Méthode de Newton locale pour l'optimisation

Michel Bierlaire

michel.bierlaire@epfl.ch

EPFL - Laboratoire Transport et Mobilité - ENAC

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
- Rappel : pour résoudre $F(x) = 0$, $F : \mathbb{R}^n \to \mathbb{R}^n$,
  1. Résoudre $\nabla F(x_k) d_{k+1} = -F(x_k)$
  2. Définir $x_{k+1} = x_k + d_{k+1}$.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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

$k = 0$

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

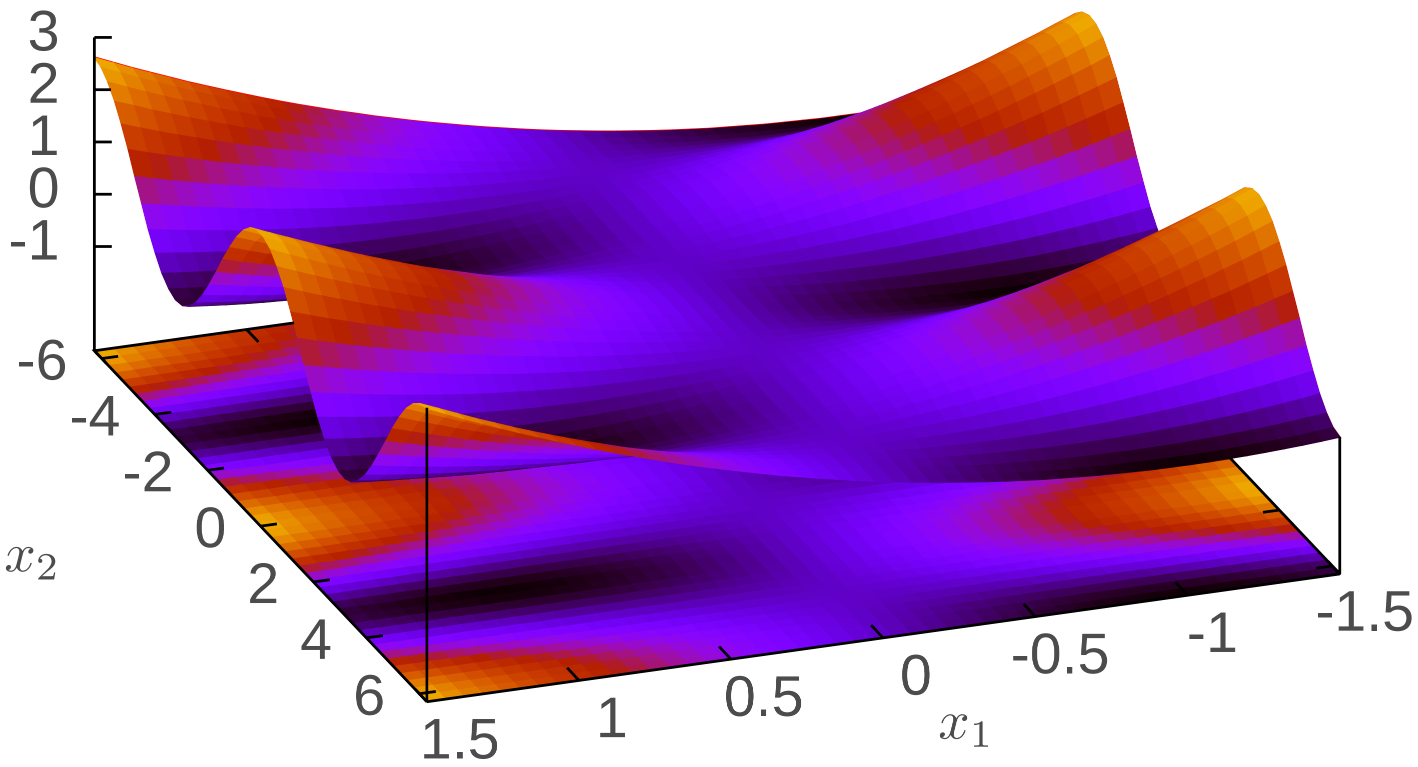
1. convergence q-quadratique dans les conditions favorables
2. divergence possible si le point de départ est trop éloigné de la solution,
3. méthode non définie si $\nabla^2 f(x_k)$ n'est pas inversible.

Inconvénient supplémentaire :

incapacité à distinguer minimum, maximum et point de selle

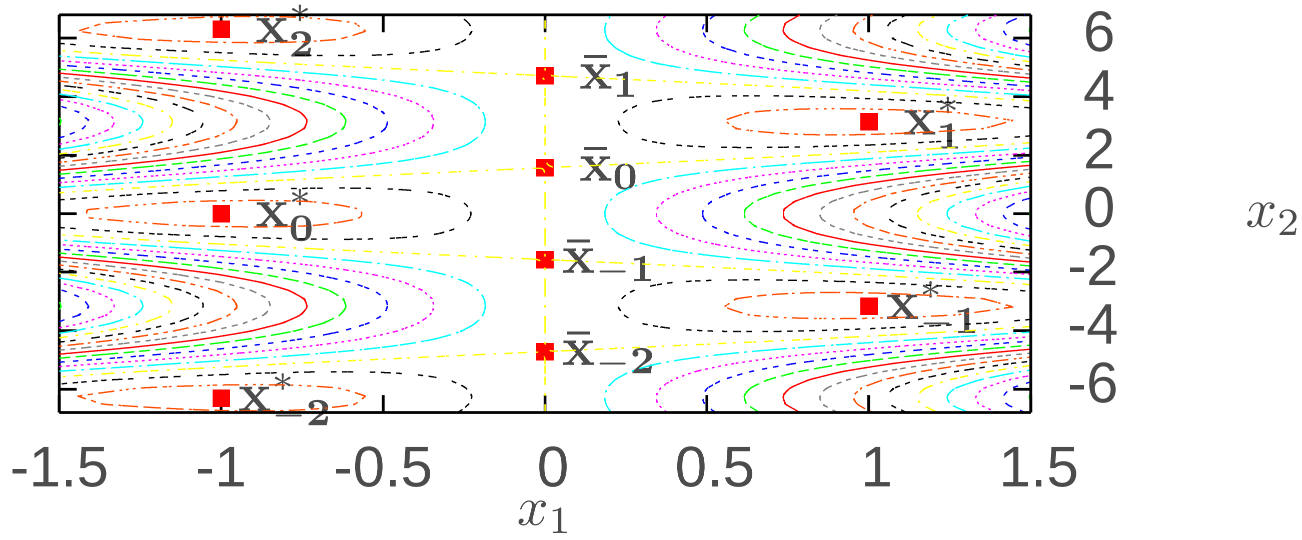
Newton locale

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



Newton locale

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

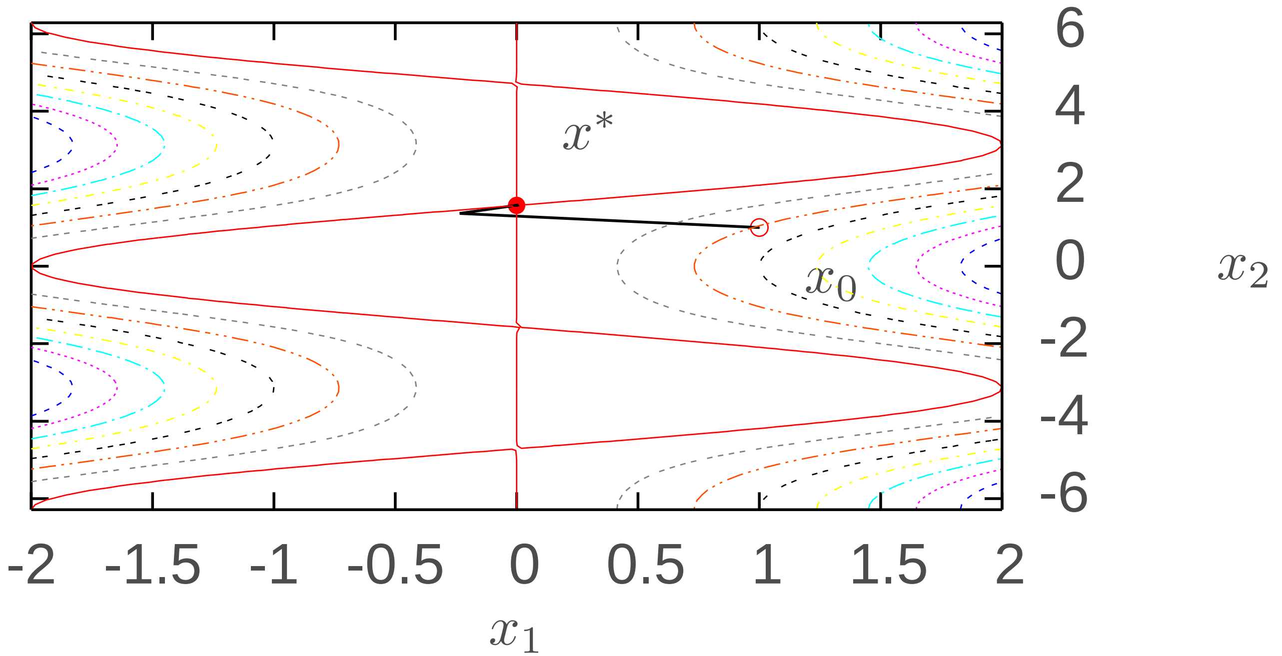


Point de départ $x_0 = (1\ 1)^T$. Convergence rapide.

Newton locale

Solution:

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



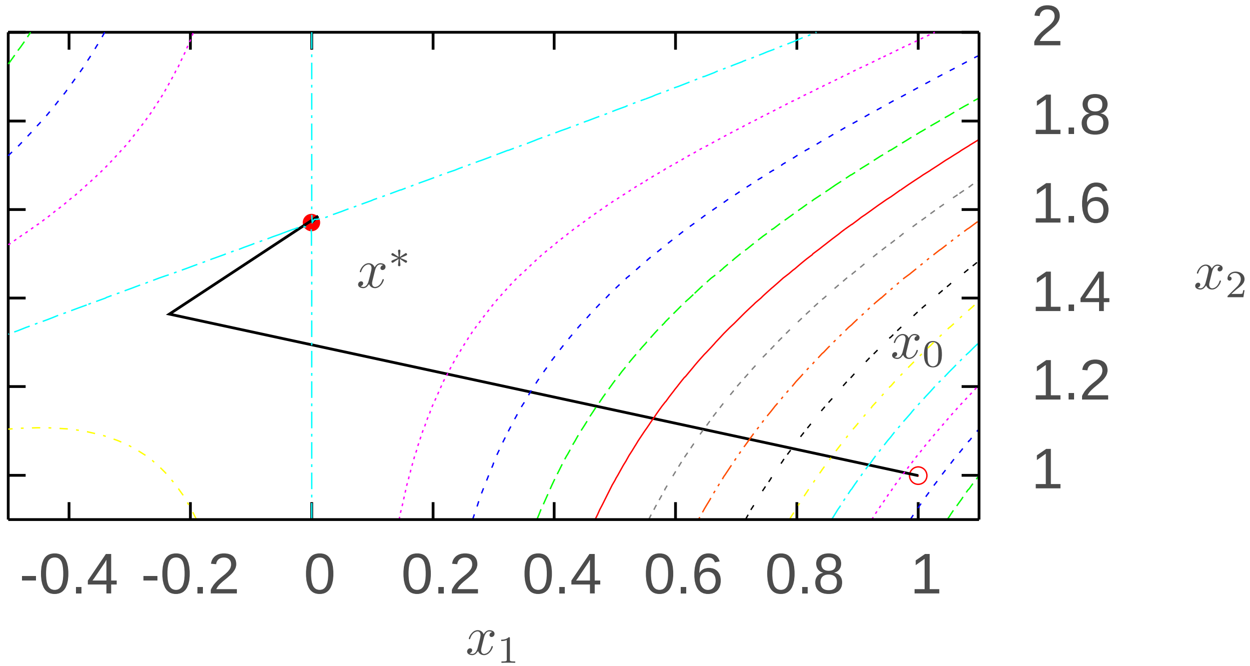
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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

- 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

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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

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

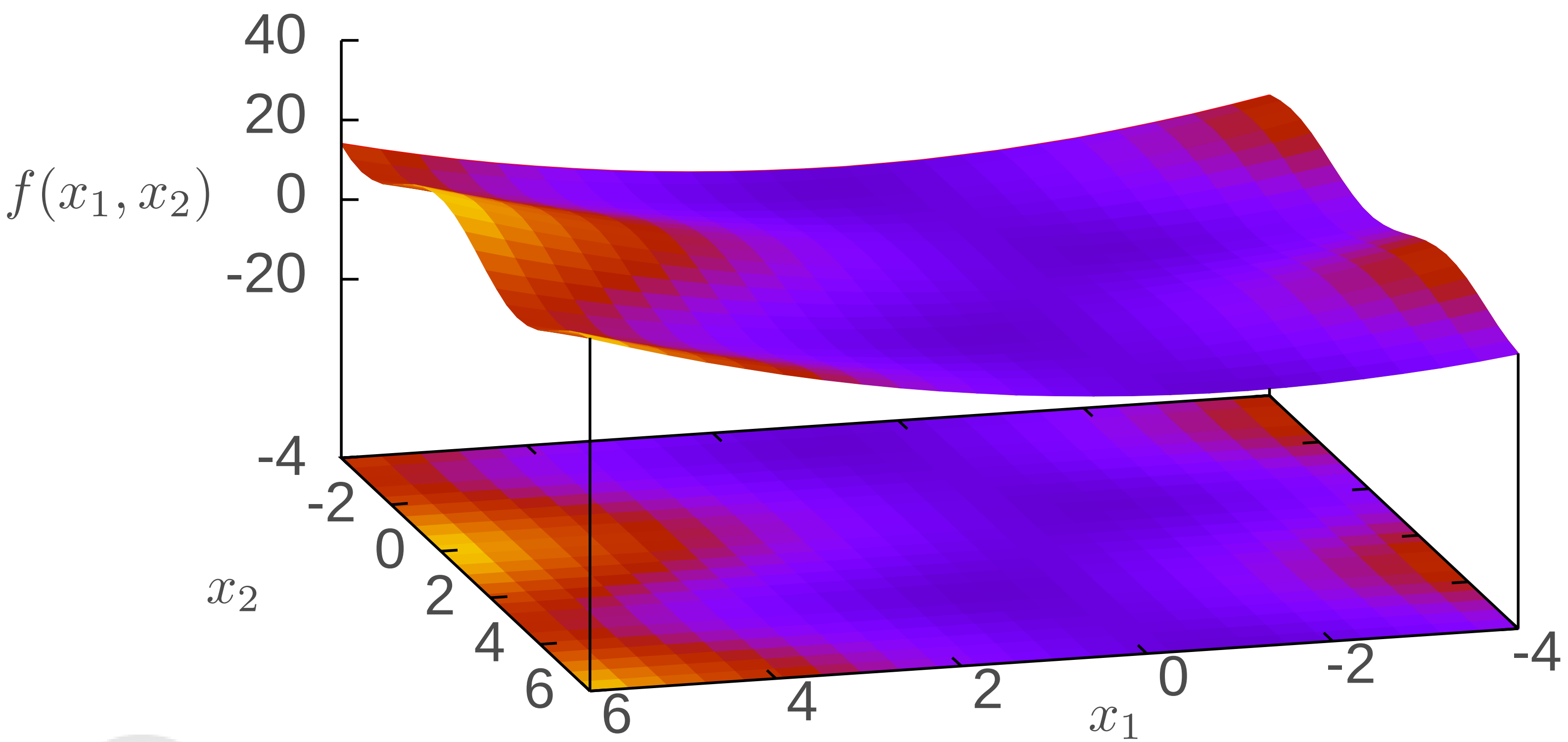
3. $x_{k+1} = x_k + d_{k+1}$,
4. $k = k + 1$.

Critère d'arrêt

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

Algorithme : Modèle quadratique

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

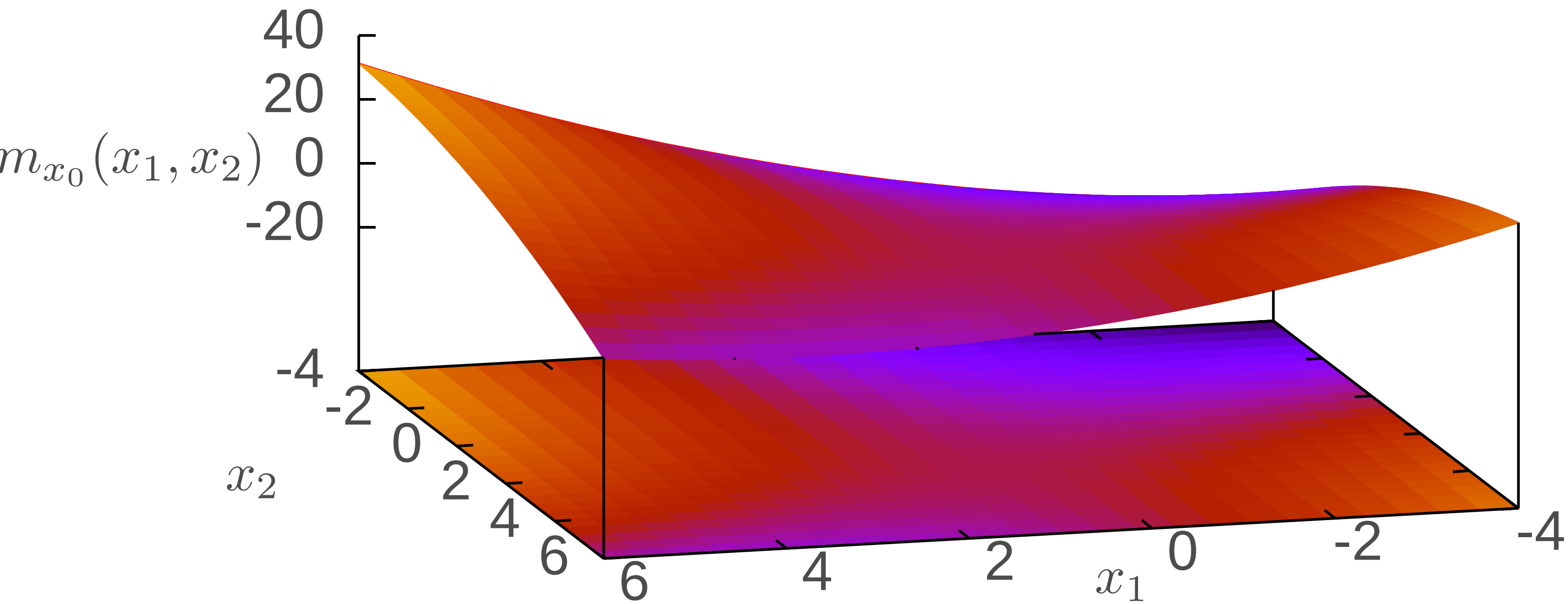


TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Algorithme : Modèle quadratique

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



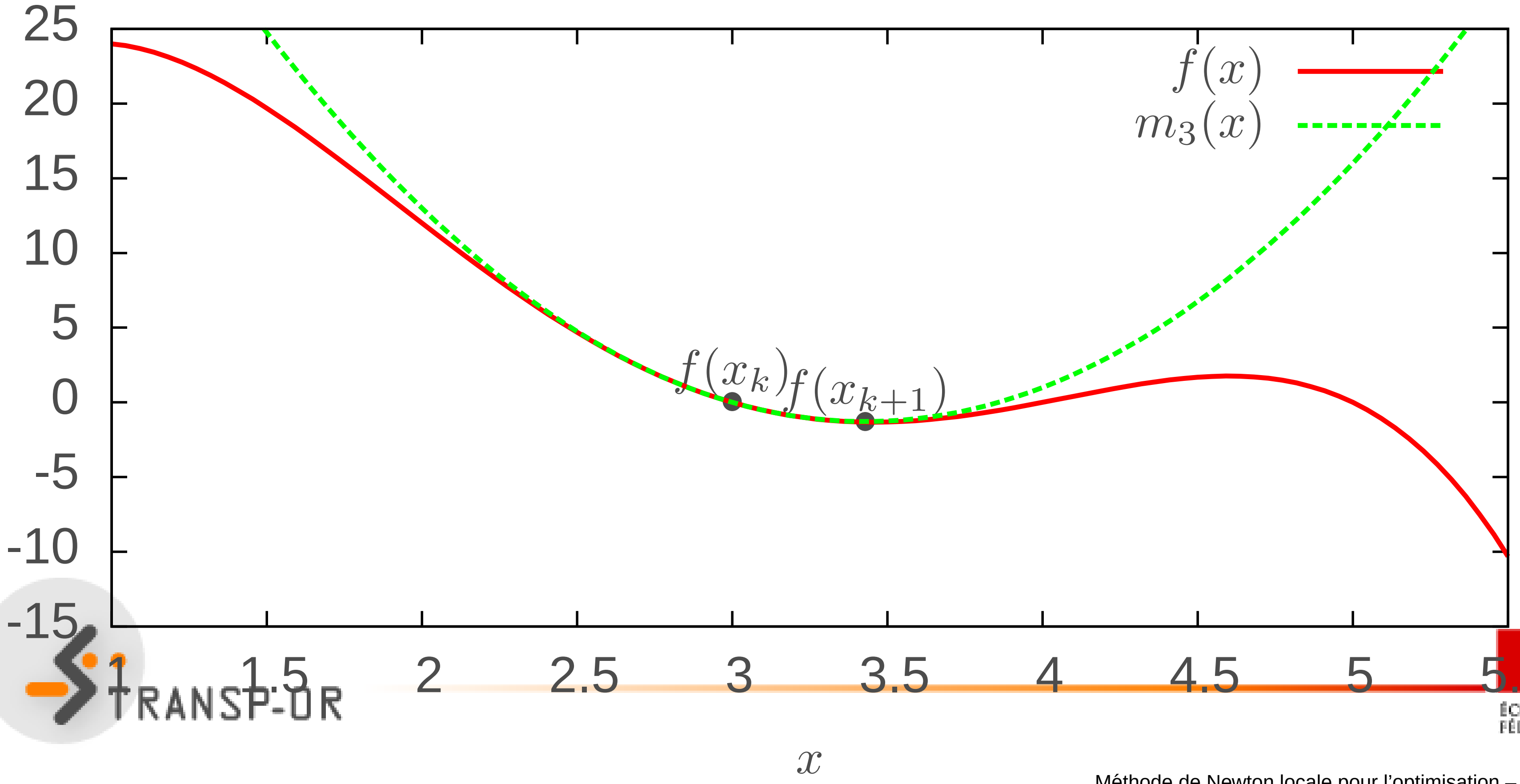
Dans ce cas, l'algorithme ne peut être appliqué.

Modèle quadratique

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

$x_k = 3$

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

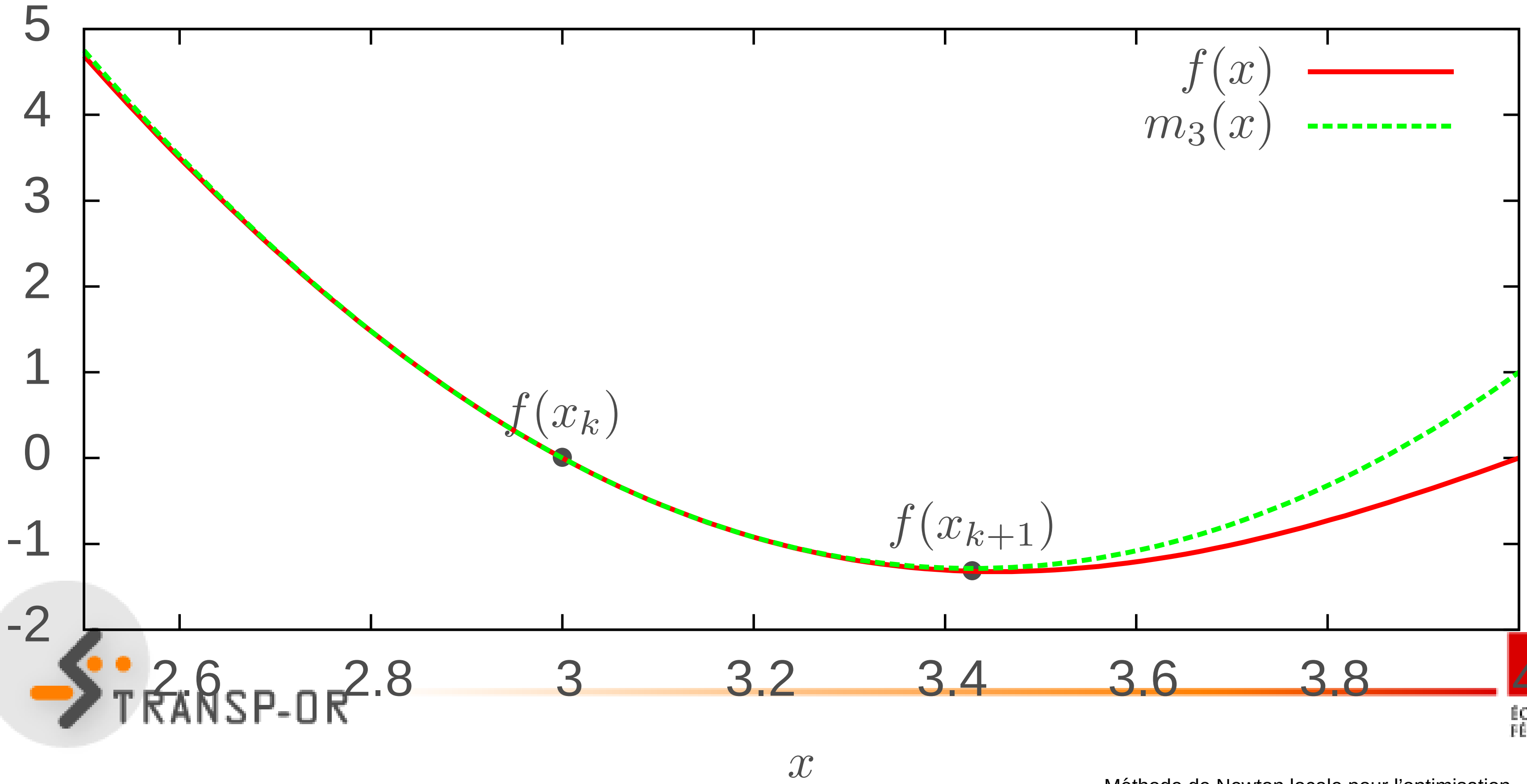


Modèle quadratique

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

$x_k = 3$

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

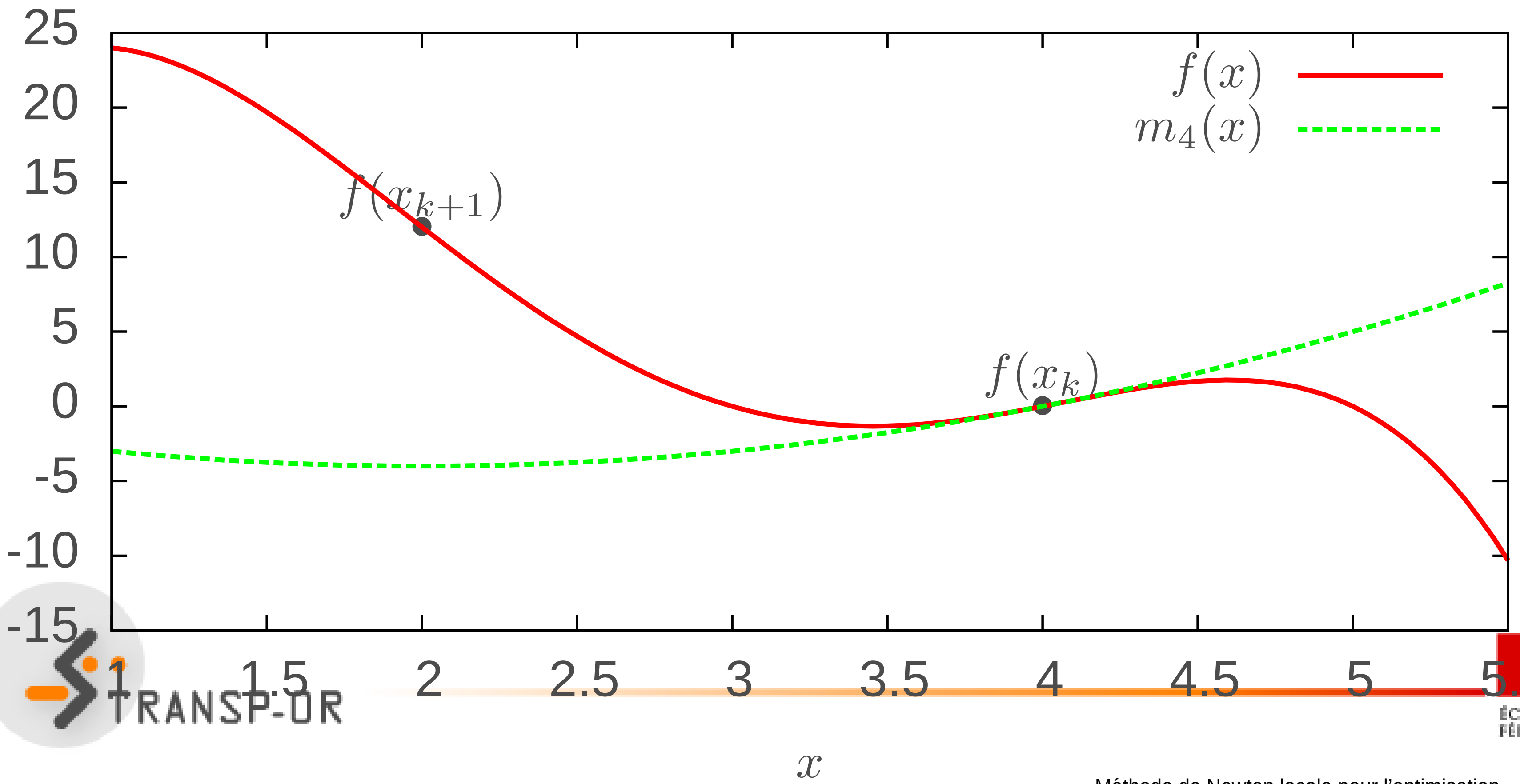


Modèle quadratique

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

$x_k = 4$

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

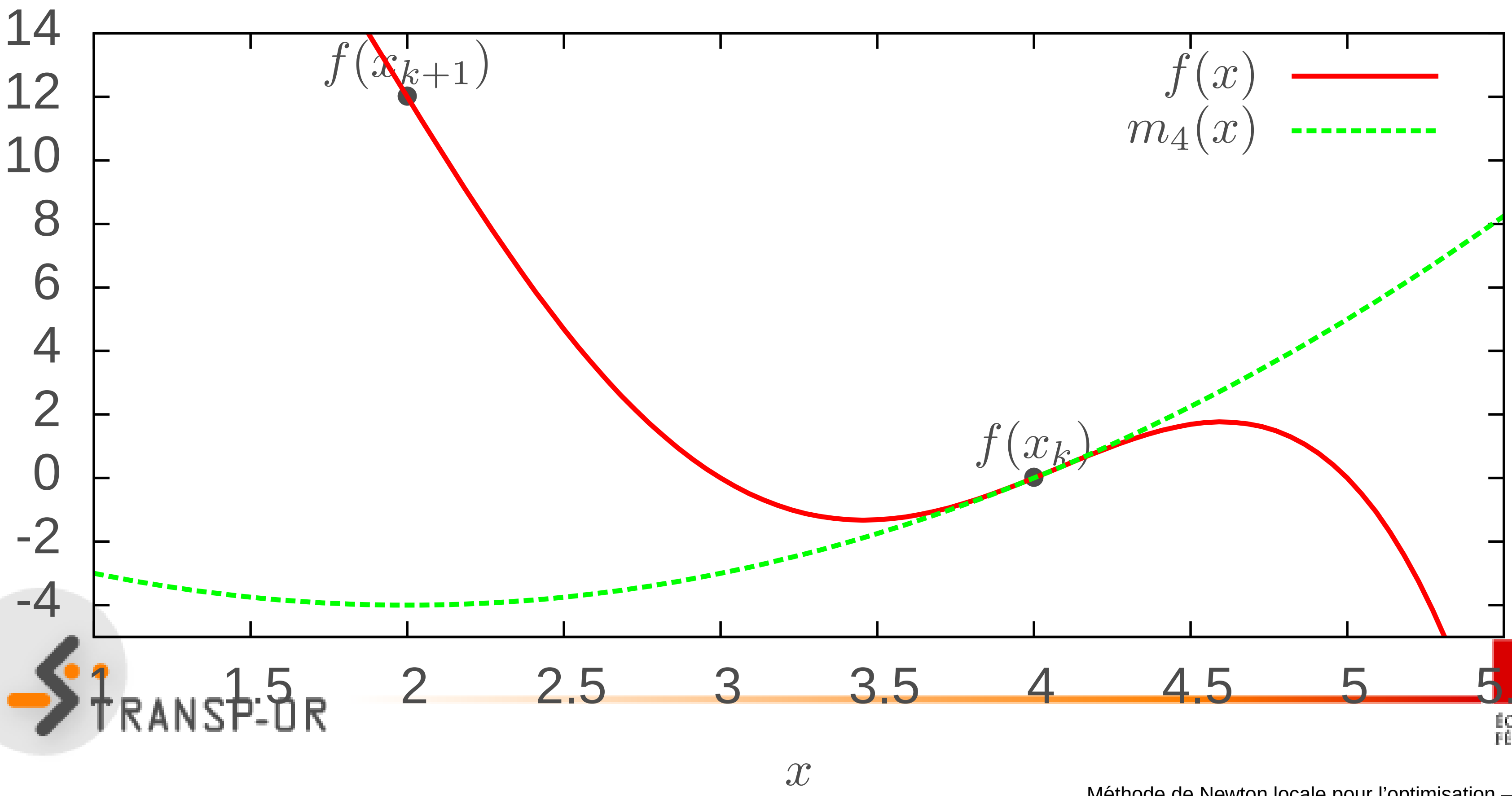


Modèle quadratique

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

$x_k = 4$

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

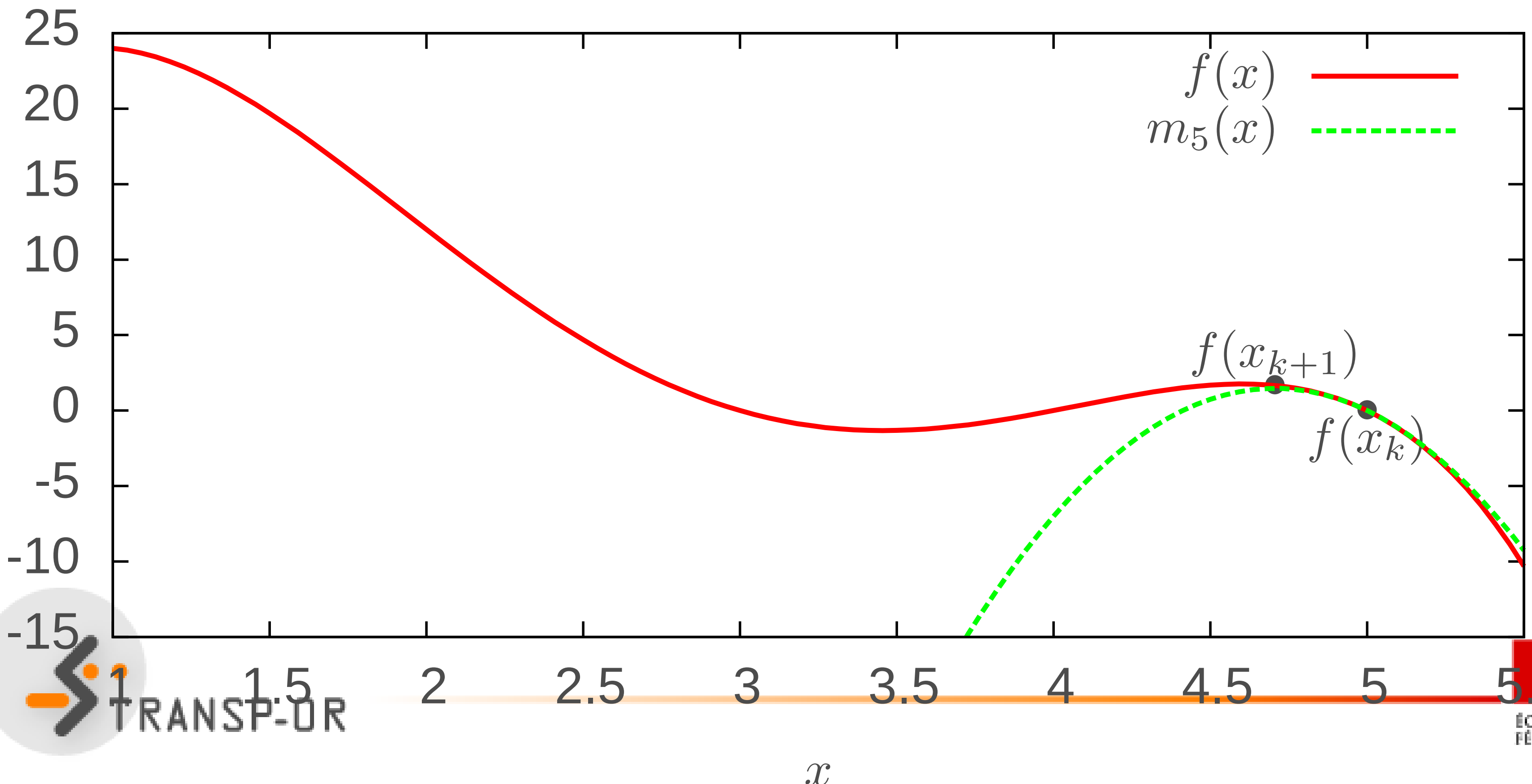


Modèle quadratique

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

$x_k = 5$

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

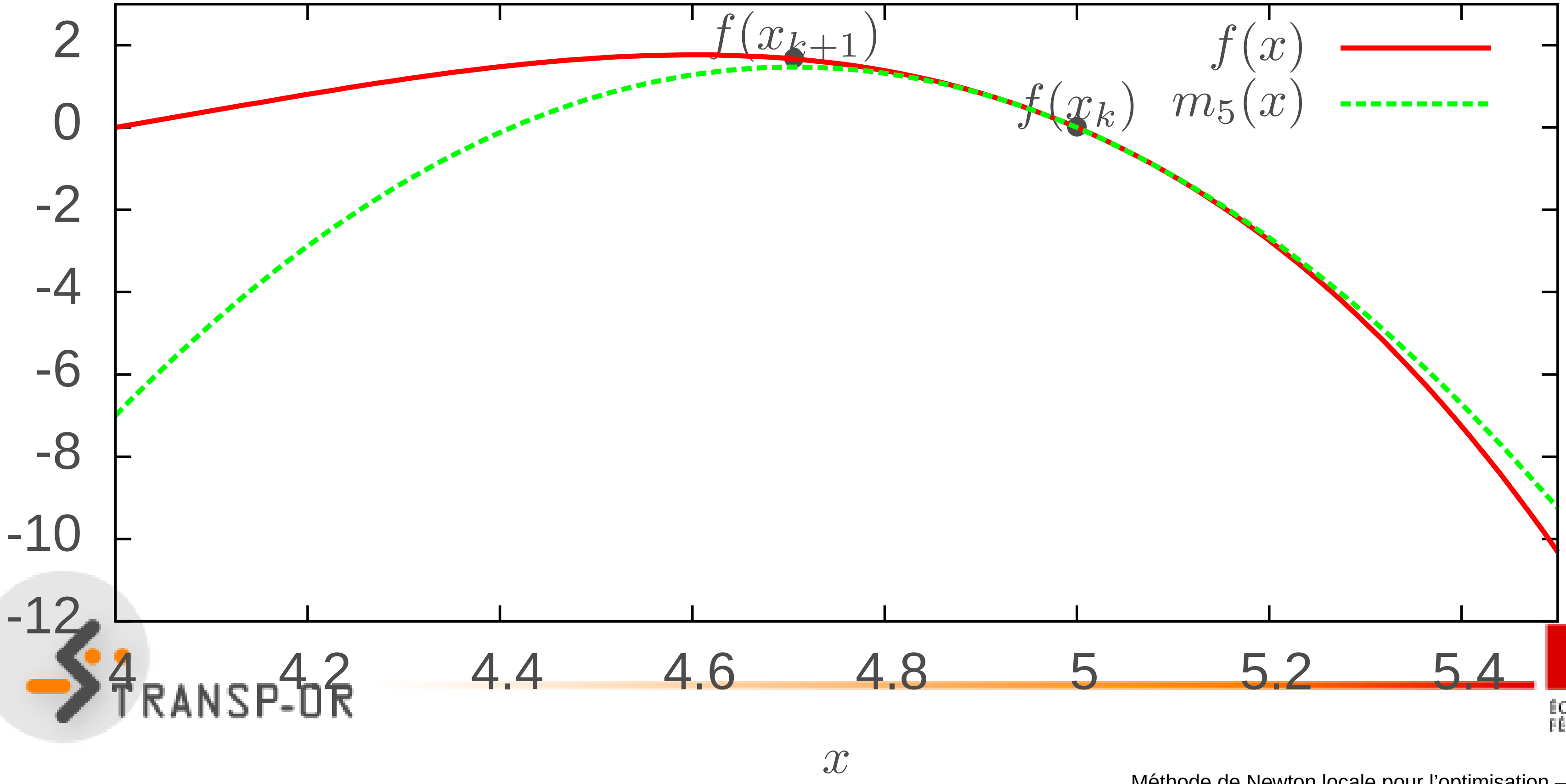


Modèle quadratique

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

$x_k = 5$

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



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*.*

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Résumé

- Conditions nécessaires d'optimalité = système d'équations.
- Méthode de Newton locale.
- Rapide ... sous conditions.
- Pas de distinctions entre minima, maxima et point de selle.