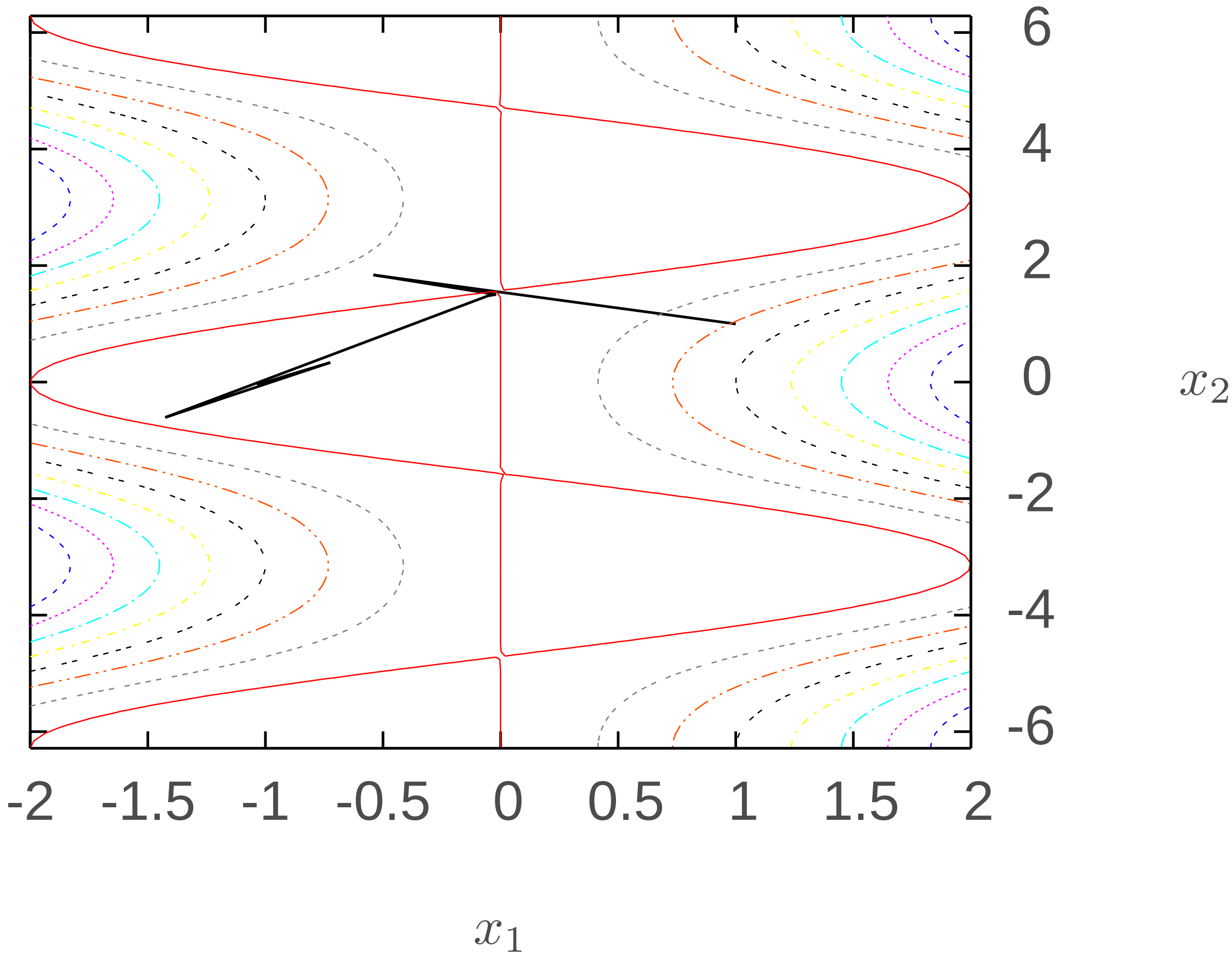
$$H_k = H_{k-1} + \frac{(y_{k-1} - H_{k-1}\bar{d}_{k-1})(y_{k-1} - H_{k-1}\bar{d}_{k-1})^T}{\bar{d}_{k-1}^T(y_{k-1} - H_{k-1}\bar{d}_{k-1})}.$$

  sinon $H_k = H_{k-1}$.

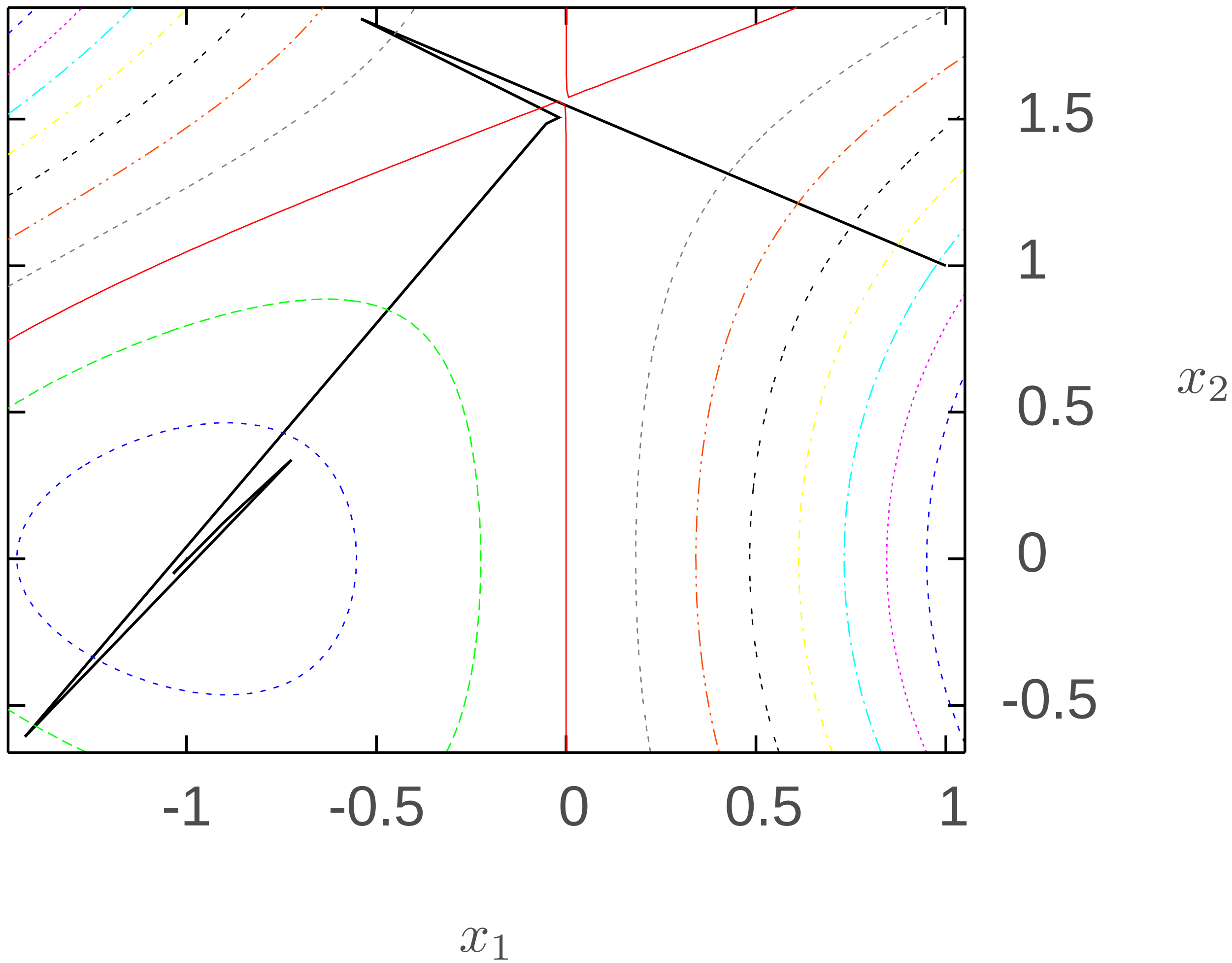
Critère d'arrêt

Si $\|\nabla f(x_k)\| \leq \varepsilon$, alors $x^* = x_k$.

SR1 : exemple



SR1 : exemple



k	$(H_k)_{1,1}$	$(H_k)_{2,2}$	$(H_k)_{1,2}$	λ_1	λ_2
1	+1.38780e+00	+1.16113e+00	-2.49977e-01	+1.00000e+00	+1.54894e+00
2	+1.50852e+00	+1.18087e+00	-2.01166e-01	+1.08526e+00	+1.60413e+00
3	+9.92963e-01	-1.42553e-02	-9.86122e-01	-6.17922e-01	+1.59663e+00
4	+1.56592e+00	+5.95949e-01	-3.94837e-01	+4.55548e-01	+1.70632e+00
5	+9.81728e-01	+8.85544e-03	-9.80476e-01	-5.99219e-01	+1.58980e+00
6	+1.33046e+00	+1.28506e+00	-3.13357e-01	+9.93580e-01	+1.62194e+00
7	+9.95038e-01	+3.68905e-02	-9.60395e-01	-5.57287e-01	+1.58922e+00
8	+1.87015e+00	+3.76238e-01	-4.15446e-01	+2.68479e-01	+1.97791e+00
9	+1.07848e+00	+2.20504e-01	-7.66576e-01	-2.28956e-01	+1.52794e+00
10	+1.86165e+00	+2.21270e-01	-7.42082e-01	-6.46135e-02	+2.14753e+00
11	+5.29518e-01	-4.83122e-01	+2.26599e-01	-5.31516e-01	+5.77911e-01
12	+5.29604e-01	-2.67251e-03	+2.20180e-01	-8.19450e-02	+6.08876e-01
13	+5.56284e-01	+1.50541e-01	+1.56244e-01	+9.73485e-02	+6.09477e-01
14	+5.85590e-01	+3.94032e-01	+2.40717e-01	+2.30739e-01	+7.48883e-01
15	+8.04073e-01	+1.08584e+00	-1.48061e-01	+7.40579e-01	+1.14934e+00
16	+1.30324e+00	+1.09341e+00	-8.66119e-02	+1.06228e+00	+1.33437e+00
17	+1.00670e+00	+1.08398e+00	-3.37434e-02	+9.94037e-01	+1.09664e+00
18	+9.99334e-01	+9.54417e-01	-2.85711e-03	+9.54236e-01	+9.99515e-01
19	+9.99555e-01	+9.99873e-01	+3.11648e-04	+9.99364e-01	+1.00006e+00
20	+1.00066e+00	+1.00000e+00	-6.64095e-05	+9.99995e-01	+1.00067e+00

Méthodes quasi-Newton

- BFGS + région de confiance
- SR1 + recherche linéaire (Cholesky modifiée)
- Autres formules proposées, notamment Davidon, Fletcher, Powell (DFP)