Optimization and Simulation

Unconstrained optimization

Michel Bierlaire

michel.bierlaire@epfl.ch

Transport and Mobility Laboratory

Optimality conditions

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

Necessary optimality conditions:

- Let x^* be a local minimum of $f : \mathbb{R}^n \to \mathbb{R}$.
- (first order condition) If f is differentiable in an open neighborhood of x^*, then

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

- (second order condition) If f is twice differentiable in an open neighborhood of x^*, then

$$\nabla^2 f(x^*) \geq 0,$$

  meaning that $\nabla^2 f(x^*)$ is *positive semidefinite*.

Optimality conditions

Sufficient optimality conditions

- Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable in an open set $V \subseteq \mathbb{R}^n$.
- Let $x^* \in V$ such that
  - (first order condition)

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

  - (second order condition)

$$\nabla^2 f(x^*) > 0,$$

    meaning that $\nabla^2 f(x^*)$ is *positive definite*.
- Then x^* is a local minimum of f.

Optimality conditions

Sufficient conditions for global optimality

- Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a continuous function
- Let $x^* \in \mathbb{R}^n$ be a local minimum of f
- If f is convex, then x^* is a global minimum of f.
- If f is strictly convex, then x^* is the unique global minimum of f.

Optimality conditions

Consider the quadratic problem:

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2} x^T Q x + g^T x + c$$

where $Q \in \mathbb{R}^{n \times n}$ is symmetric.

1. If Q is not positive semidefinite, then the problem has no solution, meaning that there is no $x^* \in \mathbb{R}^n$ which is a local minimum.
2. If Q is positive definite, then

$$x^* = -Q^{-1} g$$

is the unique global minimum.

Algorithms

- Solving systems of equations: $\nabla f(x) = 0$
  - Newton
  - Quasi-Newton
- Unconstrained optimization
  - Quadratic problems
  - Local Newton
  - Linesearch
  - Quasi-Newton

Solving systems of equations

The problem: find x^* such $F(x^*) = 0$, where $F : \mathbb{R}^n \to \mathbb{R}^n$.
Newton's method:

- Start at an arbitrary iterate $x_0 \in \mathbb{R}^n$
- At each iteration k, linearize F around x_k
- Find the root of the linear system and defines it as the next iterate

Key object: the gradient matrix, or the Jacobian matrix.

- For a function $F : \mathbb{R}^n \to \mathbb{R}^m$, the gradient and the Jacobian matrices are defined as follows.
- Note: for systems of equations, $n = m$.

Solving systems of equations

Gradient matrix

$$\nabla F(x) = \begin{pmatrix} | & & | \\ \nabla F_1(x) & \cdots & \nabla F_m(x) \\ | & & | \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\partial F_1}{\partial x_1} & \frac{\partial F_2}{\partial x_1} & \cdots & \frac{\partial F_m}{\partial x_1} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F_1}{\partial x_n} & \frac{\partial F_2}{\partial x_n} & \cdots & \frac{\partial F_m}{\partial x_n} \end{pmatrix}.$$

Solving systems of equations

Jacobian matrix

$$J(x) = \nabla F(x)^T = \begin{pmatrix} \text{———} & \nabla F_1(x)^T & \text{———} \\ & \vdots & \\ \text{———} & \nabla F_m(x)^T & \text{———} \end{pmatrix}.$$

Algorithm: Newton's method

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Newton's method

Objective

Find (an approximation of) a solution of the systems of equations:

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

Inputs

- The function $F : \mathbb{R}^n \to \mathbb{R}^n$;
- The Jacobian matrix: $J : \mathbb{R}^n \to \mathbb{R}^{n \times n}$;
- A first approximation of the solution: $x_0 \in \mathbb{R}^n$;
- The requested precision: $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$.

Output

An approximation of the solution $x^* \in \mathbb{R}^n$.

Initialization

$k = 0$.

Newton's method (ctd)

Iterations

1. Compute d_{k+1} solution of

$$J(x_k)d_{k+1} = -F(x_k).$$

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

Stopping criterion

If $\|F(x_k)\| \leq \varepsilon$, then $x^* = x_k$.

Convergence

Consider

- $X \subseteq \mathbb{R}^n$ an open convex set
- $F : X \to \mathbb{R}^n$ a function
- x^* a solution, that is $F(x^*) = 0$,
- $B(x^*, r) \subset X$ a ball of radius r
- $\rho > 0$ a constant.

If

- $J(x^*)$ is invertible
- $\|J(x^*)^{-1}\| \leq 1/\rho$
- J is Lipschitz continuous on $B(x^*, r)$, that is $\forall x, y \in B(x^*, r)$, $\exists M > 0$ such that

$$\|J(x) - J(y)\| \leq M\|x - y\|.$$

Convergence

Then, $\exists \eta > 0$ such that, if

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

then the sequence $(x_k)_k$ defined by

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

is well defined and converges to x^*. Moreover,

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

(quadratic convergence)

Secant method

- Secant, Broyden, or quasi-Newton method.
- Idea: replace the derivative by a secant approximation.
- Trivial in one dimension, more complex in n dimensions.
- Advantages:
  - does not require J anymore,
  - keep good convergence properties (superlinear).

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Secant method

Objective

Find (an approximation of) the solution of the system

$$F(x) = 0. \quad (2)$$

Inputs

- $F : \mathbb{R}^n \to \mathbb{R}^n$
- A first approximation of the solution $x_0 \in \mathbb{R}^n$;
- A first approximation of the Jacobian matrix A_0 (by default $A_0 = I$);
- Required precision $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Secant method (ctd)

Output

An approximation of the solution $x^* \in \mathbb{R}^n$.

Initialization

1. $x_1 = x_0 - A_0^{-1} f(x_0)$.
2. $d_0 = x_1 - x_0$.
3. $y_0 = f(x_1) - f(x_0)$.
4. $k = 1$.

Secant method (ctd)

Iterations

1. Broyden's update:

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

2. Compute d_k solution of $A_k d_k = -F(x_k)$.
3. $x_{k+1} = x_k + d_k$.
4. Compute $y_k = F(x_{k+1}) - F(x_k)$.
5. $k = k + 1$.

Stopping criterion

If $\|F(x_k)\| \leq \varepsilon$, then $x^* = x_k$.

Unconstrained optimization

Quadratic problem:

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2} x^T Q x + b^T x$$

where $Q \in \mathbb{R} n \times n$ is symmetric positive definite.
Optimum: solution of the linear system

$$Qx = -b.$$

Quadratic problem: direct method

Objective

Find the global minimum of the quadratic problem

Input

- $Q \in \mathbb{R}^{n \times n}$ symmetric positive definite.
- $b \in \mathbb{R}^n$.

Output

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

Solving

1. Compute the Cholesky factor: $Q = LL^T$.
2. Compute y^* solution of the lower triangular system $Ly = -b$.
3. Compute x^* solution of the upper triangular system $L^T x = y^*$.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Quadratic problem: iterative method

Conjugate gradients method

- Performs n one-dimensional optimizations
- The n directions are chosen to guarantee that the entire space is spanned
- Allows to solve large-scale problems as the matrix is not needed as such

Unconstrained optimization

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

Local Newton method: apply Newton's method to solve $\nabla f(x^*) = 0$

$$\begin{array}{ccc} F(x) & \rightarrow & \nabla f(x) \\ J(x) & \rightarrow & \nabla^2 f(x) \end{array}$$

Advantage: fast
Problems:

- not guaranteed to converge
- $\nabla^2 f(x_k)^{-1}$ may not exist
- may converge to a point which is not a minimum

Local Newton: geometric interpretation

Quadratic approximation of f

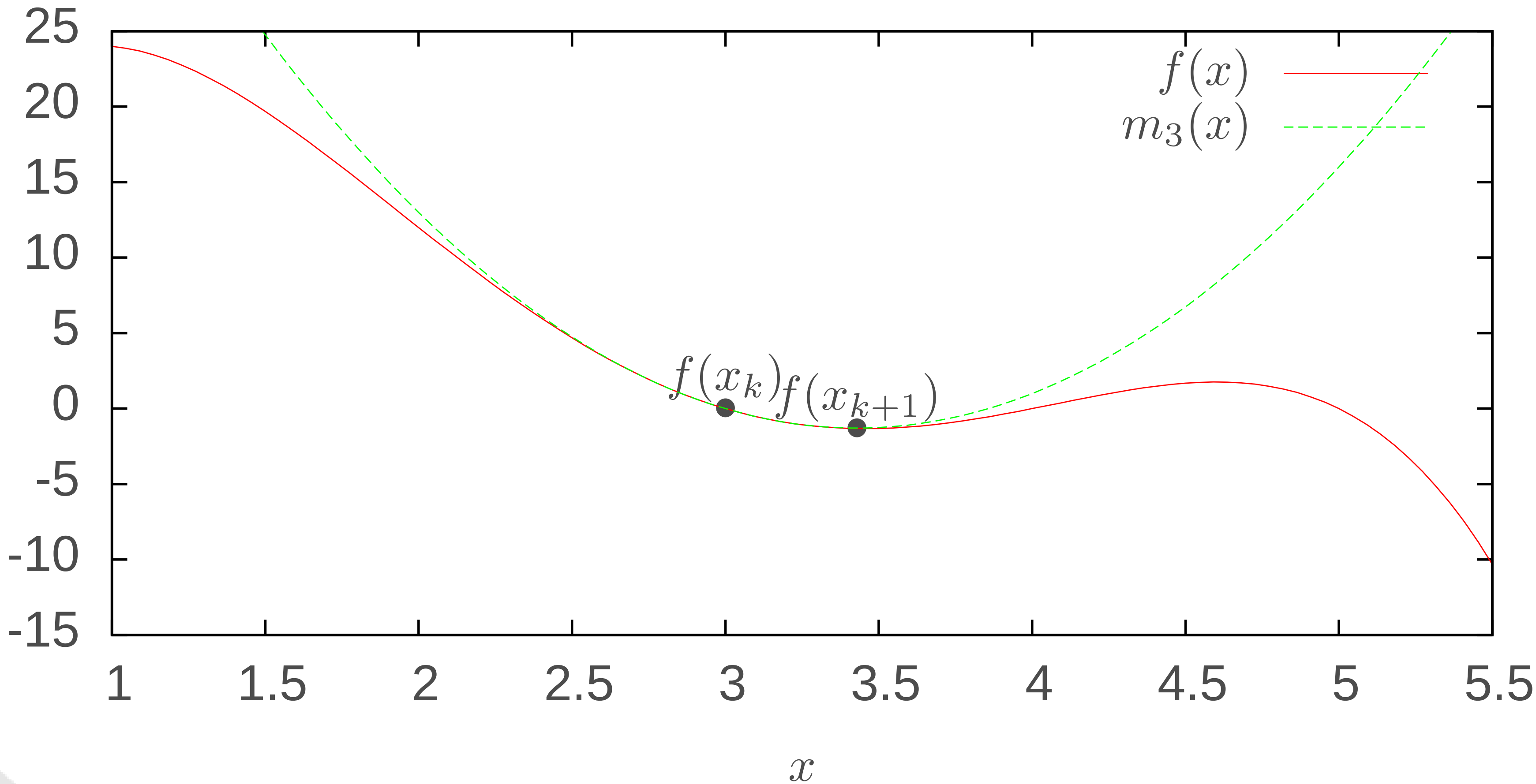
$$m_{x_k}(x) = f(x_k) + (x - x_k)^T \nabla f(x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k).$$

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

1. $x_k = 3$. Quadratic model: $m_3(x) = 7x^2 - 48x + 81$
2. $x_k = 4$. Quadratic model: $m_4(x) = x^2 - 4x$
3. $x_k = 5$. Quadratic model: $m_5(x) = -17x^2 + 160x - 375$.

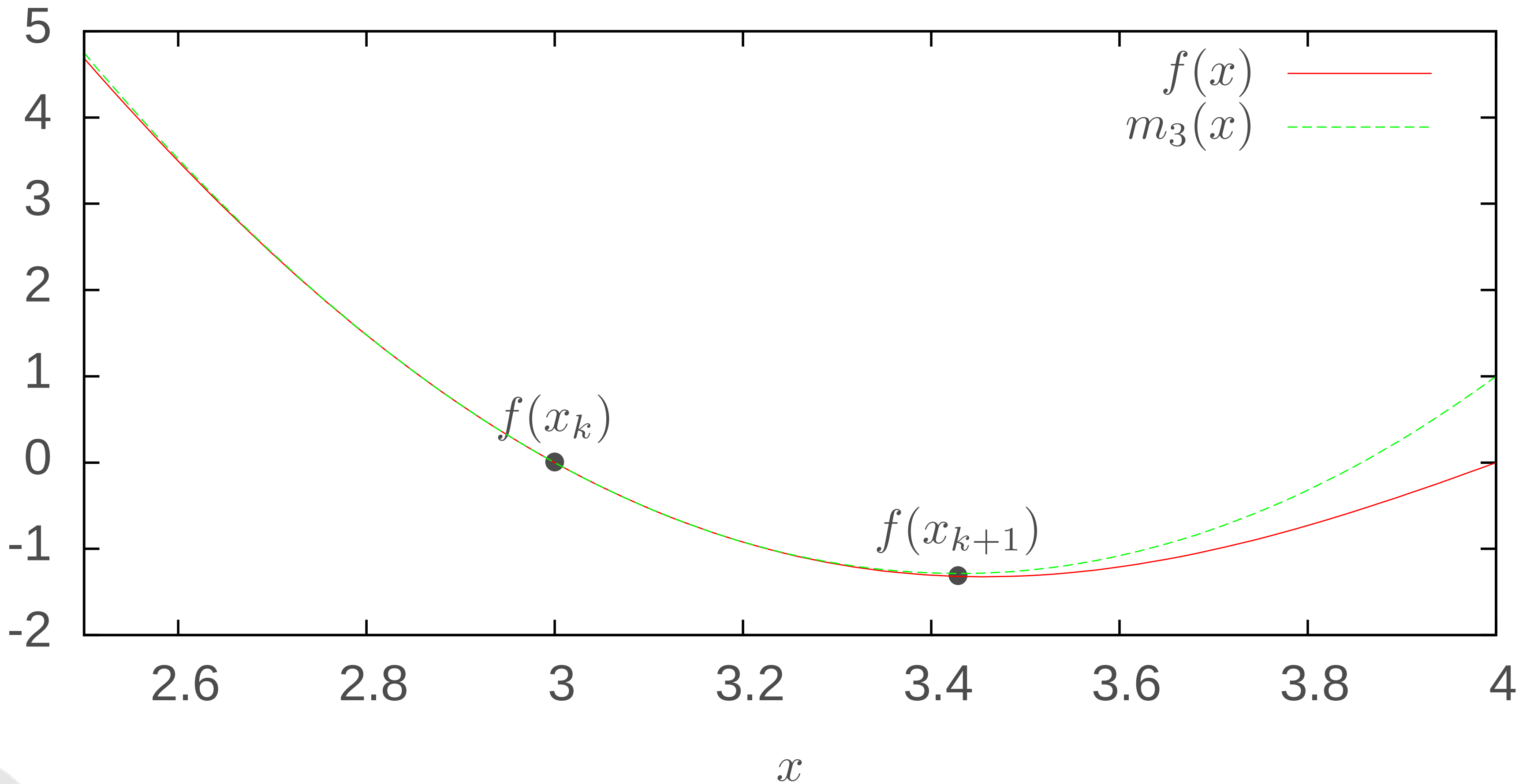
Local Newton: geometric interpretation

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



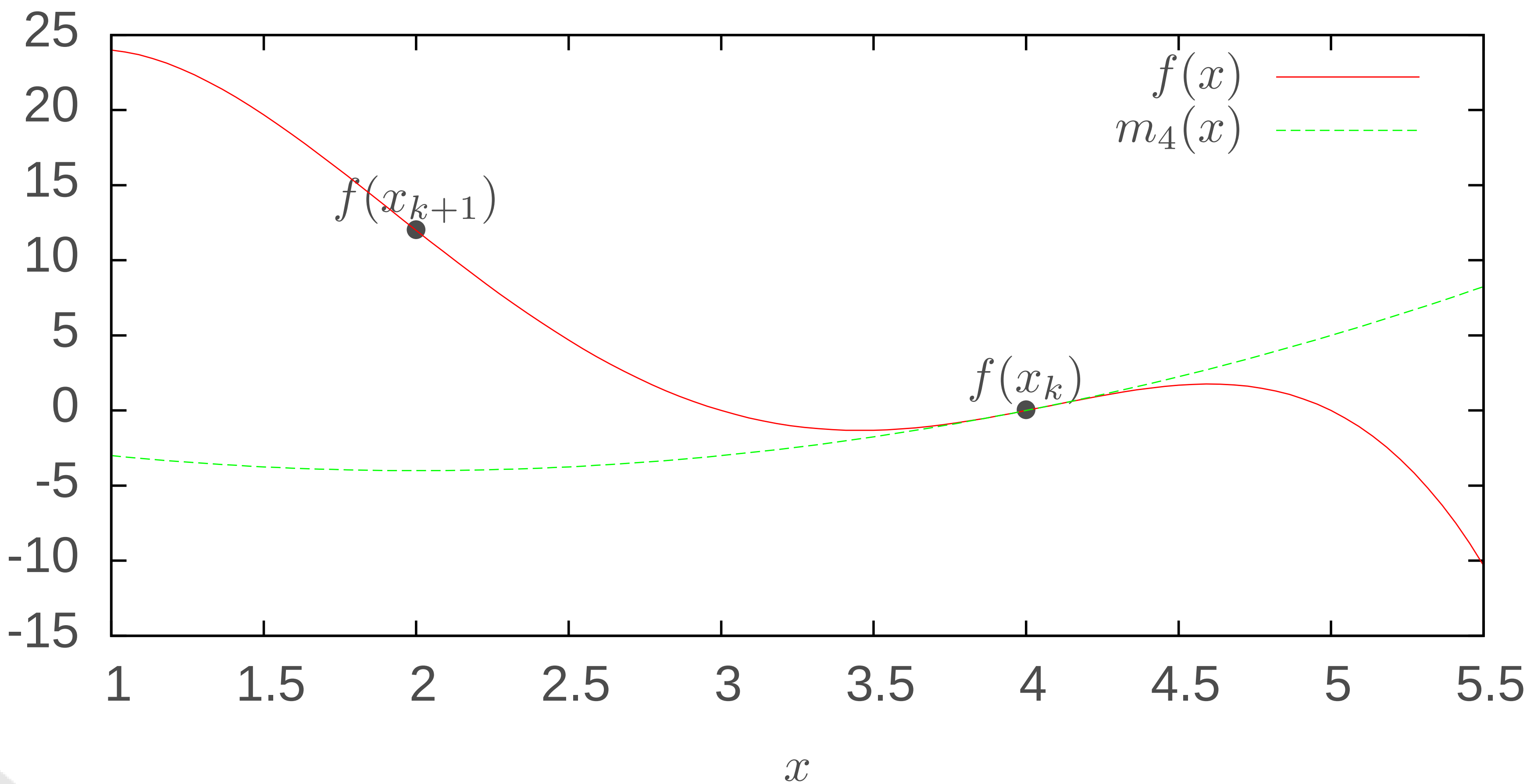
Local Newton: geometric interpretation

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



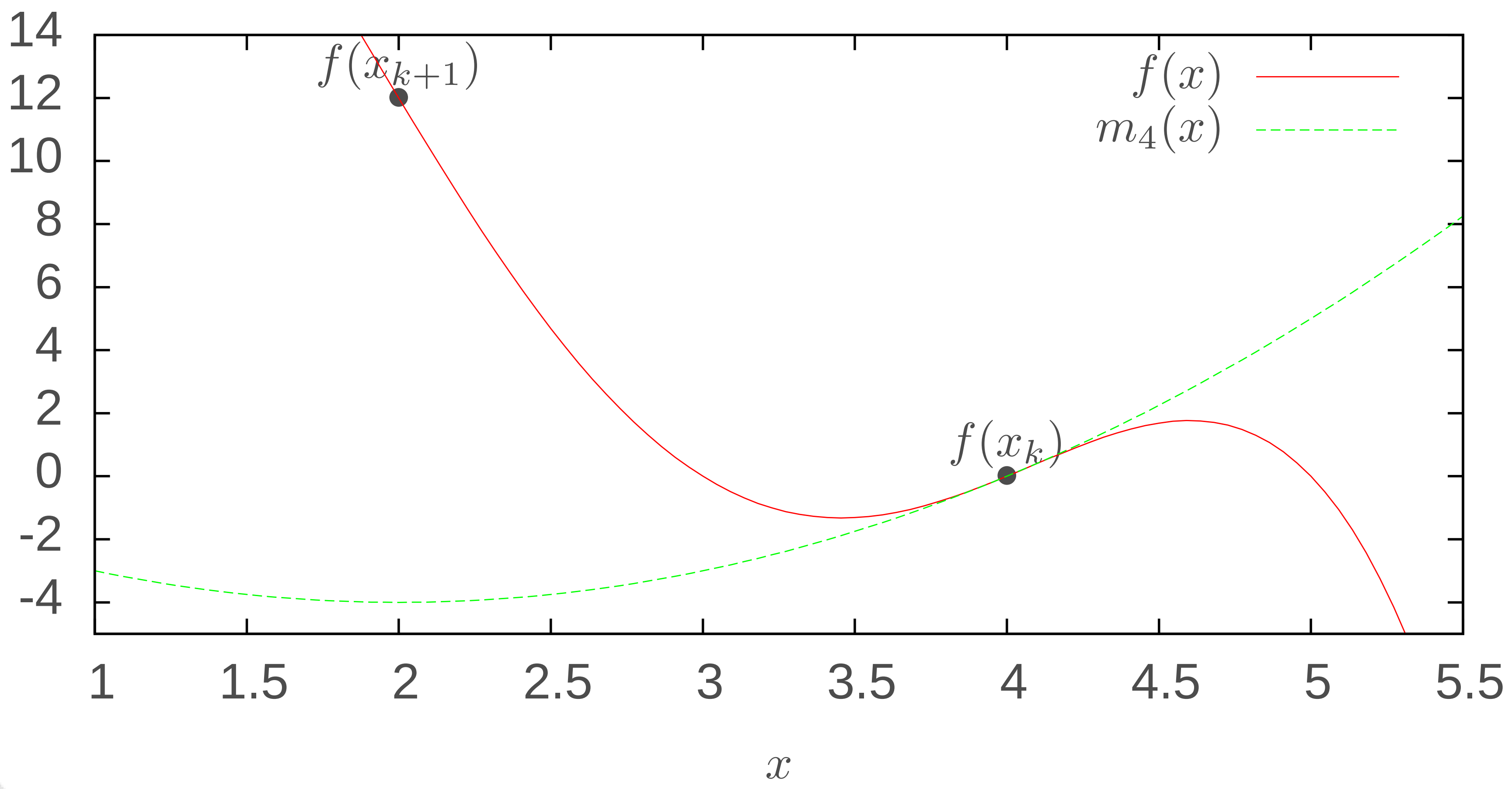
Local Newton: geometric interpretation

$m_4(x) = x^2 - 4x$: bad predictor



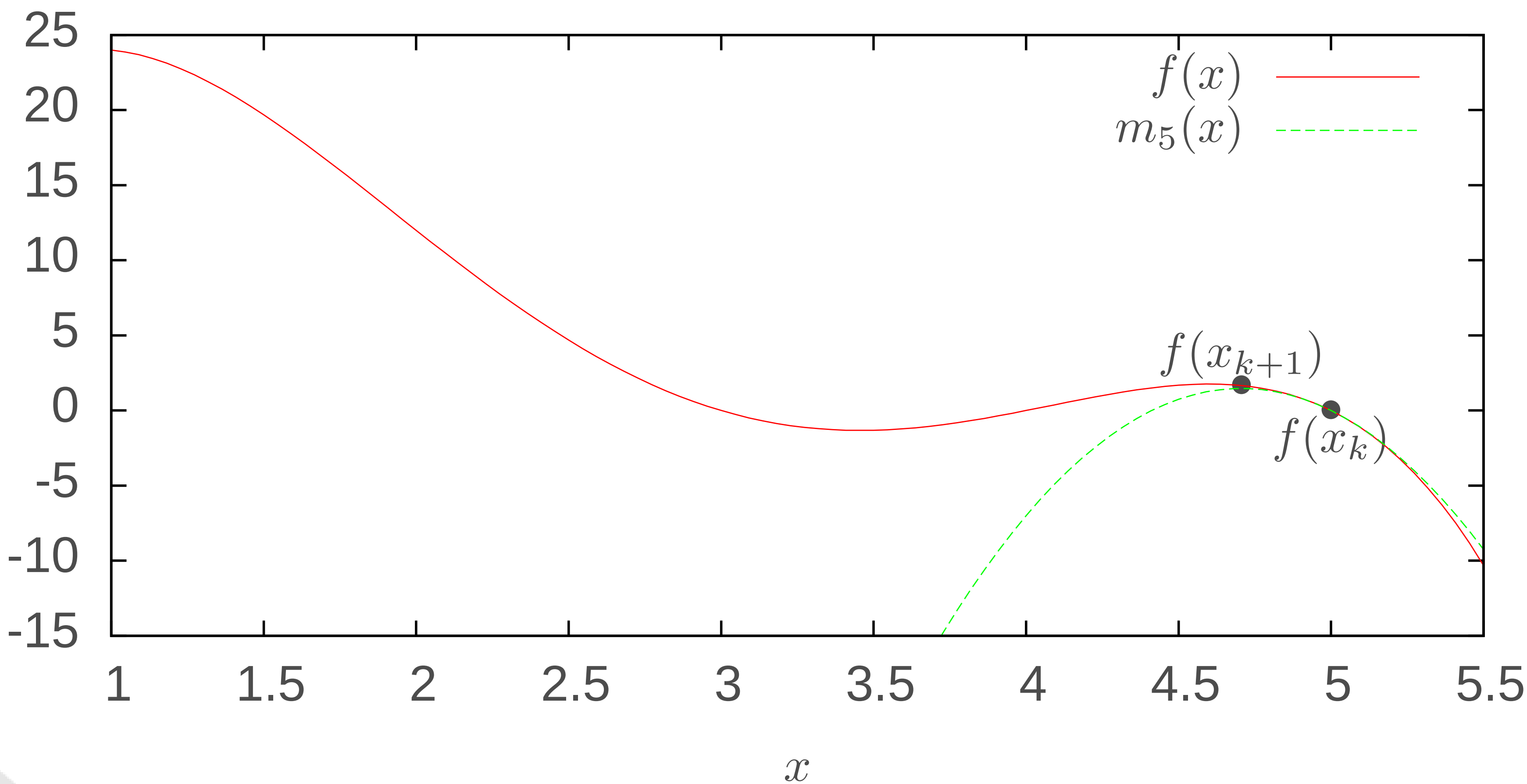
Local Newton: geometric interpretation

$m_4(x) = x^2 - 4x$: bad predictor (zoom)



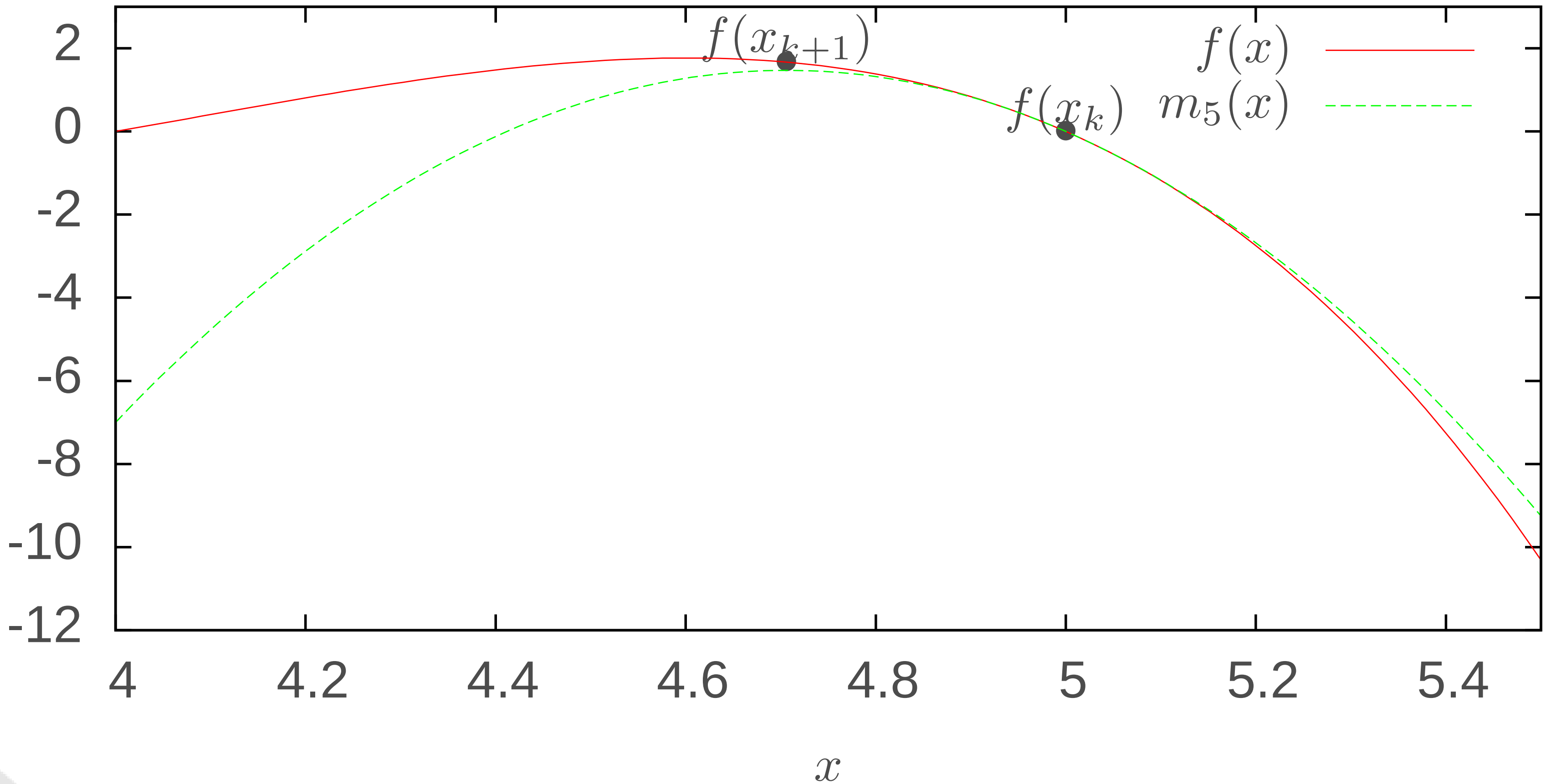
Local Newton: geometric interpretation

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



Local Newton: geometric interpretation

$m_5(x) = -17x^2 + 160x - 375$: concave (zoom)



Descent methods

Typical iteration:

- Find a descent direction d_k such that $\nabla f(x_k)^T d_k < 0$.
- Find a step α_k such that $f(x_k + \alpha_k d_k) < f(x_k)$
- Compute $x_{k+1} = x_k + \alpha_k d_k$.

Descent methods: find a direction

- Basic idea: steepest descent

$$d_k = -\nabla f(x_k)$$

- exhibits slow to very slow convergence
- Solution: precondition (change the metric)

$$d_k = -D_k \nabla f(x_k)$$

where D_k is positive definite.

- Newton:

$$d_k = -(\nabla^2 f(x_k) + \lambda I)^{-1} \nabla f(x_k)$$

where λ is such that $\nabla f(x_k) + \lambda I$ is positive definite

Descent methods: find a step

- Finding the optimal step is not cost effective
- Wolfe's conditions characterize steps guaranteeing convergence
- Use steps that verify these conditions

Wolfe 1: sufficient decrease. Consider

- Function $f : \mathbb{R}^n \to \mathbb{R}$
- Iterate $x_k \in \mathbb{R}^n$
- Direction $d_k \in \mathbb{R}^n$ such that $\nabla f(x_k)^T d_k < 0$
- Step $\alpha_k \in \mathbb{R}$, $\alpha_k > 0$

f decreases sufficiently at $x_k + \alpha_k d_k$ compared to x_k if

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \alpha_k \beta_1 \nabla f(x_k)^T d_k,$$

where $0 < \beta_1 < 1$.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Descent methods: find a step

Wolfe 2: sufficient progress. Consider

- Function $f : \mathbb{R}^n \to \mathbb{R}$
- Iterate $x_k \in \mathbb{R}^n$
- Direction $d_k \in \mathbb{R}^n$ such that $\nabla f(x_k)^T d_k < 0$
- Step $\alpha_k \in \mathbb{R}$, $\alpha_k > 0$

$x_k + \alpha_k d_k$ brings sufficient progress compared to x_k if

$$\nabla f(x_k + \alpha_k d_k)^T d_k \geq \beta_2 \nabla f(x_k)^T d_k,$$

with $0 < \beta_2 < 1$, $\beta_2 > \beta_1$.

Descent methods: find a step

Objective

Find a step α^* such that both Wolfe's conditions are verified

Input

- Function: $f : \mathbb{R}^n \to \mathbb{R}$ continuously differentiable;
- Gradient: $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$;
- Iterate: $x \in \mathbb{R}^n$;
- Descent direction d such that $\nabla f(x)^T d < 0$;
- First approximation $\alpha_0 > 0$;
- Parameters β_1, β_2 such that $0 < \beta_1 < \beta_2 < 1$ (typical example: $\beta_1 = 10^{-4}$ and $\beta_2 = 0.99$);
- Parameter $\lambda > 1$.

Output

A step α^* verifying both Wolfe's conditions.

Descent methods: find a step

Initialization

$i = 0$, $\alpha_\ell = 0$, $\alpha_r = +\infty$.

Iterations

1. If α_i verify both conditions, then $\alpha^* = \alpha_i$. STOP.
2. If α_i violates Wolfe 1, then the step is too long and

$$\begin{aligned} \alpha_r &= \alpha_i \\ \alpha_{i+1} &= \frac{\alpha_\ell + \alpha_r}{2}. \end{aligned}$$

3. If α_i verifies Wolfe 1 and violates Wolfe 2, then the step is too short and

$$\begin{aligned} \alpha_\ell &= \alpha_i \\ \alpha_{i+1} &= \begin{cases} \frac{\alpha_\ell + \alpha_r}{2} & \text{if } \alpha_r < +\infty \\ \lambda \alpha_i & \text{otherwise.} \end{cases} \end{aligned}$$

4. $i = i+1$.

Newton's method with linesearch

Objective

Find (an approximation of) a local minimum of

$$\min_{x \in \mathbb{R}^n} f(x). \quad (3)$$

Input

- Function $f : \mathbb{R}^n \to \mathbb{R}$ continuously differentiable;
- Gradient $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$;
- Hessian $\nabla^2 f : \mathbb{R}^n \to \mathbb{R}^{n \times n}$;
- First approximation $x_0 \in \mathbb{R}^n$;
- Required precision $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$.

Output

An approximation of the solution $x^* \in \mathbb{R}$.

Initialization

$k = 0$.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Newton's method with linesearch

Iterations

1. Compute a lower triangular matrix and a real parameter $\tau \geq 0$ such that

$$L_k L_k^T = \nabla^2 f(x_k) + \tau I,$$

   using a modified Cholesky factorization.
2. Find z_k by solving the triangular system $L_k z_k = \nabla f(x_k)$.
3. Find d_k by solving the triangular system $L_k^T d_k = -z_k$.
4. Find α_k with line search starting with $\alpha_0 = 1$.
5. $x_{k+1} = x_k + \alpha_k d_k$.
6. $k = k + 1$.

Stopping criterion

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

Quasi-Newton

Ideas:

- Adapt Broyden's (secant) method to optimization
- Additional constraint: the approximated matrix must be
  - symmetric
  - positive definite
- Update formula: BFGS (C. G. Broyden, R. Fletcher, D. Goldfarb and D. F. Shanno

Quasi-Newton

Objective

Find (an approximation of) a local minimum of

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

Input

- Function $f : \mathbb{R}^n \to \mathbb{R}$
- Gradient $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$;
- First approximation of the solution $x_0 \in \mathbb{R}^n$;
- First approximation of the inverse of the hessian $H_0^{-1} \in \mathbb{R}^{n \times n}$ symmetric positive definite. Typically, $H_0^{-1} = I$.
- Required precision: $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$.

Output

An approximation of the solution $x^* \in \mathbb{R}$.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Quasi-Newton

Initialization

$k = 0.$

Iterations

1. Compute $d_k = -H_k^{-1} \nabla f(x_k)$.
2. Find α_k with linesearch starting with $\alpha_0 = 1$.
3. $x_{k+1} = x_k + \alpha_k d_k$.
4. $k = k + 1$.
5. Update H_k^{-1}

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

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Quasi-Newton

Stopping criterion

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

Summary

- Solving systems of equations: $\nabla f(x) = 0$
  - Newton
  - Quasi-Newton
- Unconstrained optimization
  - Quadratic problems
  - Local Newton
  - Linesearch
  - Quasi-Newton