Méthodes Quasi-Newton

Michel Bierlaire

michel.bierlaire@epfl.ch

Laboratoire Transport et Mobilité

EPFL - ENAC - TRANSP-OR

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

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

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

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

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)

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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).

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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

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

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 $x_0 \in \mathbb{R}^n$;

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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 $x^* \in \mathbb{R}$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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

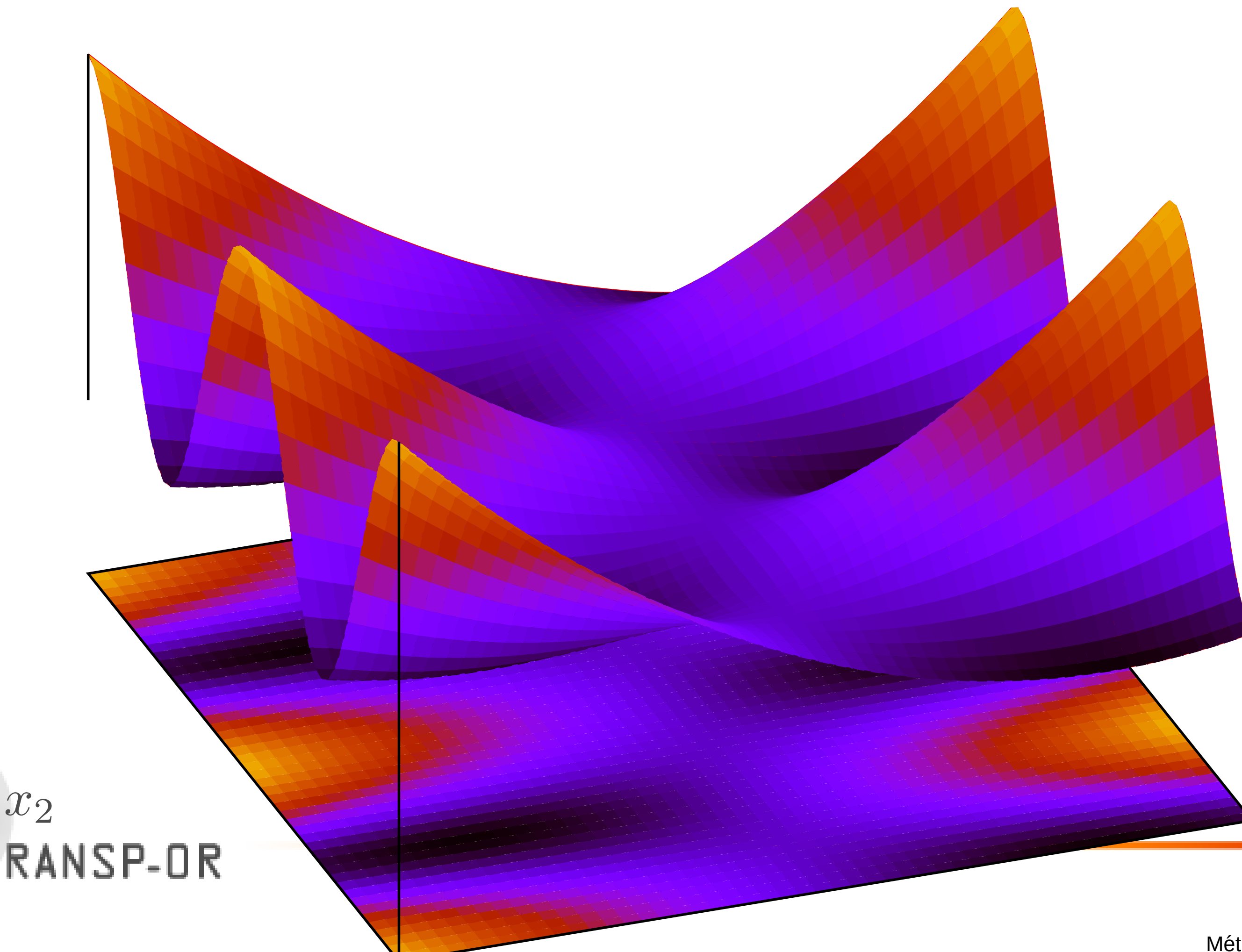
Si $\|\nabla f(x_k)\| \leq \varepsilon$, alors $x^* = x_k$.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

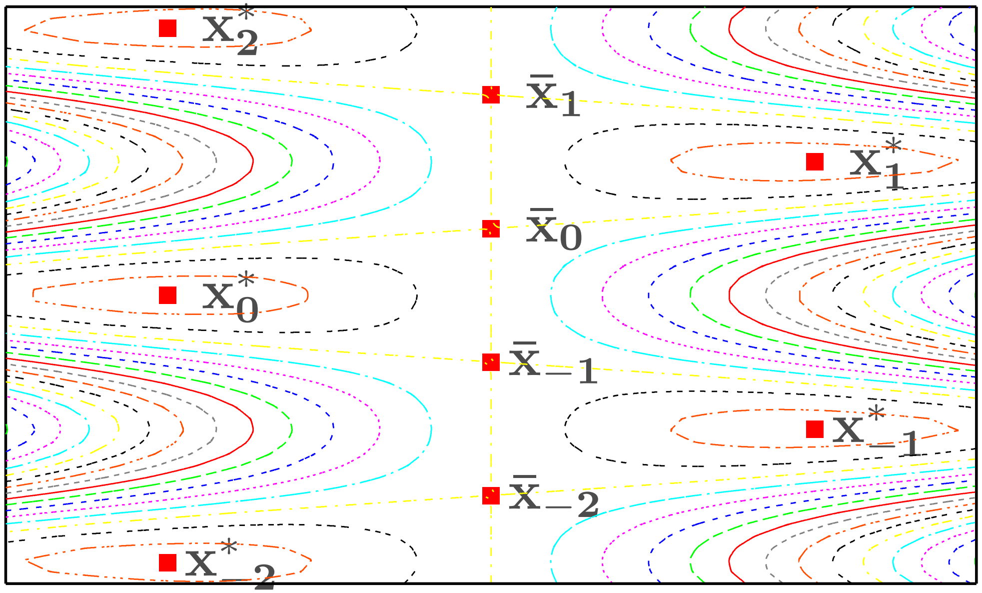
BFGS : exemple

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



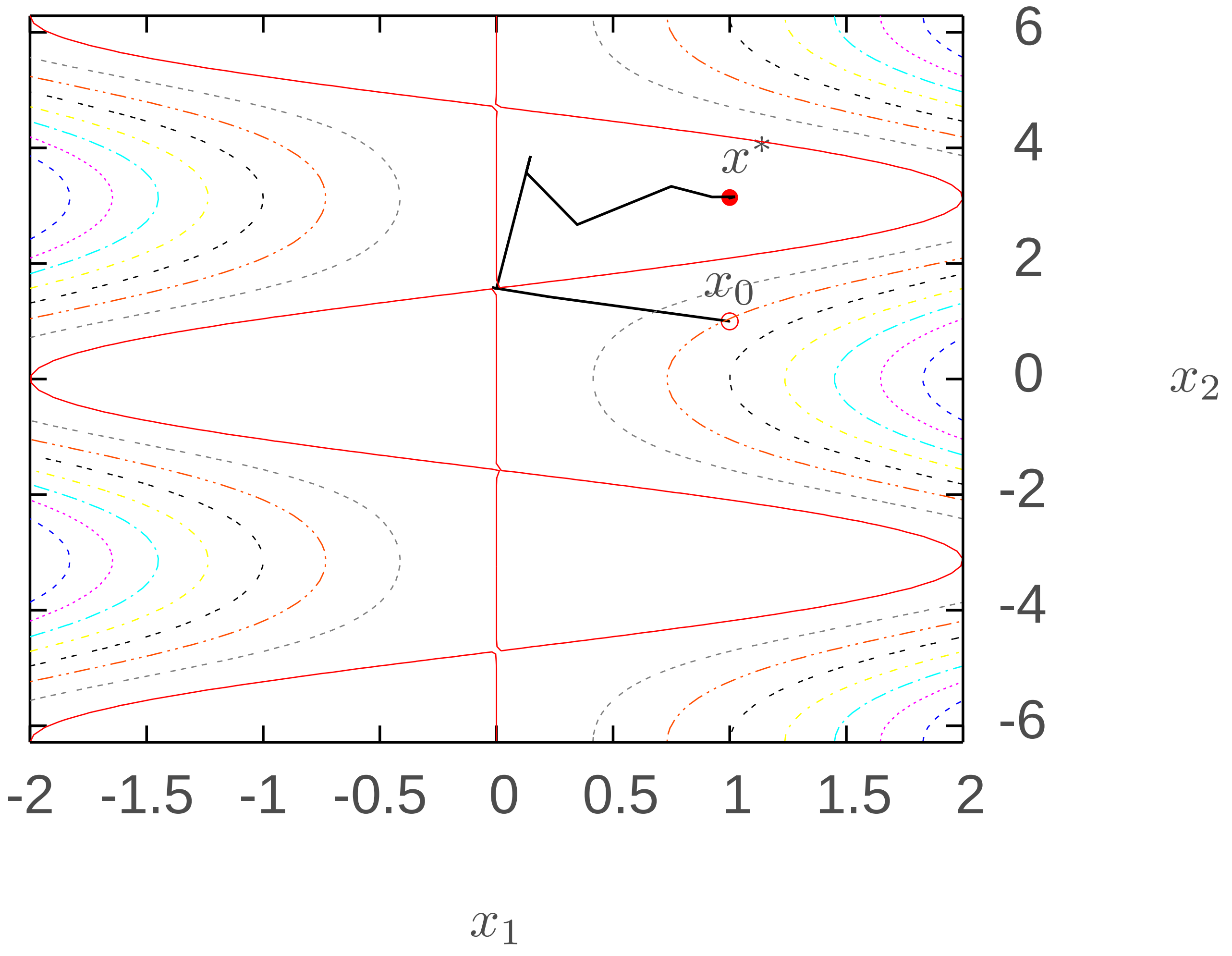
BFGS : exemple

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

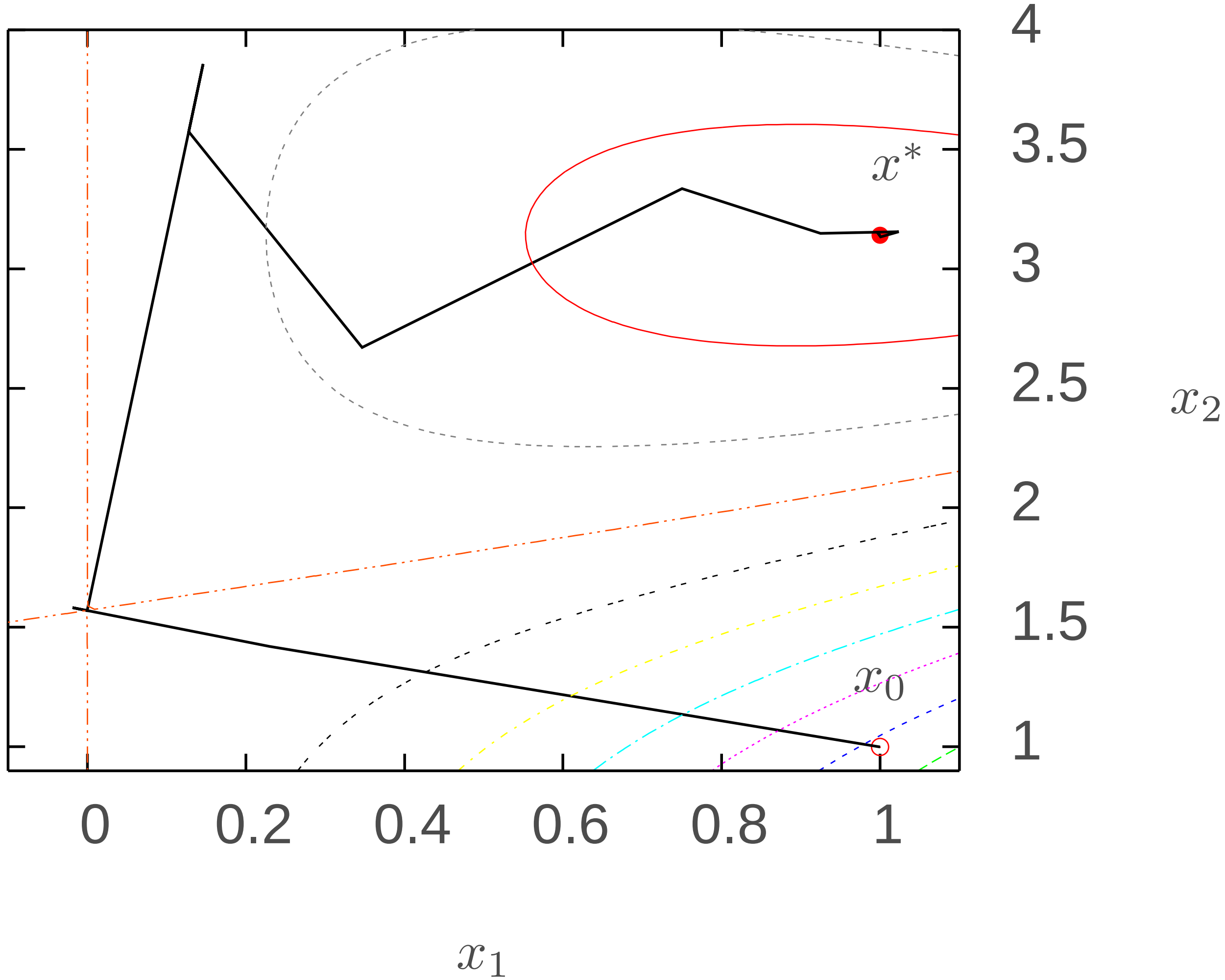


Point de départ $x_0 = (1\ 1)^T$.

BFGS : exemple



BFGS : exemple



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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)

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 (voir Chap. 12)

Méthodes quasi-Newton

- BFGS
- SR1
- Autres formules proposées, notamment Davidon, Fletcher, Powell (DFP)