Méthode de Newton

Michel Bierlaire

michel.bierlaire@epfl.ch

Laboratoire Transport et Mobilité

EPFL - ENAC - TRANSP-OR

Newton locale

- Conditions nécessaires d'optimalité

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

- Il s'agit d'un système d'équations non linéaires.
- Appliquons la méthode de Newton pour les équations.
- Voir Bierlaire (2006), chapitre 7 pour rappel.

Equation à une inconnue

Résoudre

$$F(x) = 0$$

avec

$$F(x) = x^2 - 2, \quad \widehat{x} = 2$$

Théorème de Taylor

$$\begin{aligned} F(\widehat{x} + d) &= F(\widehat{x}) + dF'(\widehat{x}) + o(|d|) \\ &= \widehat{x}^2 - 2 + 2\widehat{x}d + o(|d|) \\ &= 2 + 4d + o(|d|). \end{aligned}$$

Equation à une inconnue

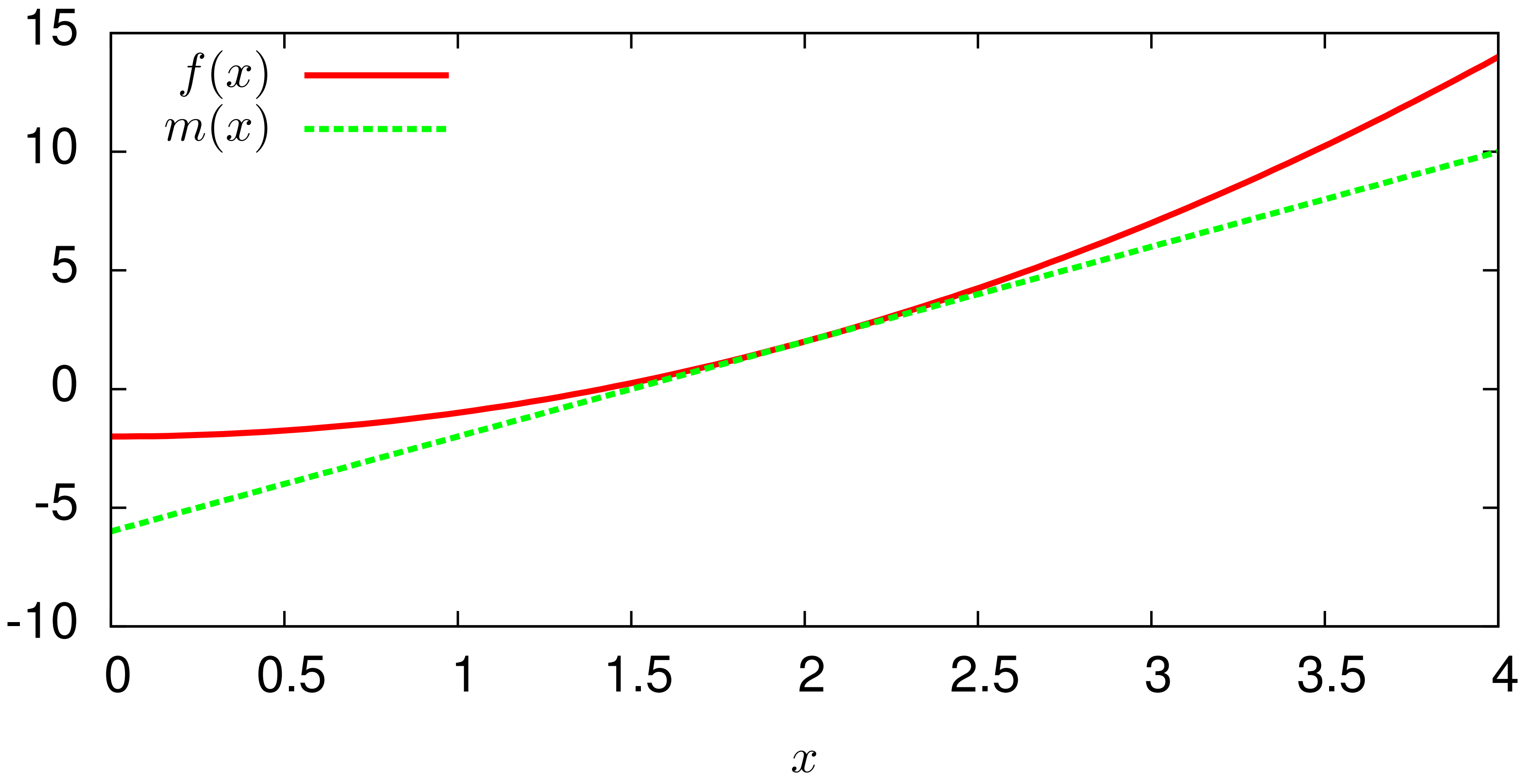
Ignorons le terme d'erreur pour obtenir un modèle :

$$m(\widehat{x} + d) = 2 + 4d.$$

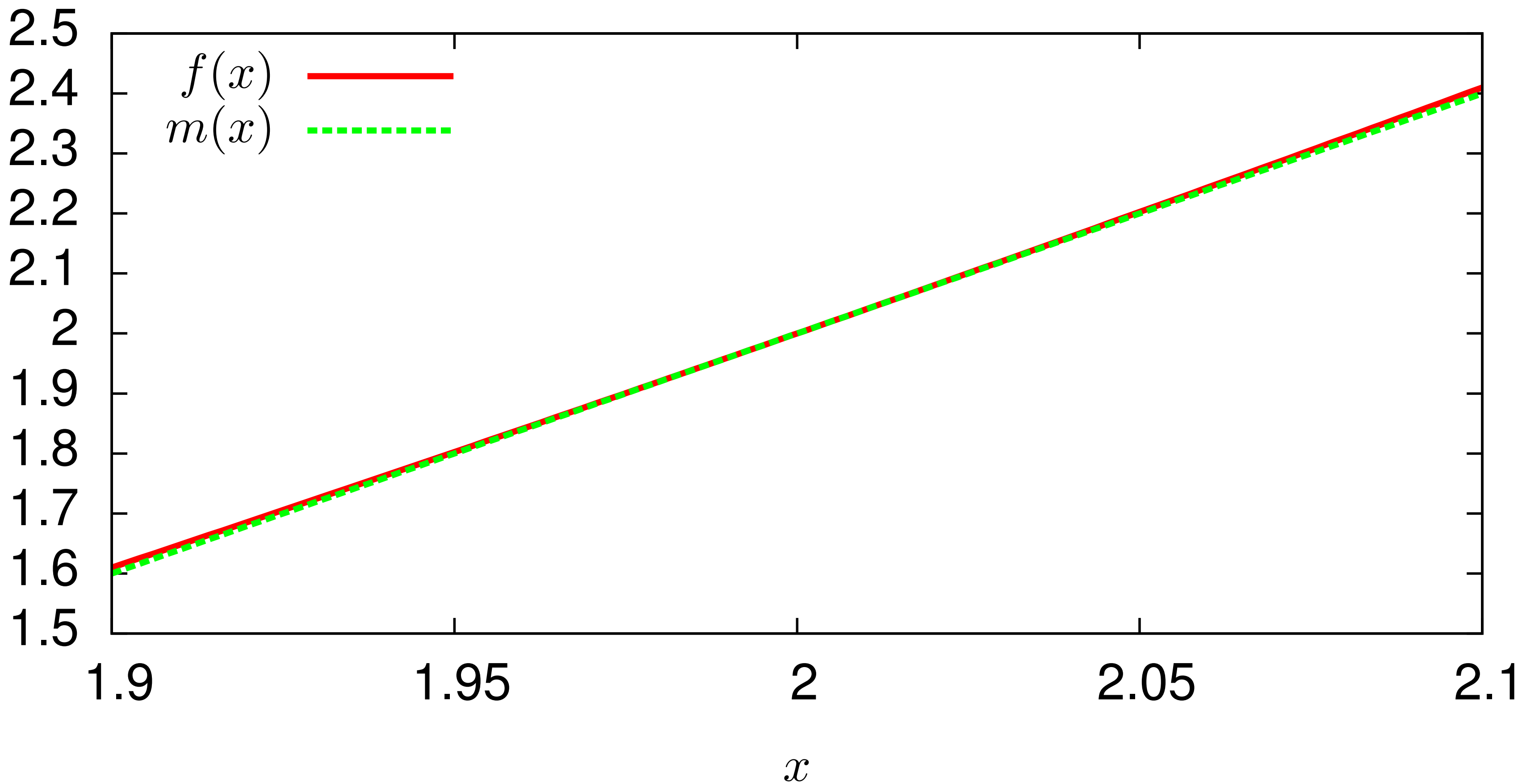
En posant $x = \widehat{x} + d$, nous obtenons

$$m(x) = 2 + 4(x - 2) = 4x - 6.$$

Equation à une inconnue



Equation à une inconnue



Equation à une inconnue

Modèle linéaire d'une fonction à une variable

Soit $F : \mathbb{R} \to \mathbb{R}$ une fonction différentiable.

Le modèle linéaire *de F en $\widehat{x}$ est une fonction $m_{\widehat{x}} : \mathbb{R} \to \mathbb{R}$ définie par*

$$m_{\widehat{x}}(x) = F(\widehat{x}) + (x - \widehat{x})F'(\widehat{x}).$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Equation à une inconnue

Algorithme :

1. Calculer le modèle linéaire en $\widehat{x}$:

$$F(\widehat{x}) + (x - \widehat{x})F'(\widehat{x}) = 0,$$

2. Calculer sa racine x^+

$$x^+ = \widehat{x} - \frac{F(\widehat{x})}{F'(\widehat{x})},$$

3. Si x^+ n'est pas une racine du problème de départ, considérer x^+ comme nouvelle approximation, et recommencer.

Equation à une inconnue

Critère d'arrêt :

- En théorie $F(x^+) = 0$.
- En pratique, arithmétique finie.
- On définit une précision ε, et la condition est

$$|F(x^+)| \leq \varepsilon.$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Rappel : Méthode de Newton — une variable

Objectif

Trouver une approximation de la solution de l'équation

$$F(x) = 0.$$

Input

- La fonction $F : \mathbb{R} \to \mathbb{R}$;
- La dérivée de la fonction $F' : \mathbb{R} \to \mathbb{R}$;
- Une première approximation de la solution $x_0 \in \mathbb{R}$;
- La précision demandée $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Rappel : Méthode de Newton — une variable

Output

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

Initialisation

$k = 0$

Itérations

1. $x_{k+1} = x_k - F(x_k)/F'(x_k)$,
2. $k = k + 1$.

Critère d'arrêt

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

Rappel : Méthode de Newton — n variables

Objectif

Trouver une approximation de la solution du système d'équations

$$F(x) = 0.$$

Input

- La fonction $F : \mathbb{R}^n \to \mathbb{R}^n$;
- La matrice jacobienne de la fonction $J : \mathbb{R}^n \to \mathbb{R}^{n \times n}$;
- Une première approximation de la solution $x_0 \in \mathbb{R}^n$;
- La précision demandée $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$.

Output

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Rappel : Méthode de Newton — n variables

Initialisation

$k = 0$

Itérations

1. Calculer d_{k+1} solution de $J(x_k)d_{k+1} = -F(x_k)$.
2. $x_{k+1} = x_k + d_{k+1}$.
3. $k = k + 1$.

Critère d'arrêt

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

Rappel: Méthode de Newton — n variables

Convergence de la méthode de Newton — n variables *Soit un ensemble convexe ouvert $X \subseteq \mathbb{R}^n$, et une fonction $F : X \to \mathbb{R}^n$. Supposons qu'il existe $x^* \in X$, une boule $B(x^*, r)$ centrée en x^* de rayon r, et une constante $\rho > 0$ tels que $F(x^*) = 0$, $B(x^*, r) \subset X$, $J(x^*)$ est inversible,*

$$\|J(x^*)^{-1}\| \leq \frac{1}{\rho},$$

et J est continue au sens de Lipschitz sur $B(x^*, r)$, la constante de Lipschitz étant M.

(suite...)

Rappel: Méthode de Newton — n variables

Convergence de la méthode de Newton — n variables (suite)

Alors, il existe $\eta > 0$ tel que si

$$x_0 \in B(x^*, \eta),$$

alors la suite $(x_k)_k$ définie par

$$x_{k+1} = x_k - J(x_k)^{-1} F(x_k) \quad k = 0, 1, \ldots$$

est bien définie et converge vers x^*. De plus,

$$\|x_{k+1} - x^*\| \leq \frac{M}{\rho} \|x_k - x^*\|^2.$$

(p. 204)

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Rappel: Méthode de Newton

Performance de la méthode de Newton

- Si la fonction n'est pas trop non-linéaire;
- Si la dérivée de f à la solution n'est pas trop proche de 0;
- Si x_0 n'est pas trop éloigné de la racine;
- Alors la méthode de Newton converge très vite vers la solution.

Algorithme : Newton locale

Objectif

Trouver une approximation de la solution du système

$$\nabla f(x) = 0.$$

Input

- Le gradient de la fonction $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$;
- Le hessien de la fonction $\nabla^2 f : \mathbb{R}^n \to \mathbb{R}^{n \times n}$;
- Une première approximation de la solution $x_0 \in \mathbb{R}^n$;
- La précision demandée $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$.

Algorithme : Newton locale

Output

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

Initialisation

$k = 0$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Algorithme : Newton locale

Itérations

1. Calculer d_{k+1} solution de $\nabla^2 f(x_k) d_{k+1} = -\nabla f(x_k)$,
2. $x_{k+1} = x_k + d_{k+1}$,
3. $k = k + 1$.

Critère d'arrêt

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

Newton locale

Mêmes propriétés que pour les équations

1. convergence q-quadratique dans les conditions favorables
2. divergence possible si le point de départ est trop éloigné de la solution,
3. méthode non définie si $\nabla^2 f(x_k)$ n'est pas inversible.

Inconvénient supplémentaire :

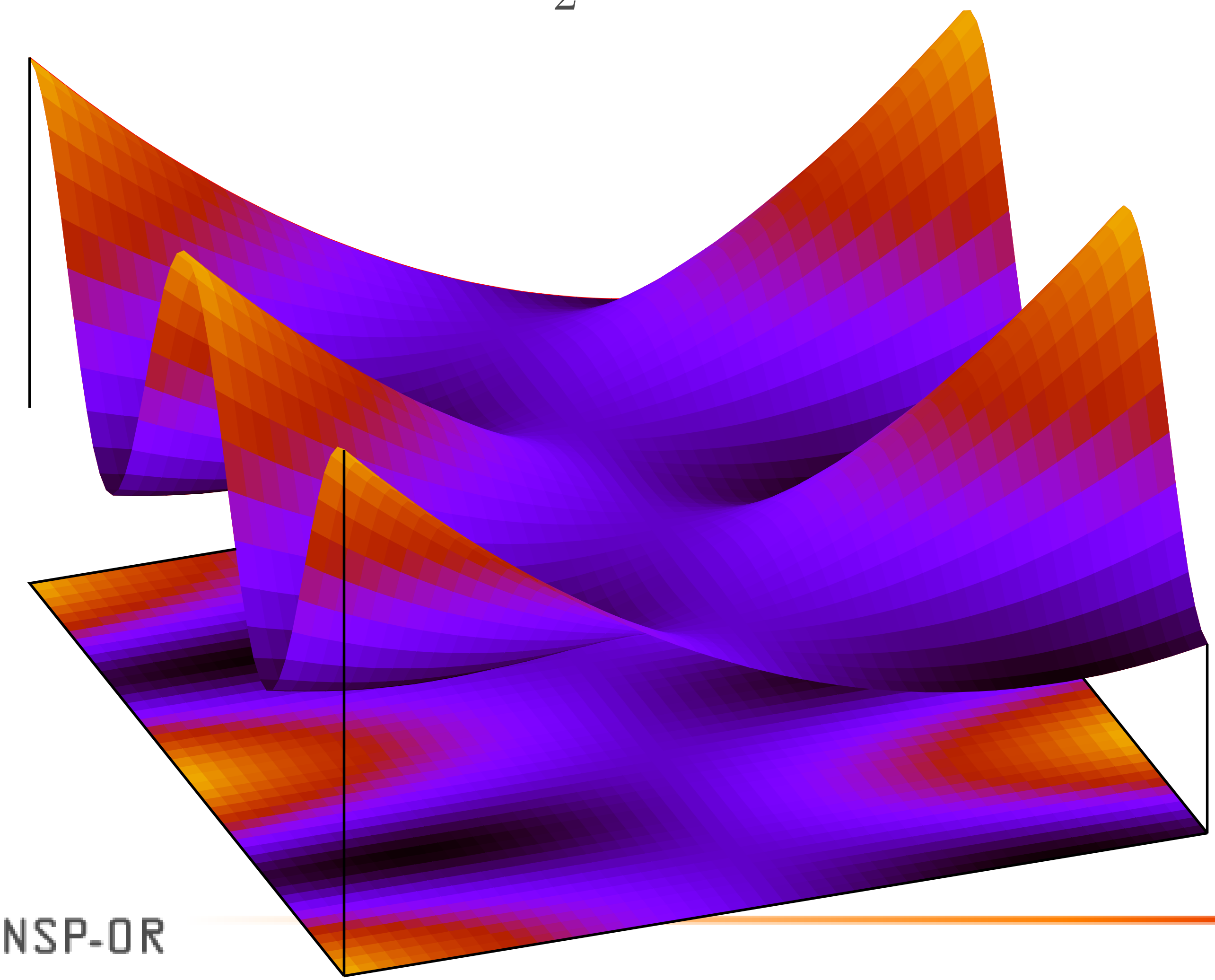
incapacité à distinguer minimum, maximum et point de selle

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

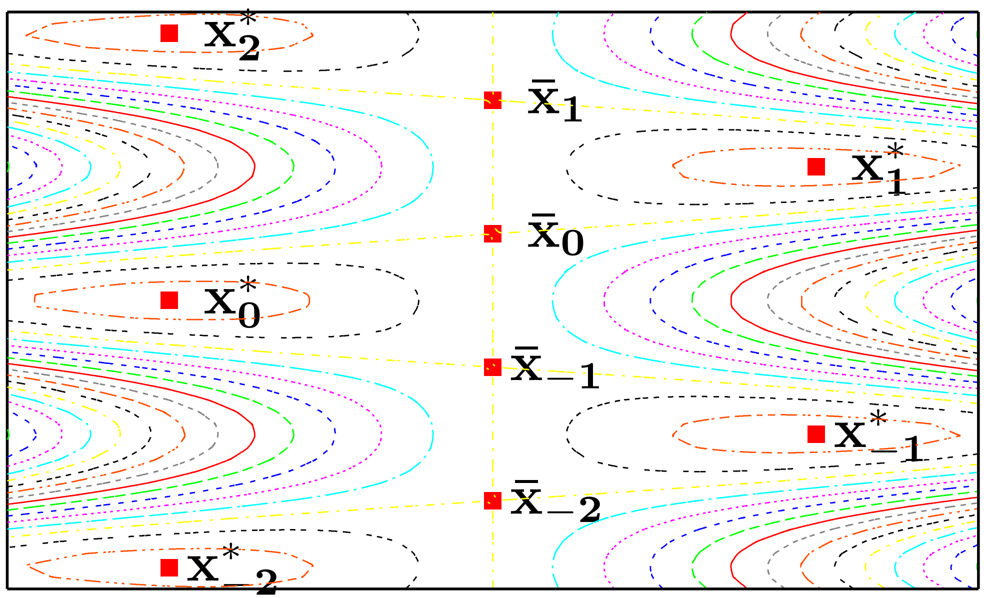
Newton locale

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



Newton locale

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

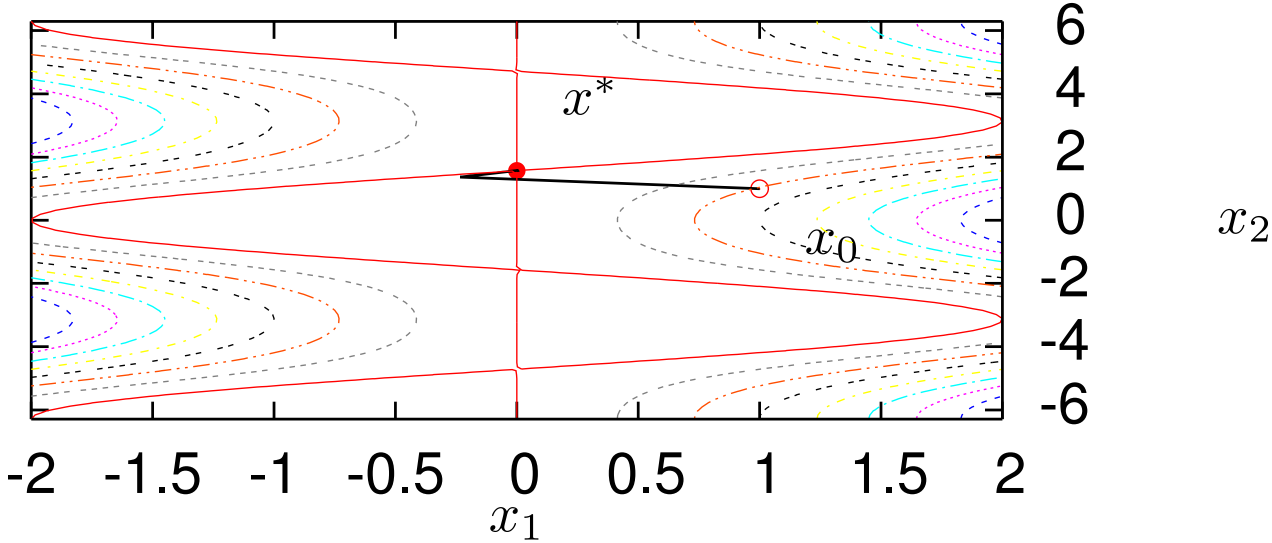


Point de départ $x_0 = (1\ 1)^T$. Convergence rapide.

Newton locale

Solution:

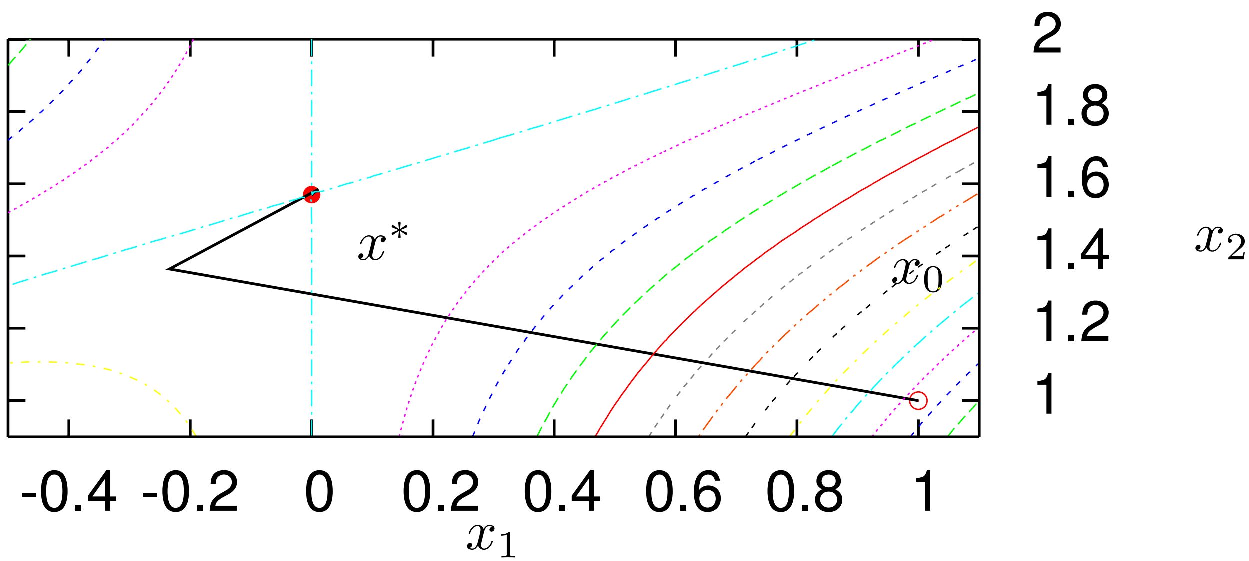
$$x^* = \begin{pmatrix} 0 \\ \frac{\pi}{2} \end{pmatrix} \quad \nabla f(x^*) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \nabla^2 f(x^*) = \begin{pmatrix} 1 & -1 \\ -1 & 0 \end{pmatrix}$$



Newton locale

Solution:

$$x^* = \begin{pmatrix} 0 \\ \frac{\pi}{2} \end{pmatrix} \quad \nabla f(x^*) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \nabla^2 f(x^*) = \begin{pmatrix} 1 & -1 \\ -1 & 0 \end{pmatrix}$$



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Newton locale

- Méthode rapide mais peu fiable
- Interprétation géométrique
  - Equations : modèle linéaire à chaque itération
  - Optimisation : modèle quadratique

Newton locale

Modèle quadratique d'une fonction

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction deux fois différentiable.
Le modèle quadratique *de f en $\widehat{x}$ est une fonction $m_{\widehat{x}} : \mathbb{R}^n \to \mathbb{R}$ définie par*

$$m_{\widehat{x}}(x) = f(\widehat{x}) + (x - \widehat{x})^T \nabla f(\widehat{x}) + \frac{1}{2}(x - \widehat{x})^T \nabla^2 f(\widehat{x})(x - \widehat{x}),$$

où $\nabla f(\widehat{x})$ est le gradient de f en $\widehat{x}$ et $\nabla^2 f(\widehat{x})$ est la matrice hessienne de f en $\widehat{x}$.
En posant $d = x - \widehat{x}$, on obtient la formulation équivalente:

$$m_{\widehat{x}}(\widehat{x} + d) = f(\widehat{x}) + d^T \nabla f(\widehat{x}) + \frac{1}{2} d^T \nabla^2 f(\widehat{x}) d.$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Newton locale

$$\min_x m_{\widehat{x}}(x) = f(\widehat{x}) + (x-\widehat{x})^T \nabla f(\widehat{x}) + \frac{1}{2}(x-\widehat{x})^T \nabla^2 f(\widehat{x})(x-\widehat{x})$$

Condition suffisante d'optimalité (premier ordre)

$$\nabla m_{\widehat{x}}(\widehat{x}+d) = \nabla f(\widehat{x}) + \nabla^2 f(\widehat{x}) d = 0$$

c'est-à-dire

$$d = -\nabla^2 f(\widehat{x})^{-1} \nabla f(\widehat{x}),$$

ou encore

$$x = \widehat{x} - \nabla^2 f(\widehat{x})^{-1} \nabla f(\widehat{x}),$$

Newton locale

Condition suffisante d'optimalité (second ordre)

$$\nabla^2 f(\widehat{x}) \text{ définie positive}$$

Lorsque la matrice hessienne de la fonction est définie positive en x_k, une itération de la méthode de Newton locale revient à minimiser le modèle quadratique de la fonction en x_k, et ainsi définir

$$x_{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^n} m_{x_k}(x).$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Algorithme : Modèle quadratique

Objectif

Trouver une approximation de la solution du système

$$\nabla f(x) = 0. \tag{1}$$

Input

- Le gradient de la fonction $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$;
- Le hessien de la fonction $\nabla^2 f : \mathbb{R}^n \to \mathbb{R}^{n \times n}$;
- Une première approximation de la solution $x_0 \in \mathbb{R}^n$;
- La précision demandée $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$.

Algorithme : Modèle quadratique

Output

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

Initialisation

$k = 0$

Algorithme : Modèle quadratique

Itérations

1. Construire le modèle quadratique

$$m_{\widehat{x}}(\widehat{x}+d) = f(\widehat{x}) + d^T \nabla f(\widehat{x}) + \frac{1}{2} d^T \nabla^2 f(\widehat{x}) d,$$

2. Calculer

$$d_{k+1} = \operatorname{argmin}_d m_{\widehat{x}}(\widehat{x}+d)$$

3. $x_{k+1} = x_k + d_{k+1}$,
4. $k = k+1$.

Critère d'arrêt

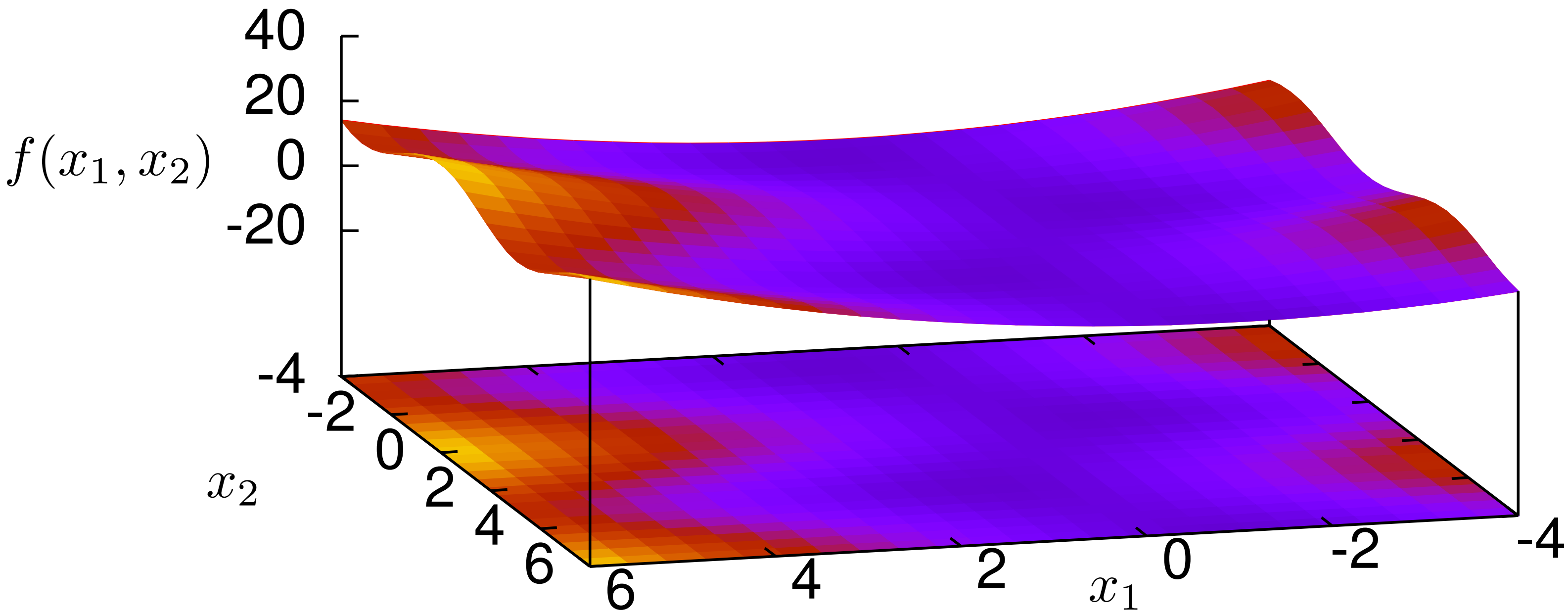
Si $\|\nabla f(x_k)\| \leq \varepsilon$, alors $x^* = x_k$.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

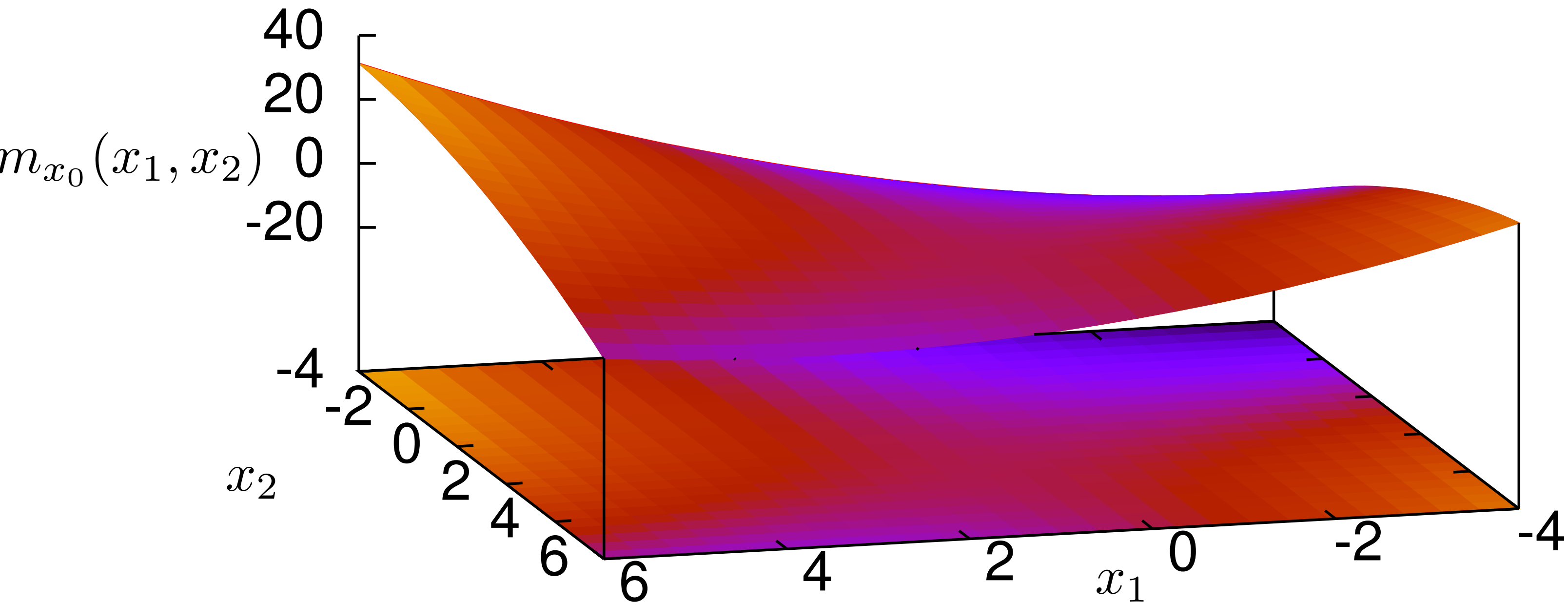
Algorithme : Modèle quadratique

Attention : si $\nabla^2 f(x_k)$ n'est pas définie positive,...



Algorithme : Modèle quadratique

Attention : si $\nabla^2 f(x_k)$ n'est pas définie positive, le modèle n'est pas borné inférieurement



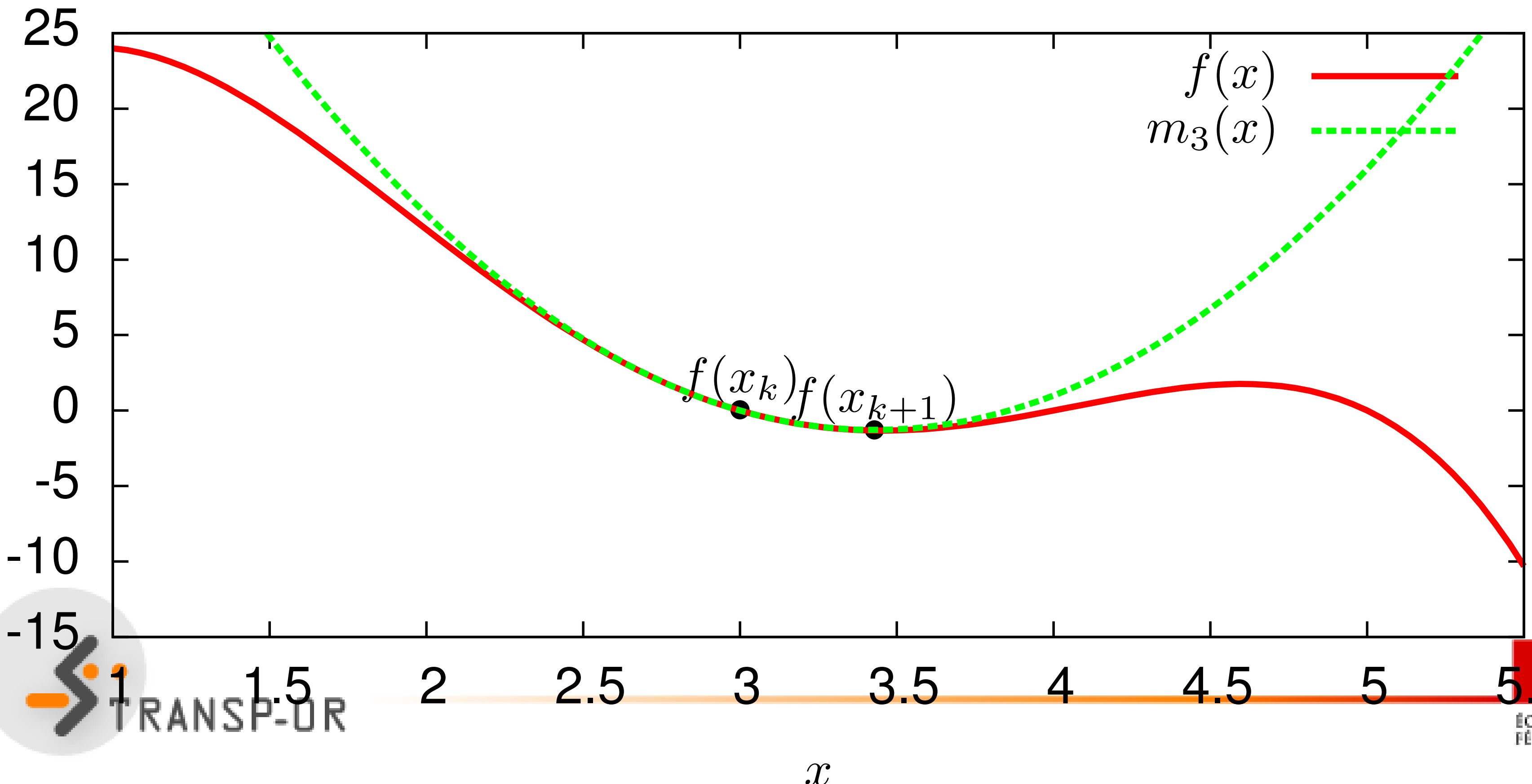
Dans ce cas, l'algorithme ne peut être appliqué.

Modèle quadratique

$$f(x) = -x^4 + 12x^3 - 47x^2 + 60x.$$

$x_k = 3$

$$m_3(x) = 7x^2 - 48x + 81$$



TRANSP-OR

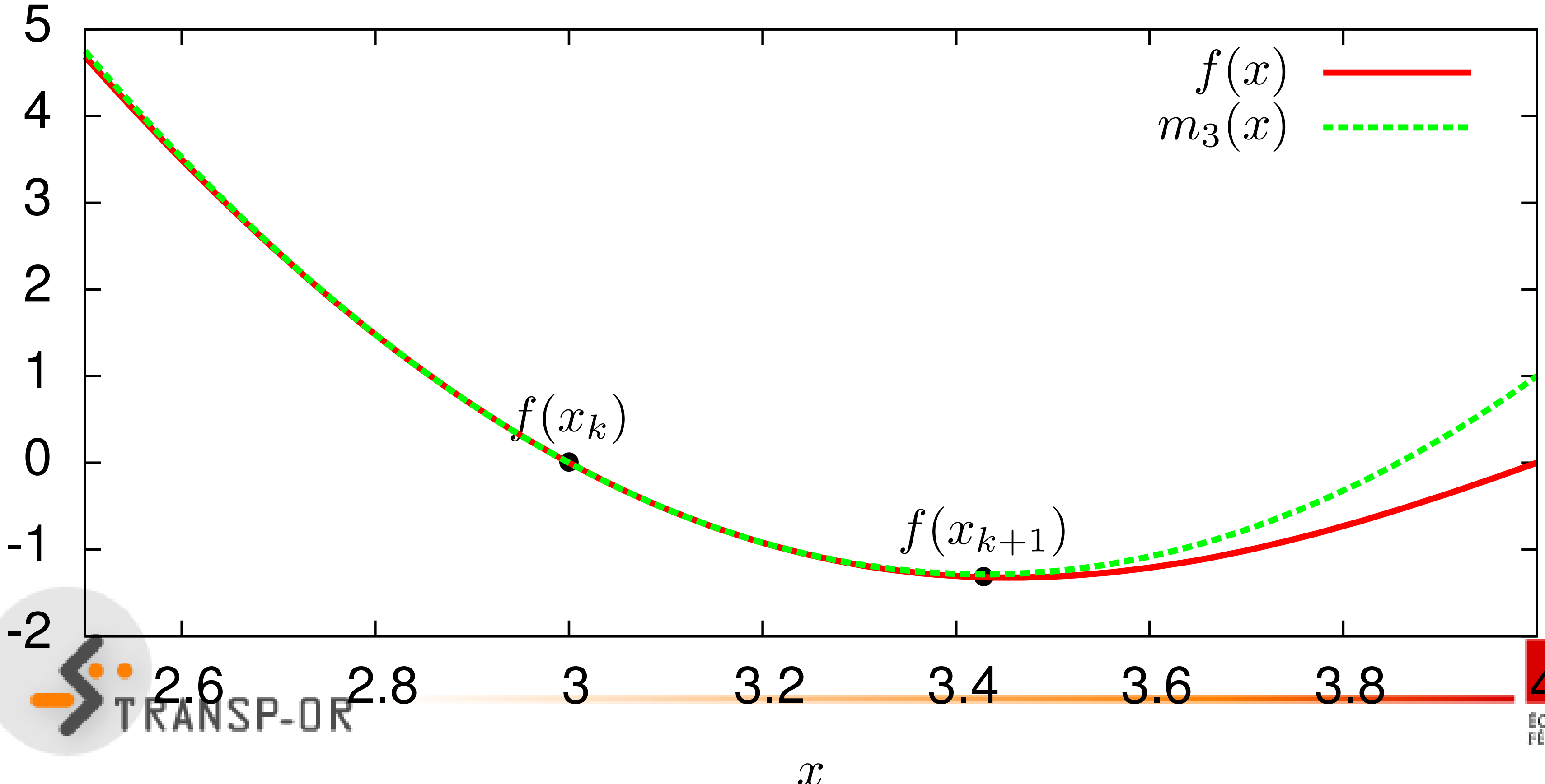
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Modèle quadratique

$$f(x) = -x^4 + 12x^3 - 47x^2 + 60x.$$

$x_k = 3$

$$m_3(x) = 7x^2 - 48x + 81$$

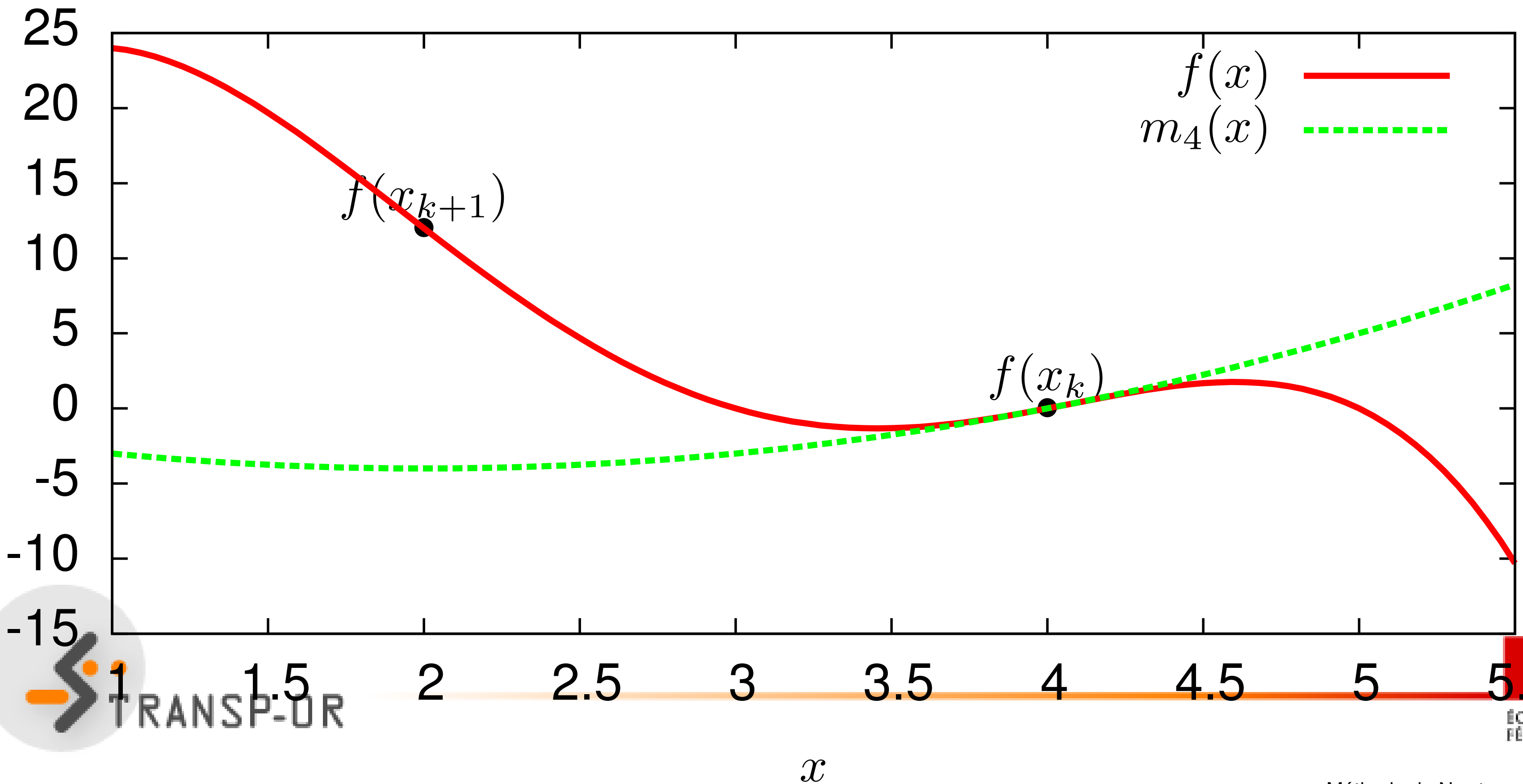


Modèle quadratique

$$f(x) = -x^4 + 12x^3 - 47x^2 + 60x.$$

$x_k = 4$

$$m_4(x) = x^2 - 4x$$

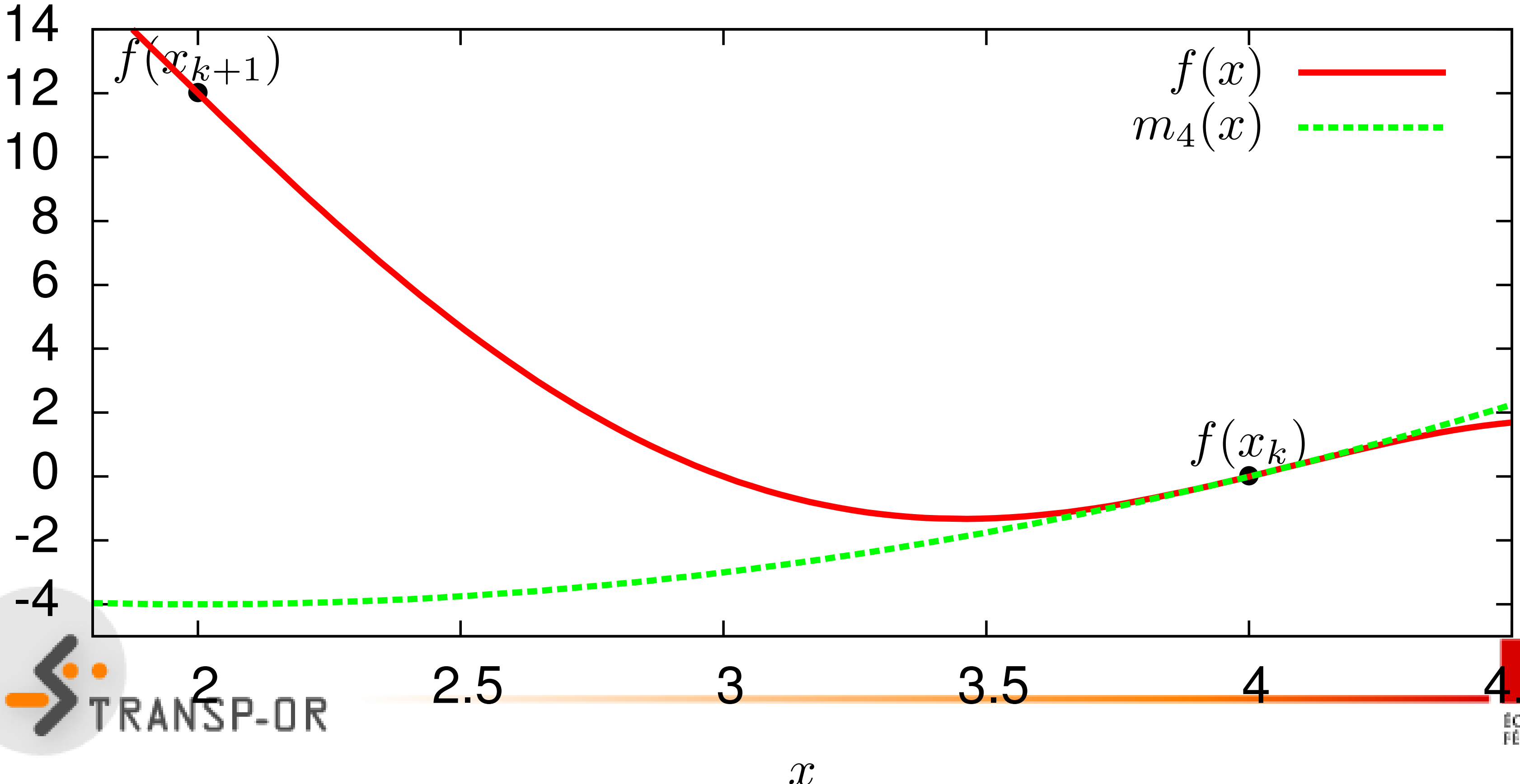


Modèle quadratique

$$f(x) = -x^4 + 12x^3 - 47x^2 + 60x.$$

$x_k = 4$

$$m_4(x) = x^2 - 4x$$

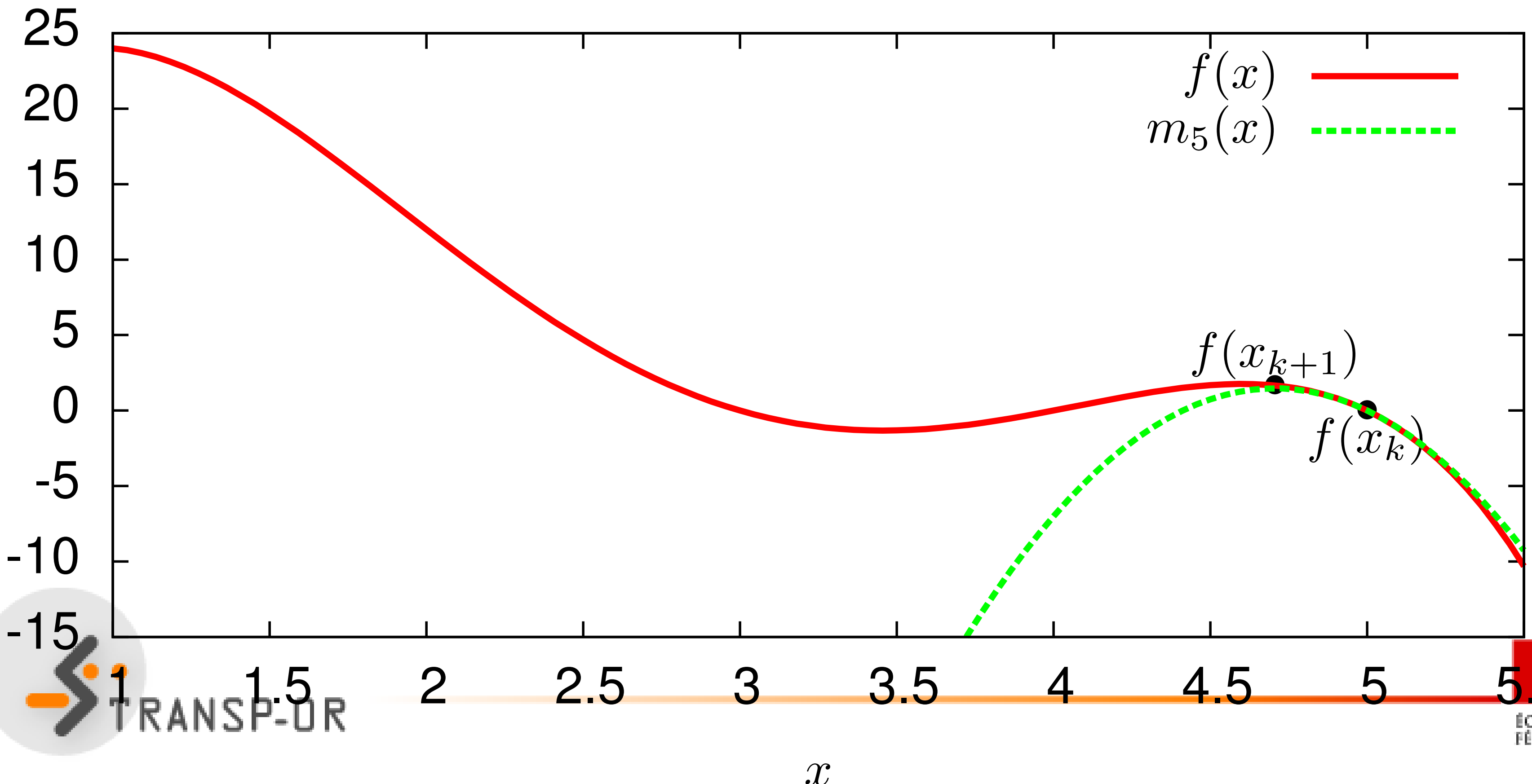


Modèle quadratique

$$f(x) = -x^4 + 12x^3 - 47x^2 + 60x.$$

$x_k = 5$

$$m_5(x) = -17x^2 + 160x - 375.$$

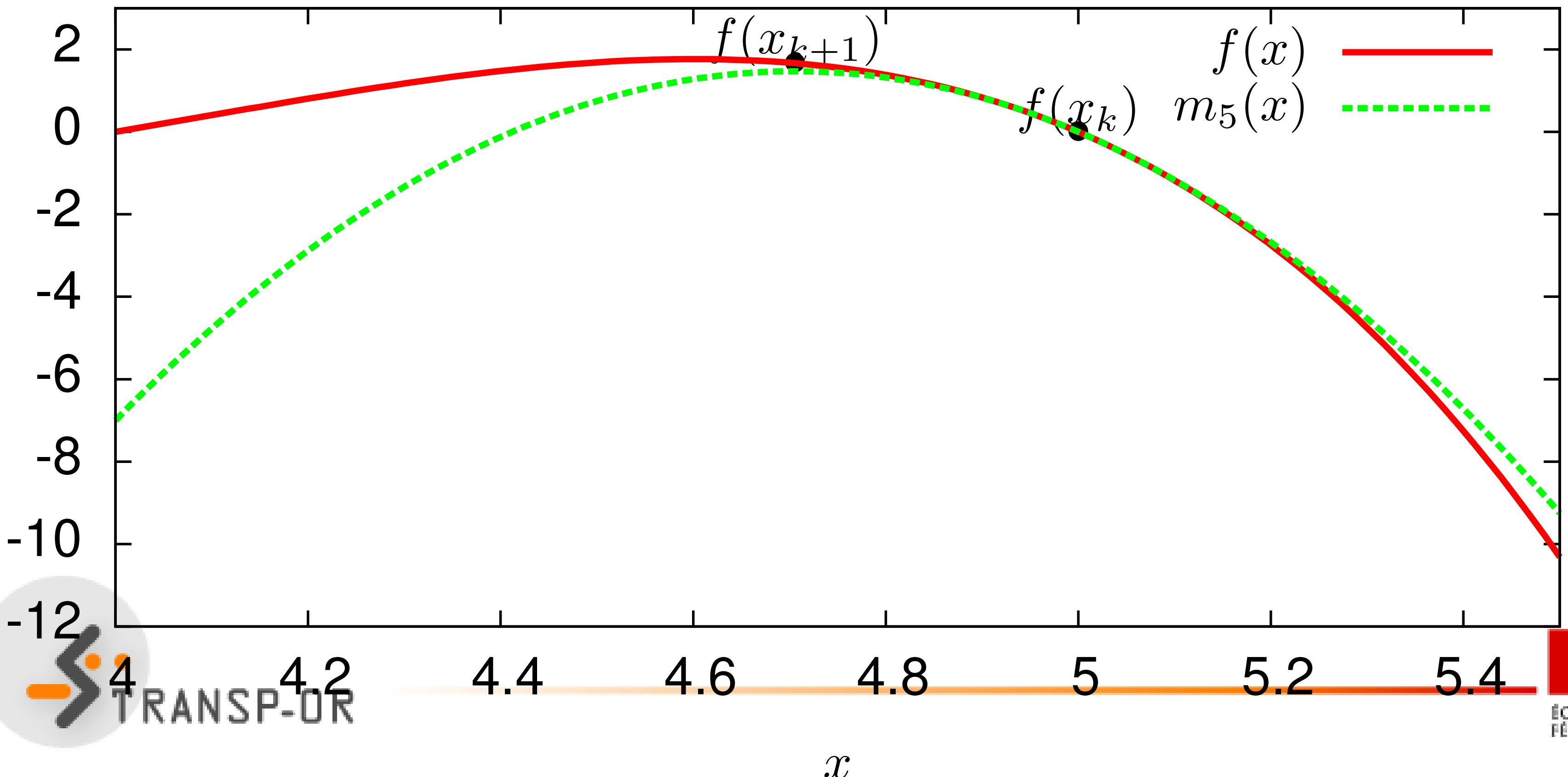


Modèle quadratique

$$f(x) = -x^4 + 12x^3 - 47x^2 + 60x.$$

$x_k = 5$

$$m_5(x) = -17x^2 + 160x - 375.$$



Modèle quadratique

Point de Newton

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction deux fois différentiable, et soit $x_k \in \mathbb{R}^n$. Le point de Newton *de f en x_k est le point*

$$x_N = x_k + d_N$$

où d_N est solution du système d'équations

$$\nabla^2 f(x_k) d_N = -\nabla f(x_k).$$

Ce système est souvent appelé équations de Newton*.*

Modèle quadratique

Point de Cauchy

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction deux fois différentiable, et soit $x_k \in \mathbb{R}^n$. Le point de Cauchy *de f en x_k est le point x_C qui minimise le modèle quadratique de f dans la direction de la plus forte descente, c'est-à-dire*

$$x_C = x_k - \alpha_C \nabla f(x_k)$$

où

$$\alpha_C = \operatorname{argmin}_{\alpha \in \mathbb{R}^+_0} m_{x_k}(x_k - \alpha \nabla f(x_k))$$

ou encore

$$\alpha_C = \frac{\nabla f(x_k)^T \nabla f(x_k)}{\nabla f(x_k)^T \nabla^2 f(x_k) \nabla f(x_k)}.$$