Région de confiance

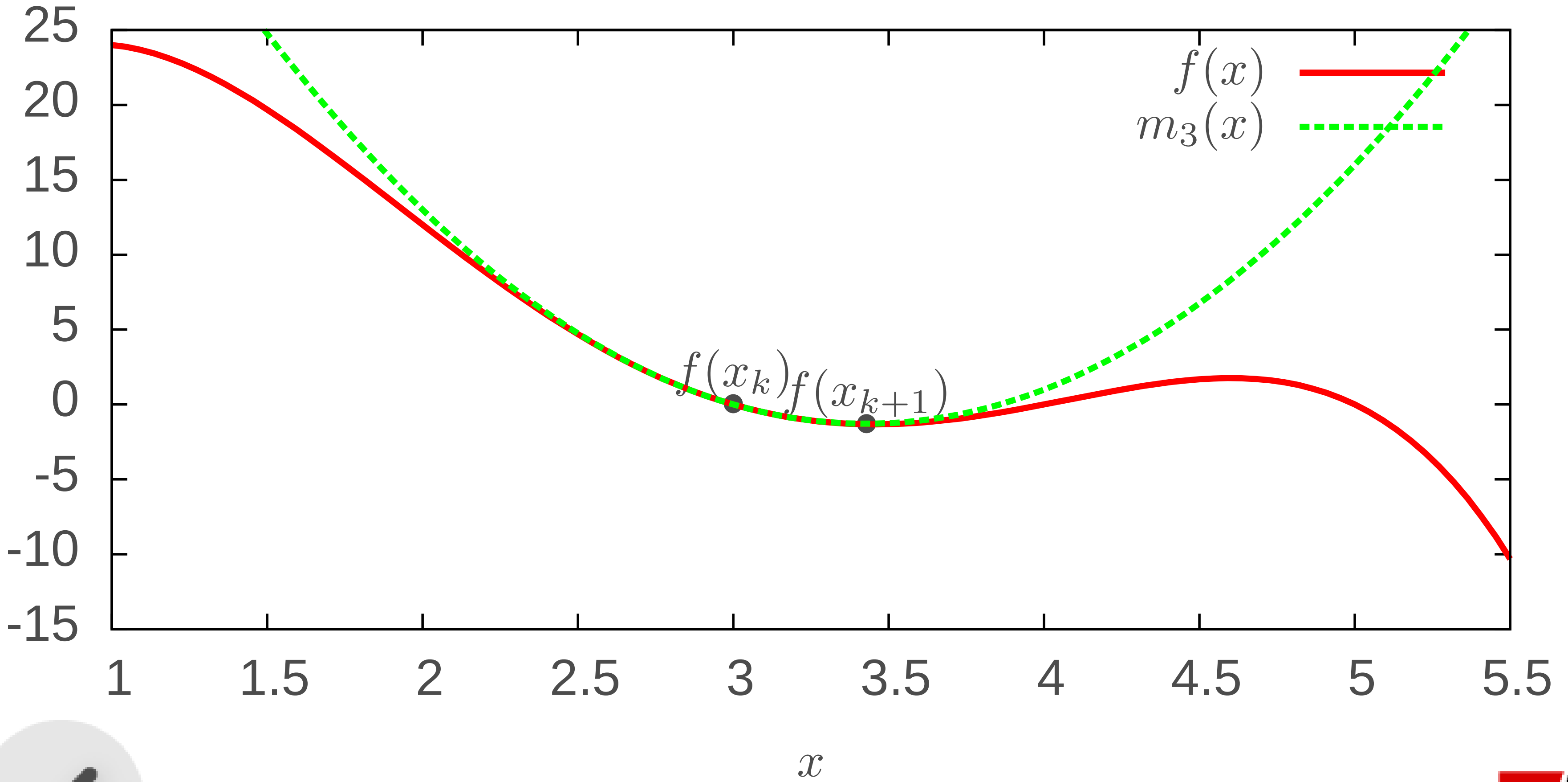
- Méthode de Newton pure, modifiée pour qu'elle fonctionne tout le temps
- Première manière : recherche linéaire
- Seconde manière : région de confiance

Modèle quadratique

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

$x_k = 3$

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

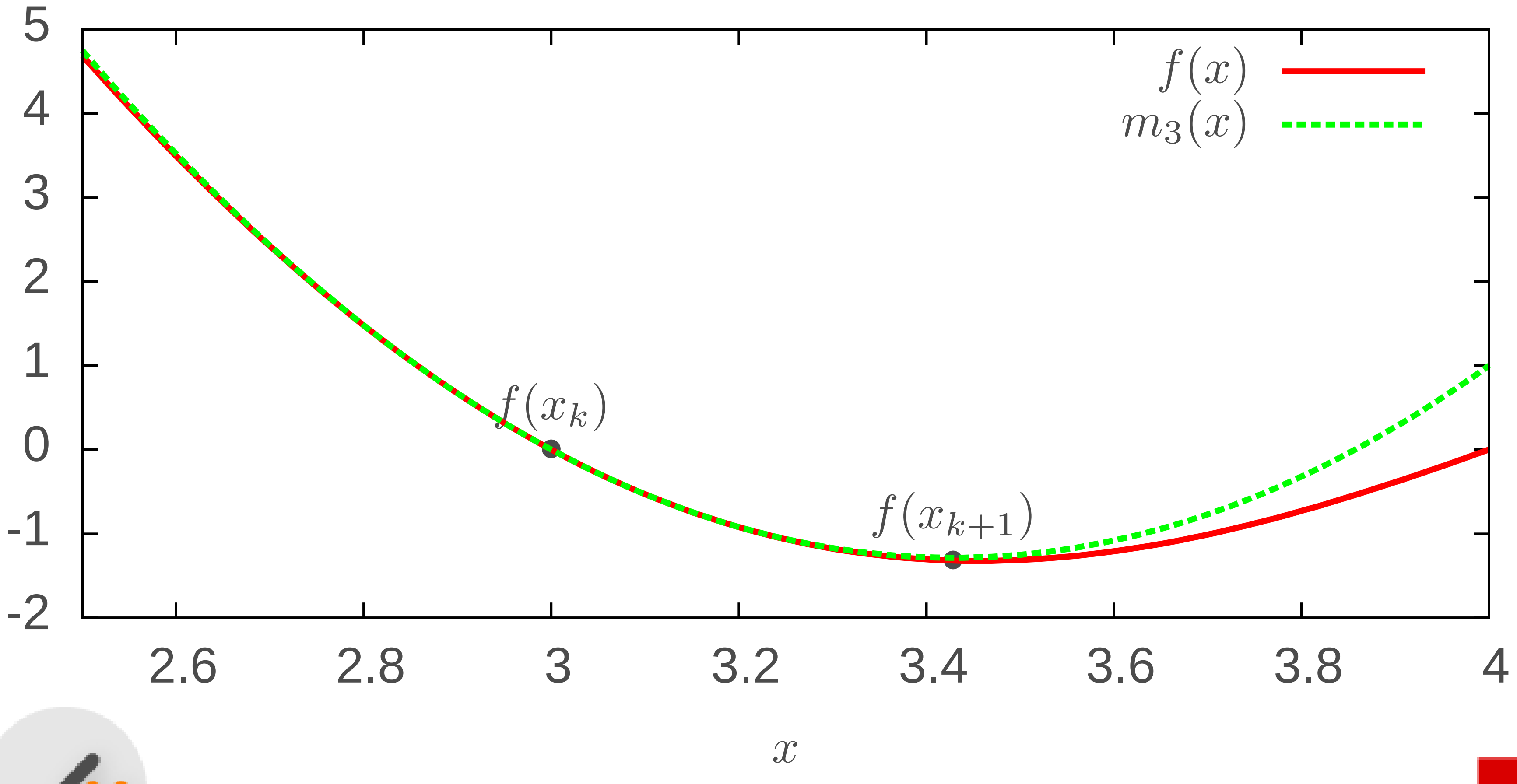


Modèle quadratique

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

$x_k = 3$

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

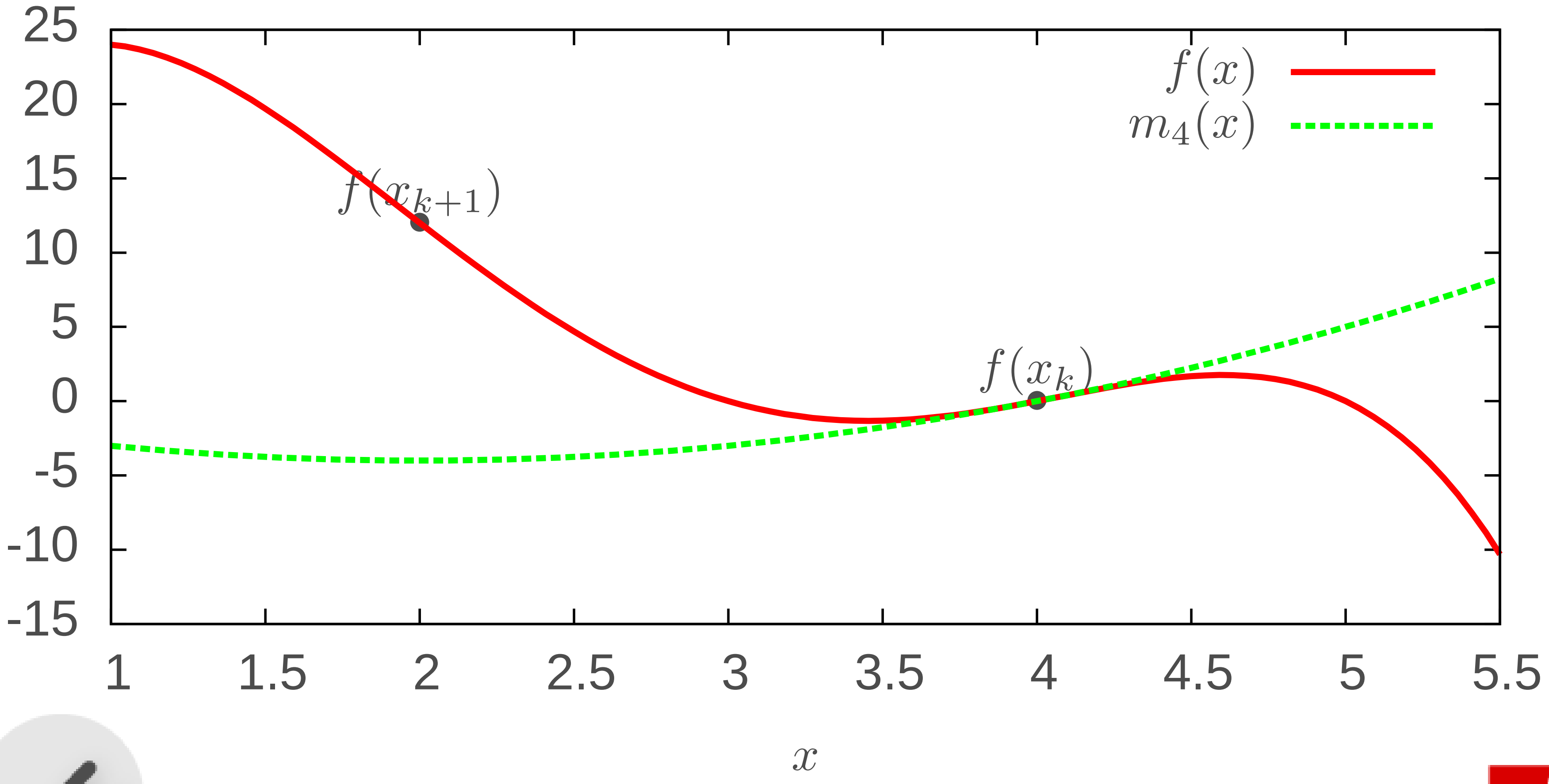


Modèle quadratique

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

$x_k = 4$

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

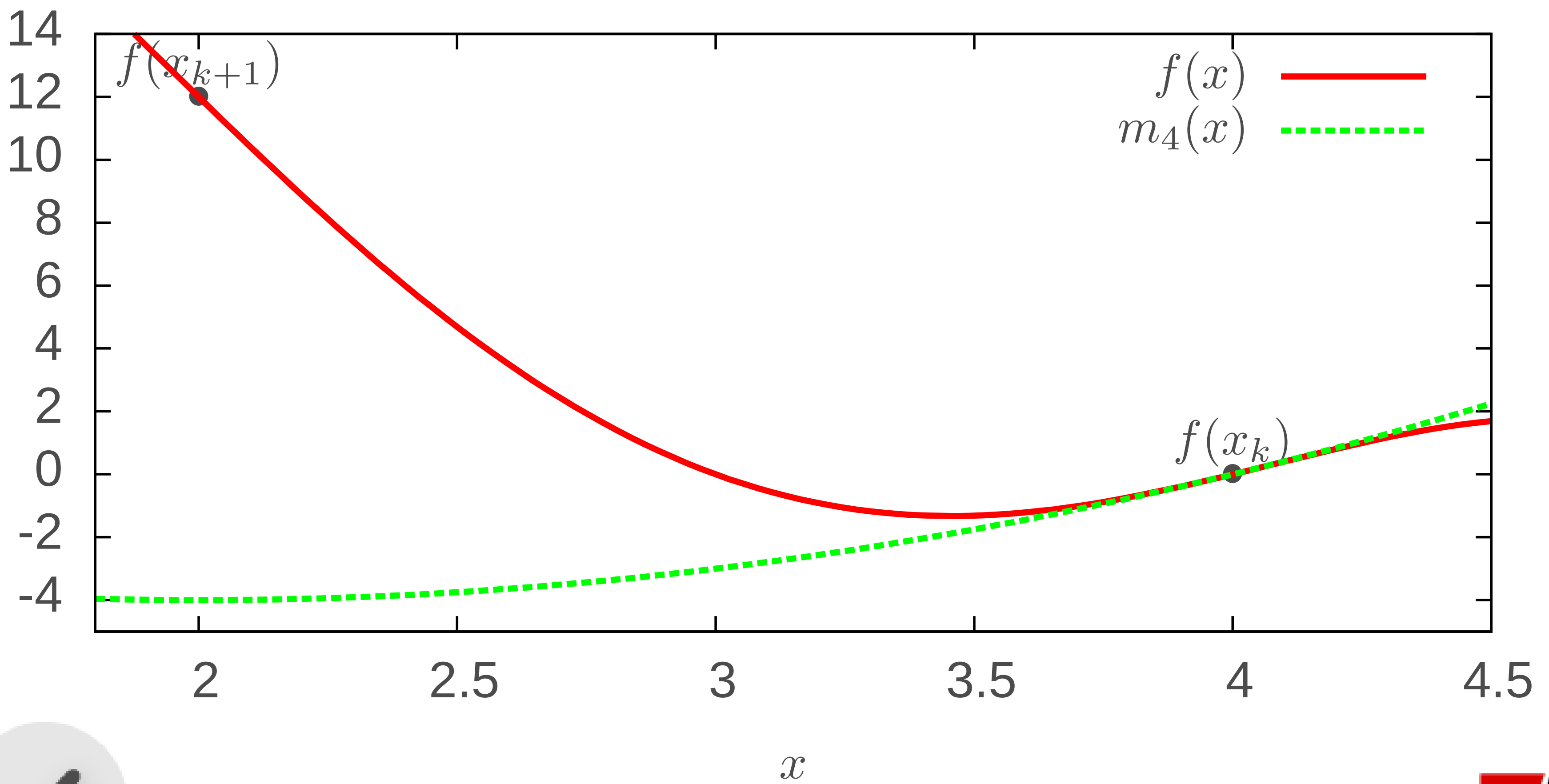


Modèle quadratique

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

$x_k = 4$

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

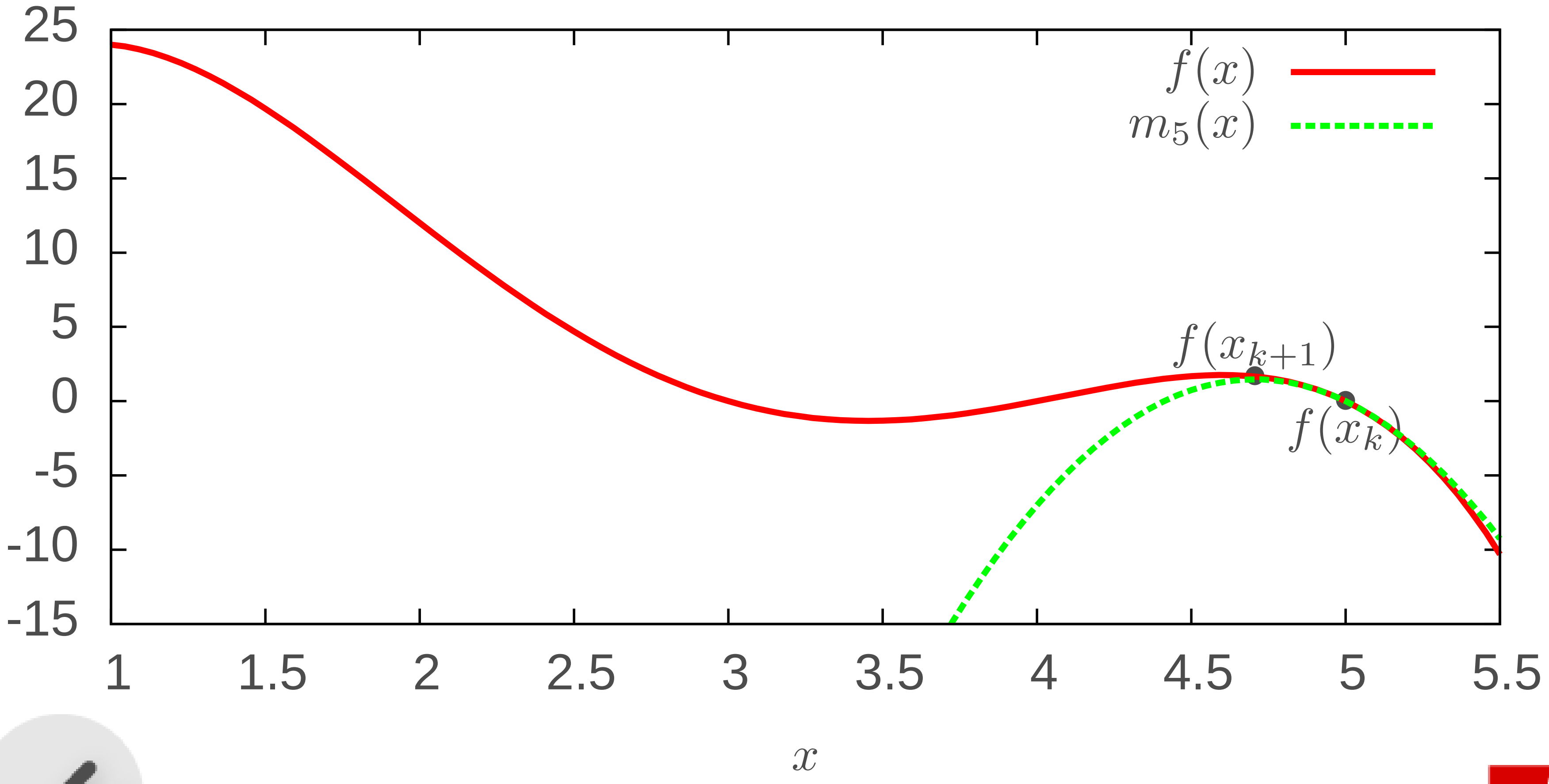


Modèle quadratique

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

$x_k = 5$

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



TRANSP-OR

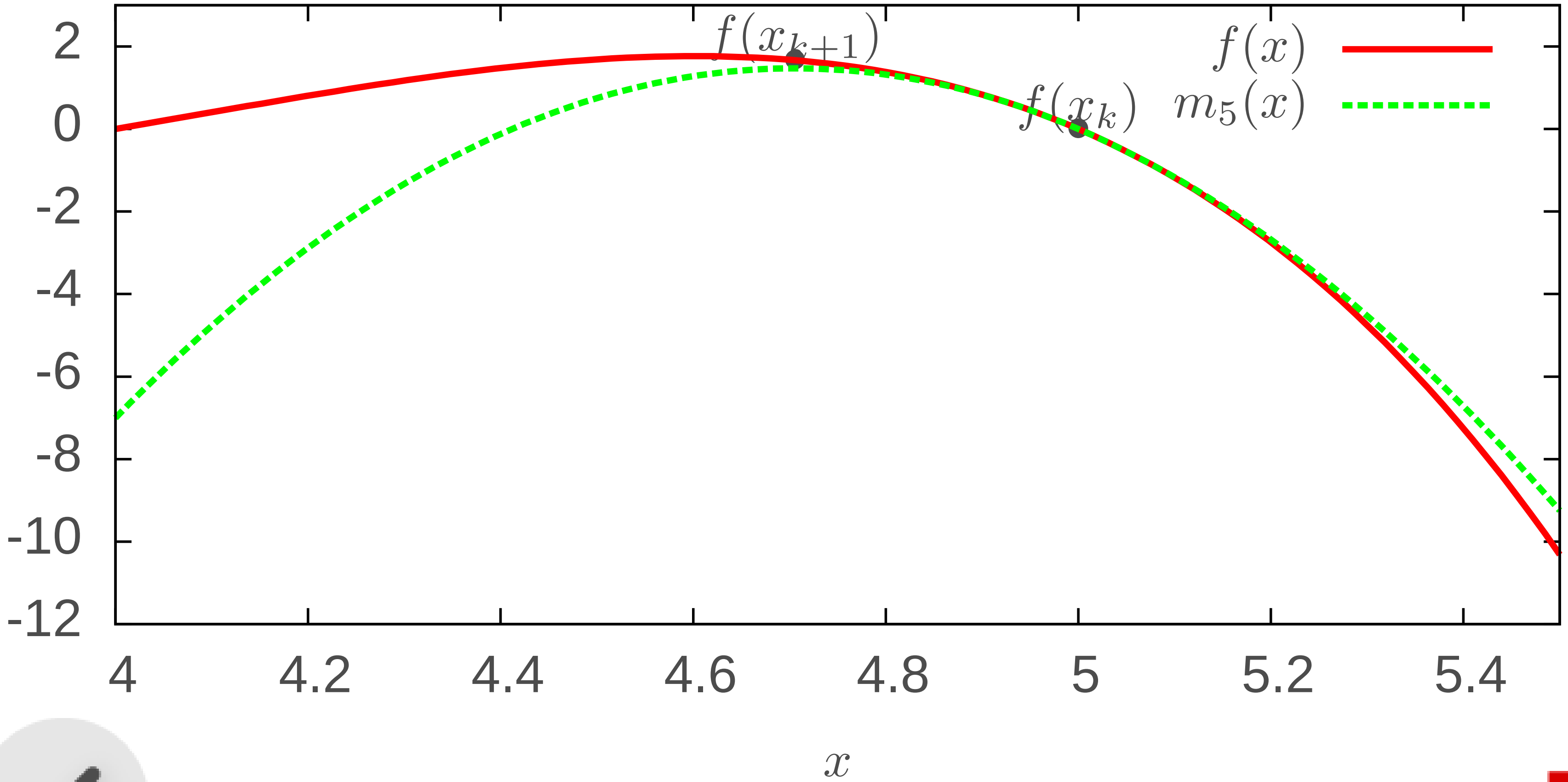
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Modèle quadratique

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

$x_k = 5$

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



Région de confiance

Idées:

- Modèle quadratique à chaque itération
- Bonne approximation dans un voisinage (Taylor)
- Région de confiance:

$$\|x_k - x\| \leq \Delta_k$$

Région de confiance

Sous-problème de région de confiance

Soit $f : \mathbb{R}^n \to \mathbb{R}$ *une fonction deux fois différentiable,* $\widehat{x} \in \mathbb{R}^n$*,* $m_{\widehat{x}}$ *le modèle quadratique de* f *en* $\widehat{x}$*, et* $\Delta_k > 0$*.*

Le sous-problème de région de confiance *est le problème de minimisation suivant*

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

sous contraintes

$$\|d\| \leq \Delta_k.$$

Région de confiance

Questions :

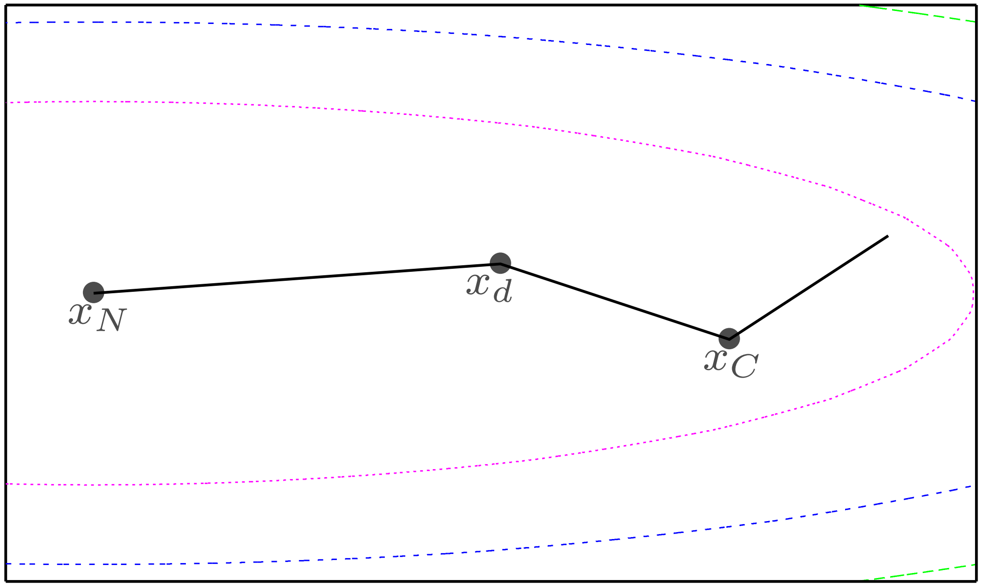
- Comment résoudre ce sous-problème ?
- Comment déterminer Δ_k ?

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Méthode “dogleg”

- Inutile de résoudre le sous-problème exactement
- Si la région de confiance est petite, l’approximation de Taylor du 1er ordre est déjà bonne : Point de Cauchy
- Si la région de confiance est large : Point de Newton



Méthode “dogleg”

Trajectoire :

$$\widehat{x} + p(\alpha) = \begin{cases} \alpha d_C & 0 \leq \alpha \leq 1 \\ d_C + (\alpha - 1)(x_d - x_C) & 1 \leq \alpha \leq 2 \\ (\eta(3-\alpha) + \alpha - 2) d_N & 2 \leq \alpha \leq 3 \end{cases}$$

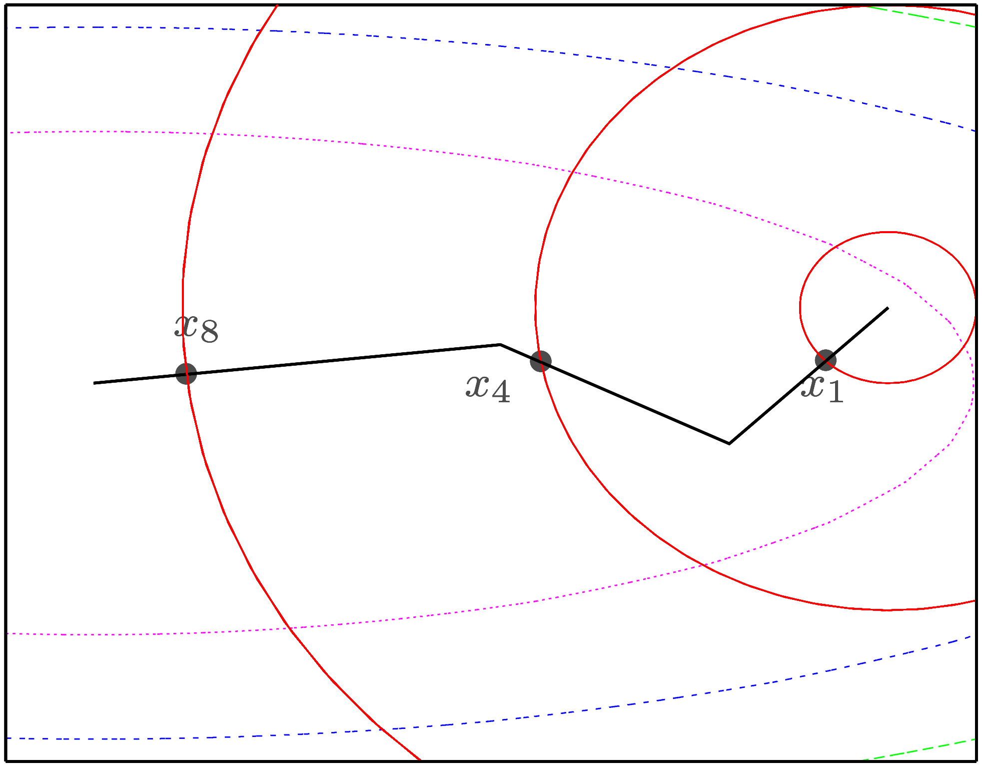
$$x_C = \widehat{x} + d_C \qquad d_C = -\frac{\nabla f(\widehat{x})^T \nabla f(\widehat{x})}{\nabla f(\widehat{x})^T \nabla^2 f(\widehat{x}) \nabla f(\widehat{x})} \nabla f(\widehat{x})$$

$$x_d = \widehat{x} + \eta d_N \qquad d_N = -\nabla^2 f(\widehat{x})^{-1} \nabla f(\widehat{x})$$

$$\eta \leq 1 \qquad \eta = 0.8 \|d_C\| / \|d_N\| + 0.2$$

Méthode “dogleg”

Suivre la trajectoire jusqu’au bord de la région de confiance



Méthode “dogleg”

Lemme 1 *Soient* x*,* $d \in \mathbb{R}^n$*,* $d \neq 0$*, et* $\Delta > 0$ *tels que* $\|x\|_2 \leq \Delta$*, c'est-à-dire* x *est à l'intérieur de la région de confiance.*
Le pas λ *tel que*

$$\|x + \lambda d\|_2 = \Delta$$

est donné par

$$\lambda = \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

avec

$$\begin{aligned} a &= d^T d \\ b &= 2x^T d \\ c &= x^T x - \Delta^2. \end{aligned}$$

(sans preuve)

Méthode “dogleg”

Attention : concavité

- Si concave dans la direction de Cauchy : suivre cette direction jusqu’au bord de la région de confiance
- Si concave dans la direction de Newton : considérer uniquement la direction de Cauchy

Algorithme : Méthode "dogleg"

Objectif

Trouver une approximation de la solution du sous-problème de région de confiance

$$\min_{d \in \mathbb{R}^n} d^T \nabla f(\widehat{x}) + \frac{1}{2} d^T \nabla^2 f(\widehat{x}) d.$$

sous contrainte

$$\|d\|_2 \leq \Delta$$

Input

- La valeur du gradient de la fonction au point courant: $\nabla f(\widehat{x}) \in \mathbb{R}^n \neq 0$.

Algorithme : Méthode “dogleg”

Input (suite)

- La valeur du Hessien de la fonction au point courant: $\nabla^2 f(\widehat{x}) \in \mathbb{R}^{n \times n}$.
- Le rayon de la région de confiance: $\Delta > 0$.

Output

Solution approximative d^*

Point de Cauchy

- Calculer la courbure dans la direction de plus forte pente

$$\beta = \nabla f(\widehat{x})^T \nabla^2 f(\widehat{x}) \nabla f(\widehat{x})$$

Algorithme : Méthode "dogleg"

Point de Cauchy (suite)

- Si $\beta \leq 0$, le modèle n'est pas convexe. STOP avec

$$d^* = -\frac{\Delta}{\|\nabla f(\widehat{x})\|} \nabla f(\widehat{x})$$

- Sinon, calculer la direction du point de Cauchy

$$d_C = -\frac{\nabla f(\widehat{x})^T \nabla f(\widehat{x})}{\beta} \nabla f(\widehat{x}) = -\frac{\nabla f(\widehat{x})^T \nabla f(\widehat{x})}{\nabla f(\widehat{x})^T \nabla^2 f(\widehat{x}) \nabla f(\widehat{x})} \nabla f(\widehat{x})$$

Algorithme : Méthode "dogleg"

Point de Cauchy (suite)

- Si $\|d_C\| \geq \Delta$, le point de Cauchy est hors de la région de confiance. STOP avec

$$d^* = \frac{\Delta}{\|d_C\|} d_C$$

Point de Newton

- Sinon, calculer d_N en résolvant

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

Algorithme : Méthode “dogleg”

Point de Newton (suite)

- Si $d_N^T \nabla^2 f(\widehat{x}) d_N \leq 0$, le modèle n’est pas convexe. STOP avec le point de Cauchy,

$$d^* = d_C.$$

- Si $\|d_N\| \leq \Delta$, le point de Newton est dans la région de confiance. STOP avec $d^* = d_N$.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Algorithme : Méthode "dogleg"

Point dogleg

- Calculer

$$\eta = 0.2 + \frac{0.8\alpha^2}{\beta\,|\nabla f(\widehat{x})^T d_N|} \quad et \quad d_d = \eta d_N$$

- Si $\|d_d\| \leq \Delta$, le point dogleg est à l'intérieur de la région de confiance. STOP avec

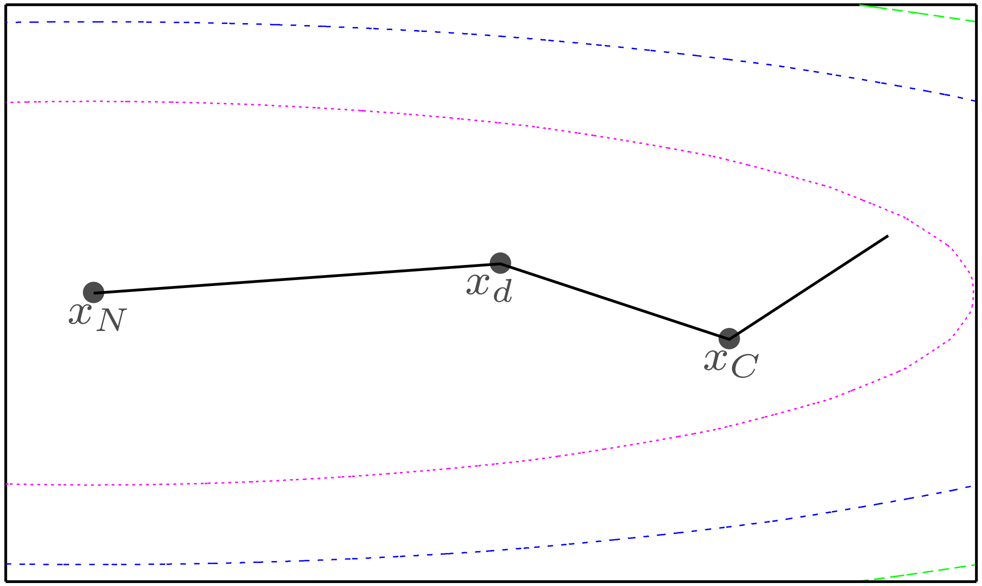
$$d^* = \frac{\Delta}{\|d_N\|} d_N$$

Algorithme : Méthode “dogleg”

Entre Cauchy et dogleg

- Calculer le point d’intersection $d_C + \lambda(d_d - d_C)$ entre le segment reliant le point de Cauchy et le point dogleg, avec la frontière de la région de confiance.
- STOP avec $d^* = d_C + \lambda(d_d - d_C)$.

Méthode “dogleg”



Méthode de Steihaug-Toint

- Problème de région de confiance = problème quadratique
- Pas besoin de résoudre complètement
- Utilisation d'une méthode itérative
- Méthode des gradients conjugués
- Attention à la contrainte
- Attention à l'hypothèse "définie positive"

Méthode de Steihaug-Toint

A chaque itération

- vérifier la courbure dans la direction courante.
  Si négative : avancer jusqu'au bord et STOP
- vérifier que le pas est admissible.
  Si non : avancer jusqu'au bord et STOP.

Algorithme : Méthode de Steihaug-Toint

Objectif

Trouver une approximation de la solution du sous-problème de région de confiance

$$\min_x \frac{1}{2} x^T Q x + x^T b$$

sous contrainte

$$\|x\|_2 \leq \Delta$$

Input

- La matrice $Q \in \mathbb{R}^{n \times n}$.
- Le vecteur $b \in \mathbb{R}^n$.
- Le rayon de la région de confiance Δ

Algorithme : Méthode de Steihaug-Toint

Output

La solution approximative $x^* \in \mathbb{R}^n$.

Initialisation

$k = 1$, $x_1 = 0$, $d_1 = -b$.

Itérations

- Vérifier la courbure de la fonction dans la direction d_k. Si

$$d_k^T Q d_k \leq 0$$

alors $x^* = x_k + \lambda d_k$ où λ est obtenu pour que x^* soit sur le bord de la région

Algorithme : Méthode de Steihaug-Toint

Itérations (suite)

- Calculer le pas

$$\alpha_k = -\frac{d_k^T(Qx_k + b)}{d_k^T Q d_k}$$

- Calculer l'itéré suivant

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

- Si $\|x_{k+1}\| > \Delta$, alors $x^* = x_k + \lambda d_k$ où λ est obtenu pour que x^* soit sur le bord de la région

Algorithme : Méthode de Steihaug-Toint

Itérations (suite)

- Calculer

$$\beta_{k+1} = \frac{\nabla f(x_{k+1})^T \nabla f(x_{k+1})}{\nabla f(x_k)^T \nabla f(x_k)} = \frac{(Qx_{k+1}+b)^T(Qx_{k+1}+b)}{(Qx_k+b)^T(Qx_k+b)}$$

- Calculer la nouvelle direction

$$d_{k+1} = -Qx_{k+1} - b + \beta_{k+1} d_k$$

- $k = k+1$

Critère d'arrêt

Si $\|\nabla f(x_k)\| = 0$ ou $k = n+1$, alors $x^* = x_k$.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Rayon de la région de confiance

- A ce stade, on a deux manières de résoudre le sous-problème de la région de confiance
- Mais comment déterminer Δ, le rayon ?
- Réponse : par essais-erreurs.
  1. On commence par une valeur arbitraire
  2. Evaluation de la qualité du point obtenu
  3. Adaptation du rayon en fonction de la qualité

Rayon de la région de confiance

Appelons d^* la solution du sous-problème de région de confiance

- Réduction prédite par le modèle :

$$m_{\widehat{x}}(\widehat{x}) - m_{\widehat{x}}(\widehat{x} + d^*)$$

- Réduction effective de la fonction :

$$f(\widehat{x}) - f(\widehat{x} + d^*).$$

Si le modèle est fiable, ces deux quantités devraient être proches.
Soit

$$\rho = \frac{f(\widehat{x}) - f(\widehat{x} + d^*)}{m_{\widehat{x}}(\widehat{x}) - m_{\widehat{x}}(\widehat{x} + d^*)}$$

Rayon de la région de confiance

Soient $0 < \eta_1 \leq \eta_2 < 1$ ($\eta_1 = 0.01$, $\eta_2 = 0.9$)
Adéquation entre le modèle et la fonction en d^*

$\rho \geq \eta_2$ très bonne augmenter le rayon

$\eta_1 \leq \rho < \eta_2$ bonne maintenir le rayon

$\rho < \eta_1$ mauvaise réduire le rayon

Algorithme : Newton avec région de confiance

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 $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$;
- Le hessien de la fonction $\nabla^2 f : \mathbb{R}^n \to \mathbb{R}^{n \times n}$;
- Une première approximation de la solution $x_0 \in \mathbb{R}^n$;

Algorithme : Newton avec région de confiance

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

Initialisation

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

Algorithme : Newton avec région de confiance

Itérations

- Calculer d_k en résolvant (approximativement) le sous-problème de région de confiance en utilisant “dogleg” ou “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 : Newton avec région de confiance

Itérations

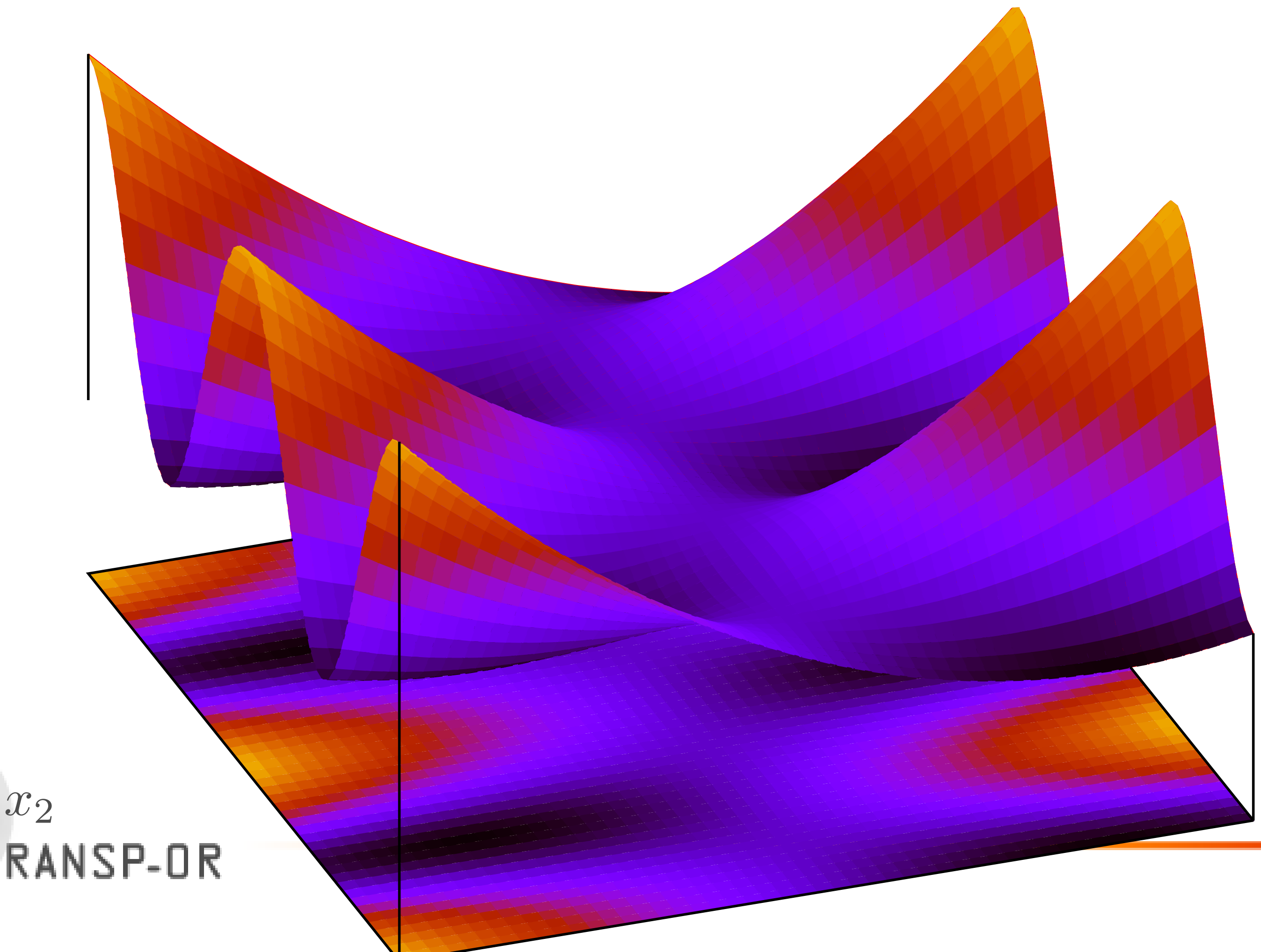
- Si $\rho < \eta_1$, alors
  - $x_{k+1} = x_k$,
  - $\Delta_{k+1} = \frac{1}{2}\|d_k\|$.
- Si $\rho \geq \eta_1$, alors
  - $x_{k+1} = x_k + d_k$,
  - Si $\rho \geq \eta_2$, alors
    - $\Delta_{k+1} = 2\Delta_k$

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

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

Newton avec région de confiance

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

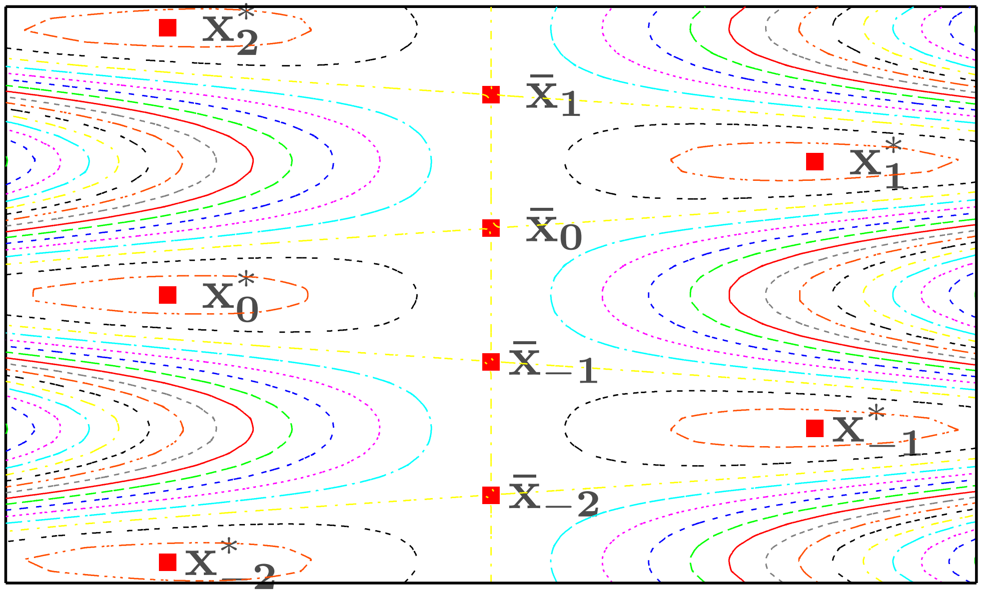


TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

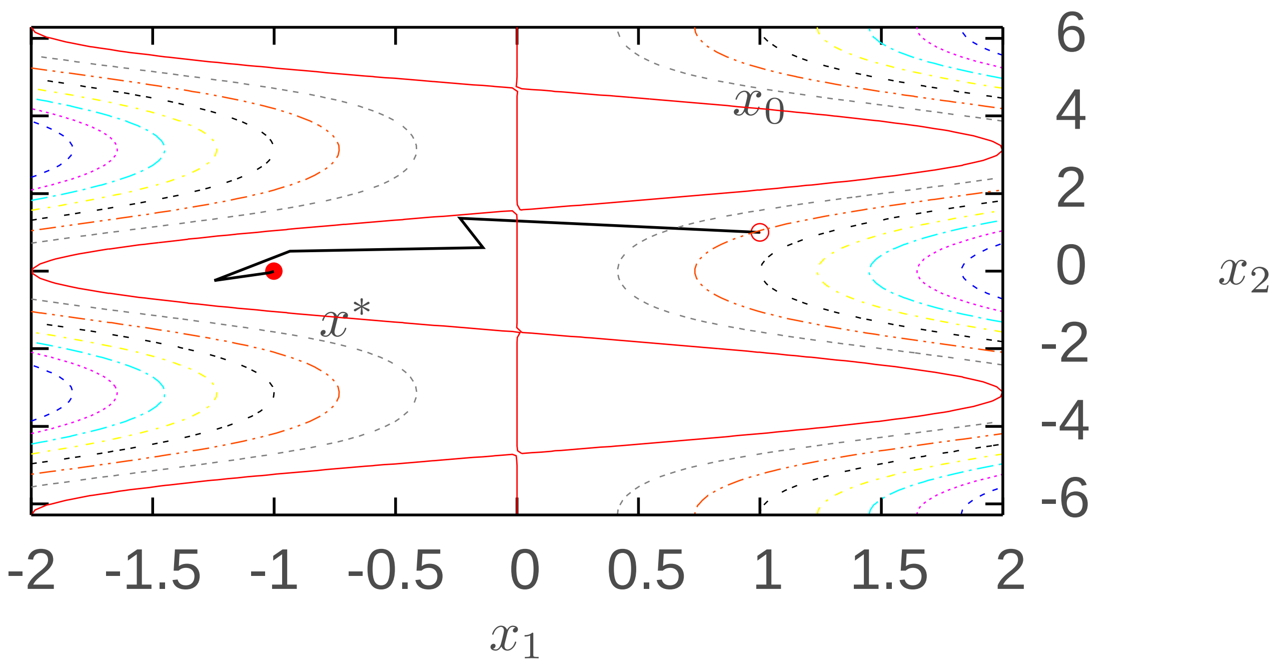
Newton avec région de confiance

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



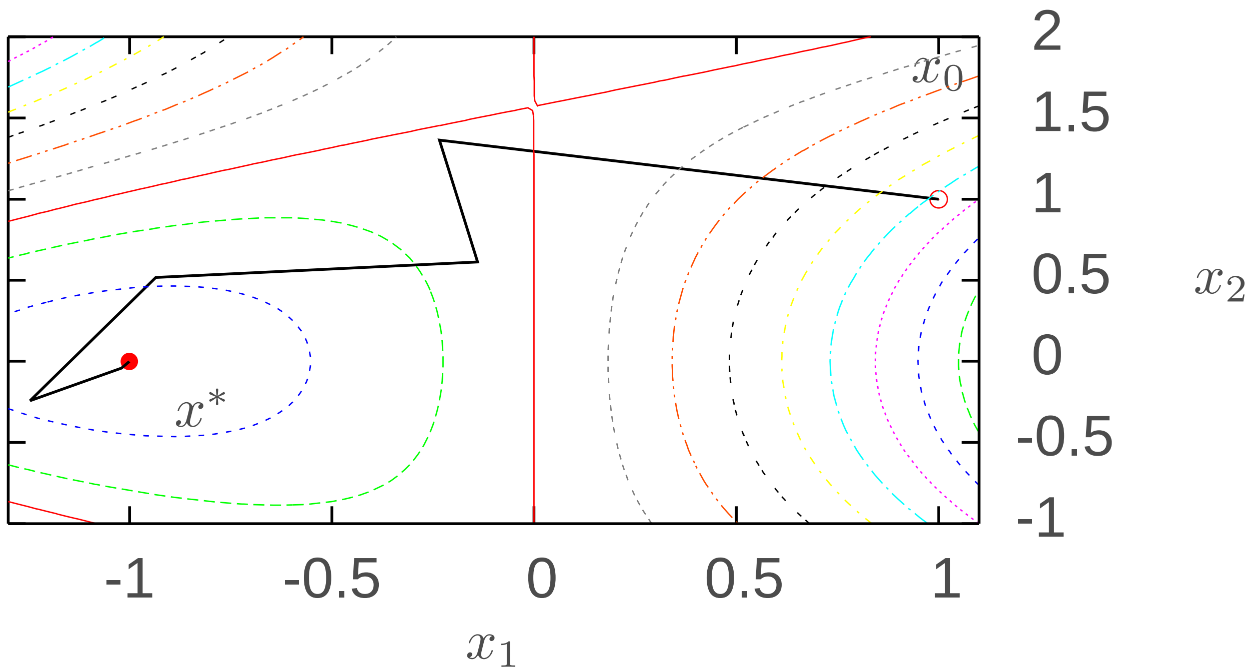
Point de départ $x_0 = (1\ 1)^T$.

Newton avec région de confiance



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

Newton avec région de confiance



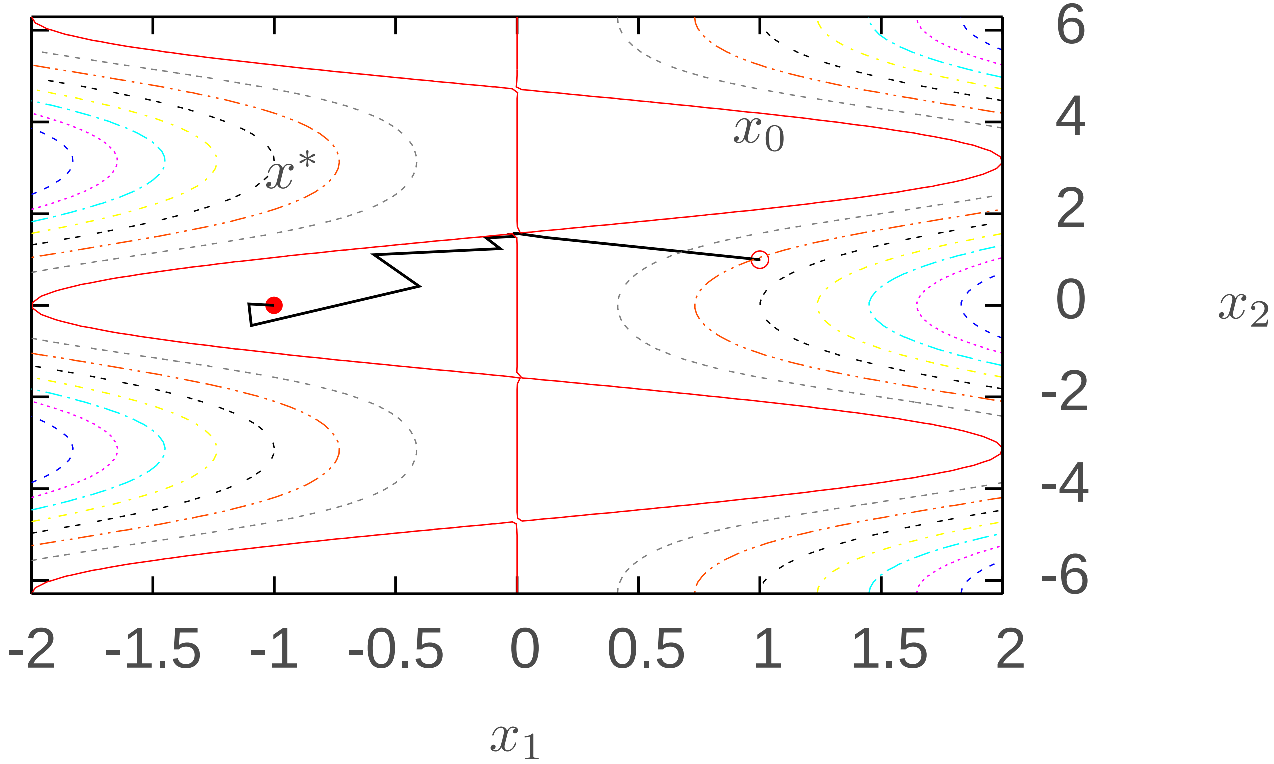
$$x^* = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \quad \nabla f(x^*) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \nabla^2 f(x^*) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

x_k		$f(x_k)$	Δ_k	ρ		
+1.00000e+00	+1.00000e+00	+1.04030e+00	+1.0e+01			
-2.33845e-01	+1.36419e+00	-2.06286e-02	+2.0e+01	+9.61445e-01	2	++
-1.39549e-01	+6.12415e-01	-1.04451e-01	+4.0e+01	+9.59237e-01	-2	++
-9.34497e-01	+5.18458e-01	-3.75047e-01	+8.0e+01	+9.89241e-01	-2	++
-1.24534e+00	-2.41828e-01	-4.33668e-01	+8.0e+01	+3.53577e-01	2	+
-1.01925e+00	-3.99531e-02	-4.99001e-01	+1.6e+02	+1.06883e+00	2	++
-1.00077e+00	-7.03374e-04	-4.99999e-01	+3.2e+02	+1.01414e+00	2	++
-1.00000e+00	-5.40691e-07	-5.00000e-01	+6.4e+02	+1.00035e+00	2	++

- point de Cauchy suite à la courbure négative de la direction de Newton
- pas de Newton pur

Newton avec région de confiance

$\Delta_0 = 1$

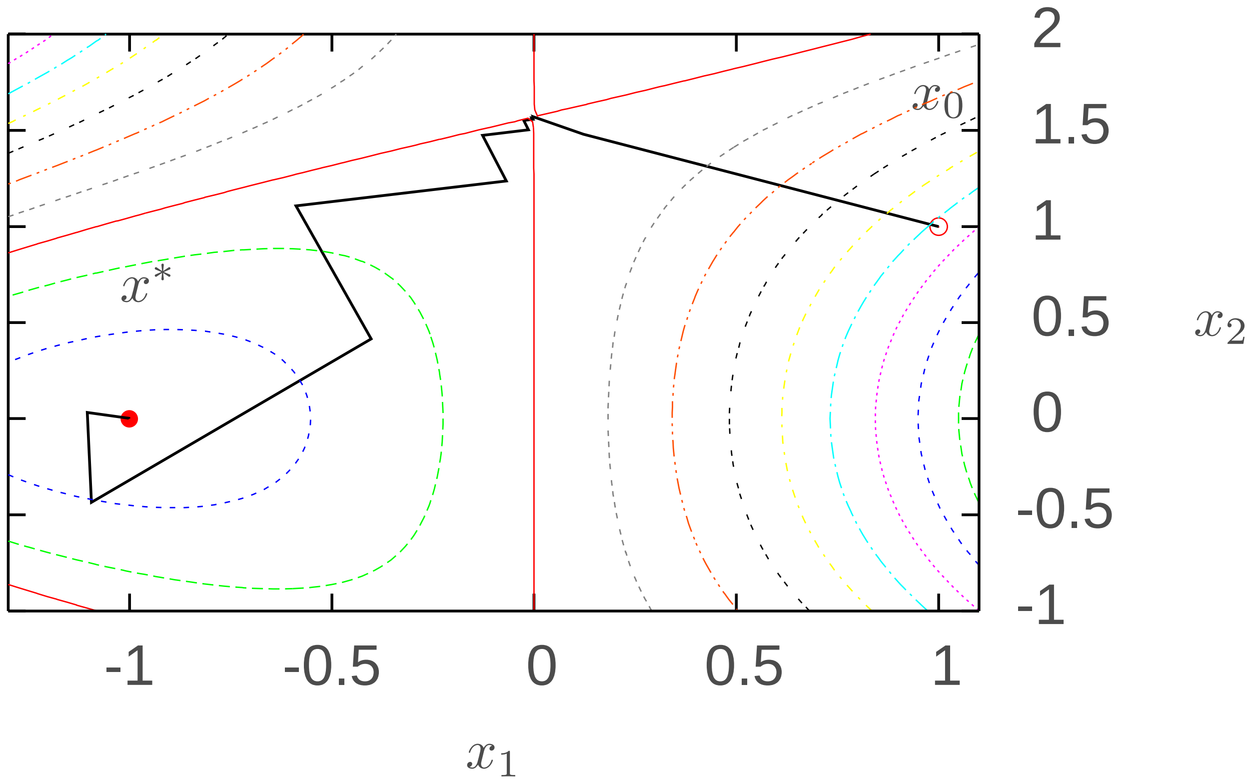


TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Newton avec région de confiance

$\Delta_0 = 1$



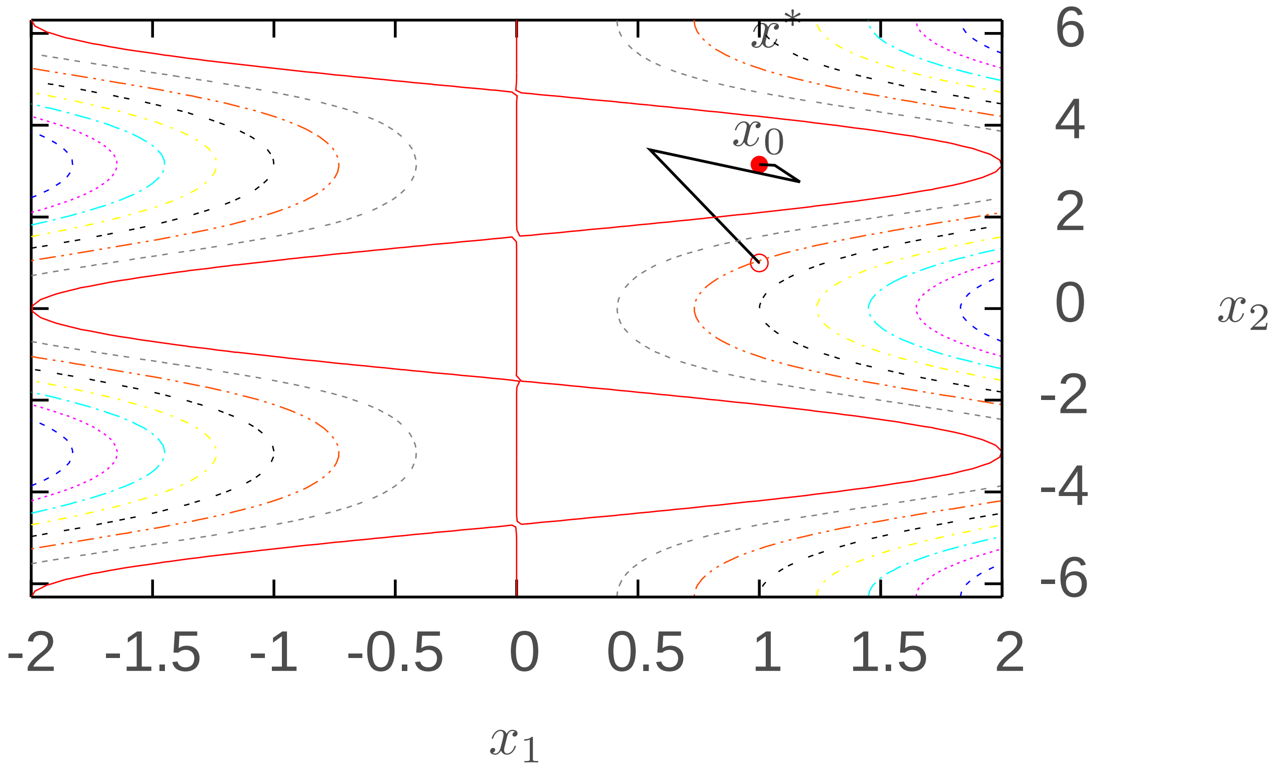
- pas de Cauchy partiel
- point de Cauchy suite à la courbure négative de la direction de Newton

x_k		$f(x_k)$	Δ_k	ρ		
+1.00000e+00	+1.00000e+00	+1.04030e+00	+1.00000e+00			
+1.22417e-01	+1.47943e+00	+1.86628e-02	+2.00000e+00	+9.47588e-01	1	++
-1.01629e-03	+1.57003e+00	-2.61464e-07	+4.00000e+00	+9.97536e-01	2	++
-5.36408e-04	+1.56809e+00	-1.30949e-06	+8.00000e+00	+1.00000e+00	-2	++
-5.08985e-03	+1.56696e+00	-6.55830e-06	+1.60000e+01	+9.99998e-01	-2	++
-2.68657e-03	+1.55723e+00	-3.28448e-05	+3.20000e+01	+1.00000e+00	-2	++
-2.54882e-02	+1.55160e+00	-1.64466e-04	+6.40000e+01	+9.99957e-01	-2	++
-1.34638e-02	+1.50289e+00	-8.22887e-04	+1.28000e+02	+1.00002e+00	-2	++
-1.27230e-01	+1.47480e+00	-4.10176e-03	+2.56000e+02	+9.98929e-01	-2	++
-6.84750e-02	+1.23764e+00	-2.00488e-02	+5.12000e+02	+1.00051e+00	-2	++
-5.88466e-01	+1.10750e+00	-8.98399e-02	+1.02400e+03	+9.77015e-01	-2	++
-4.02533e-01	+4.16075e-01	-2.87173e-01	+2.04800e+03	+1.01116e+00	-2	++
-4.02533e-01	+4.16075e-01	-2.87173e-01	+1.09534e+00	-2.88565e+00	2	-
-1.09350e+00	-4.33824e-01	-3.94333e-01	+1.09534e+00	+2.99489e-01	4	+
-1.10395e+00	+3.38629e-02	-4.93964e-01	+2.19067e+00	+9.35399e-01	2	++
-1.00047e+00	+3.16268e-03	-4.99995e-01	+4.38135e+00	+1.00813e+00	2	++
-1.00000e+00	+1.44712e-06	-5.00000e-01	+8.76269e+00	+1.00045e+00	2	++
-1.00000e+00	+7.23075e-12	-5.00000e-01	+1.75254e+01	+1.00001e+00	2	++

- dogleg entre Cauchy et Newton
- pas de Newton pur

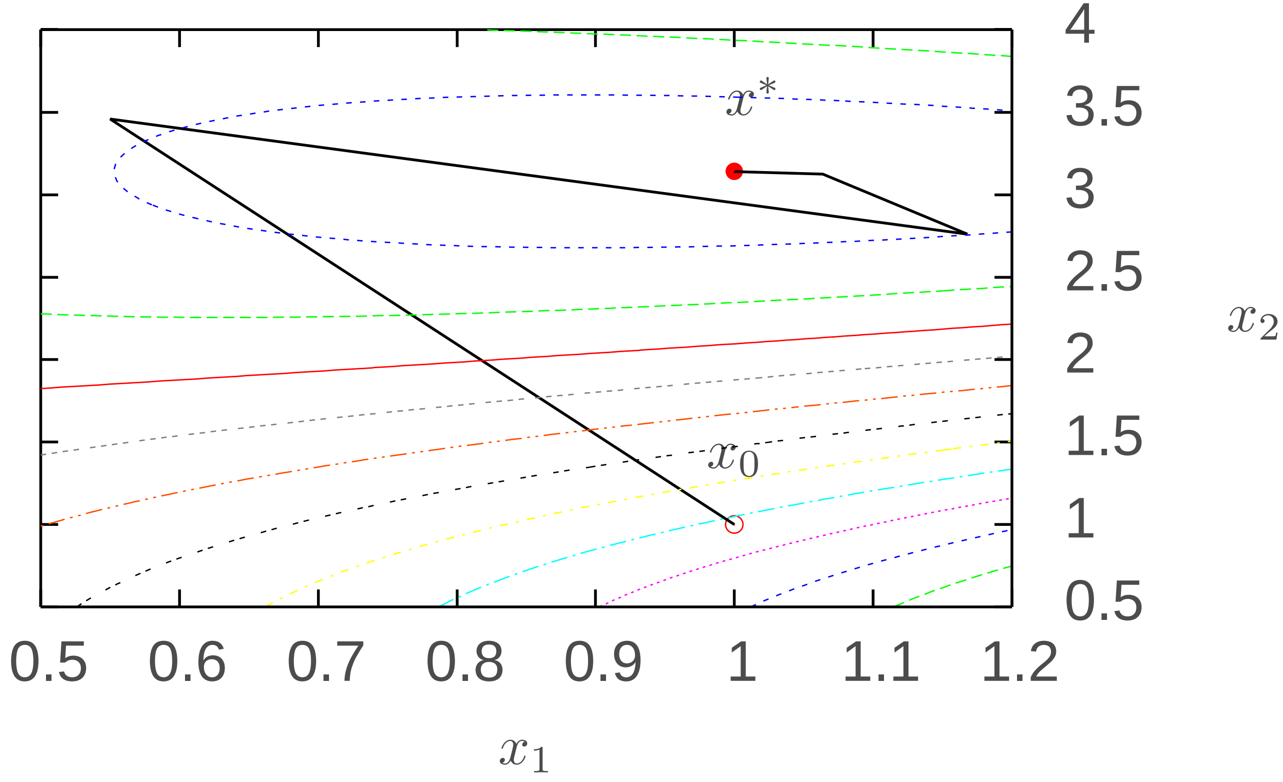
Newton avec région de confiance

$\Delta_0 = 10$, Steihaug-Toint



Newton avec région de confiance

$\Delta_0 = 10$, Steihaug-Toint



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

courbure négative

x_k		$f(x_k)$	Δ_k	ρ		
+1.00000e+00	+1.00000e+00	+1.04030e+00	+1.00000e+01			
+1.00000e+00	+1.00000e+00	+1.04030e+00	+5.00000e+00	-7.51975e-02	3	-
+1.00000e+00	+1.00000e+00	+1.04030e+00	+2.50000e+00	-1.23991e-01	3	-
+5.50230e-01	+3.45921e+00	-3.71332e-01	+2.50000e+00	+4.19624e-01	3	+
+1.16790e+00	+2.76142e+00	-4.02518e-01	+2.50000e+00	+1.70028e-01	1	+
+1.06365e+00	+3.12536e+00	-4.97834e-01	+5.00000e+00	+1.04357e+00	1	++
+1.00012e+00	+3.14062e+00	-5.00000e-01	+1.00000e+01	+1.00343e+00	1	++
+1.00000e+00	+3.14159e+00	-5.00000e-01	+2.00000e+01	+1.00011e+00	1	++

convergence