MVE165/MMG630, Applied Optimization
Lecture 11
Unconstrained nonlinear programming

Ann-Brith Strömberg

2010–04–22

An overview of nonlinear optimization

General notation of nonlinear programs

$$\begin{aligned} \text{minimize}_{\mathbf{x}\in\Re^n} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \le 0, \quad i = 1, \ldots, m. \end{aligned}$$

Some special cases

- Unconstrained problems ($m = 0$):
  $$\text{minimize } f(\mathbf{x}) \text{ subject to } \mathbf{x} \in \Re^n$$
- Convex programming: f convex, g_i convex, $i = 1, \ldots, m$
- Linear constraints: $g_i(\mathbf{x}) = \mathbf{a}_i^T\mathbf{x} - b_i, \quad i = 1, \ldots, m$
  - Quadratic programming: $f(\mathbf{x}) = \mathbf{c}^T\mathbf{x} + \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x}$
  - Linear programming: $f(\mathbf{x}) = \mathbf{c}^T\mathbf{x}$

Areas of applications, examples (Ch. 9.1)

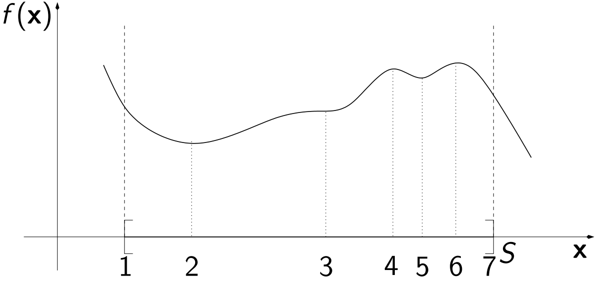
- Structural optimization
  - Design of aircraft, ships, bridges, etc
  - Decide on the material and the thickness of a mechanical structure
  - Minimize weight, maximize stiffness, constraints on deformation at certain loads, strength, etc
- Analysis and design of traffic networks
  - Estimate traffic flows and discharges
  - Detect bottlenecks
  - Analyze effects of traffic signals, tolls, etc
- Least squares—adaptation of data
- Engine development, design of antennas, …
  for each function evaluation a simulation may be needed
- Maximize the volume of a cylinder
  while keeping the surface area constant
- Assessment of cutting patterns
- …

Properties of nonlinear programs

- The mathematical properties of nonlinear optimization problems can be very different
- No algorithm exists that solves all nonlinear optimization problems
- An optimal solution does *not* have to be located at an extreme point
- Nonlinear programs can be unconstrained (what if a linear program has no constraints?)
- In this course: We study models with f differentiable (which is not always the case)
- For **convex** problems: Algorithms converge to an optimal solution
- Nonlinear problems can have local optima that are not global optima

Possible extremal points for

minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$



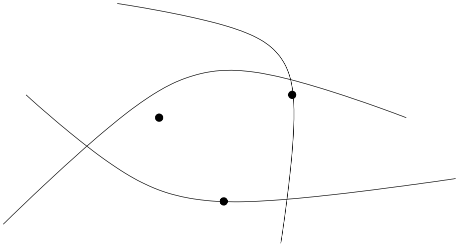
- boundary points of S
- stationary points, where $f'(\mathbf{x}) = 0$
- discontinuities in f or f' DRAW!

Boundary and stationary points (Ch. 10.0)

- $\bar{\mathbf{x}}$ is a *boundary* point to the feasible set

$$S = \{\mathbf{x} \in \Re^n \mid g_i(\mathbf{x}) \leq 0, i = 1, \ldots, m\}$$

if $g_i(\bar{\mathbf{x}}) \leq 0$, $i = 1, \ldots, m$, and $g_i(\bar{\mathbf{x}}) = 0$ for at least one index i



- $\bar{\mathbf{x}}$ is a *stationary* point to f if $\nabla f(\mathbf{x}) = \mathbf{0}$
(in one dimension: if $f'(x) = 0$)

Local and global minima (maxima) (Ch. 2.4)

$$\text{minimize } f(\mathbf{x}) \text{ subject to } \mathbf{x} \in S$$

- $\bar{\mathbf{x}}$ is a local minimum if $\bar{\mathbf{x}} \in S$ and $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$ for all $\mathbf{x} \in S$ sufficiently close to $\bar{\mathbf{x}}$
  - In words: A solution is a *local* minimum if it is *feasible* and no other feasible solution in a sufficiently *small neighbourhood* has a lower objective value
  - Formally: $\exists \varepsilon > 0$ such that $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$ for all $\mathbf{x} \in S \cap \{\mathbf{x} \in \Re^n : \|\mathbf{x} - \bar{\mathbf{x}}\| \leq \varepsilon\}$
  - DRAW!!
- $\bar{\mathbf{x}}$ is a global minimum if $\bar{\mathbf{x}} \in S$ and $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$ for all $\mathbf{x} \in S$
  - In words: A solution is a *global* minimum if it is *feasible* and no other feasible solution has a lower objective value

Unconstrained optimization

$$\text{minimize } f(\mathbf{x}) \text{ subject to } \mathbf{x} \in \Re^n$$

- Assume that $f : \Re^n \mapsto \Re$ is continuously differentiable on $\Re^n$
- **Necessary conditions** for a local optimum:

  $\overline{\mathbf{x}}$ is a local minimum/maximum for $f \Rightarrow \nabla f(\overline{\mathbf{x}}) = \mathbf{0}$
- This is not sufficient, since $\nabla f(\tilde{\mathbf{x}}) = \mathbf{0}$ also when $\tilde{\mathbf{x}}$ is a saddle point, e.g.
- If f is twice continuously differentiable on $\Re^n$ then the Hessian matrix exists: $H_f(\mathbf{x}) = \nabla^2 f(\mathbf{x})$
- **Sufficient conditions** for a local optimum:

$$\left.\begin{array}{l} \nabla f(\overline{\mathbf{x}}) = \mathbf{0} \\ H_f(\overline{\mathbf{x}}) \text{ pos/neg definite} \end{array}\right\} \Rightarrow \overline{\mathbf{x}} \text{ is a local min/max for } f$$

When is a local optimum also a global optimum? (Ch. 9.3)

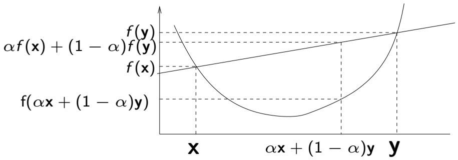
- The concept of **convexity** is essential
- Functions: convex (minimization), concave (maximization)
- Sets: convex (minimization and maximization)
- The minimization (maximization) of a convex (concave) function over a convex set is referred to as a convex optimization problem
- How conclude whether sets and functions are convex, concave, or neither?

Convex functions

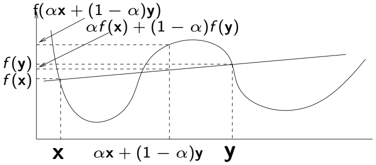
- A function f is *convex* on S if, for any $\mathbf{x}, \mathbf{y} \in S$ it holds that

$$f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y}) \text{ for all } 0 \leq \alpha \leq 1$$

A CONVEX FUNCTION



A NON-CONVEX FUNCTION



- f is *strictly convex* on S if, for any $\mathbf{x}, \mathbf{y} \in S$ such that $\mathbf{x} \neq \mathbf{y}$ it holds that

$$f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) < \alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y}) \text{ for all } 0 < \alpha < 1$$

Convex/concave functions

- f is (strictly) *concave* on S if $-f$ is (strictly) *convex* on S
- f is convex $\Leftrightarrow$ H_f is positive semi-definite
- H_f is positive definite $\Rightarrow$ f is strictly convex
- Definition: The quadratic matrix H is positive definite (semi-definite) if $d^T H d > 0$ for all $d \neq 0$ ($d^T H d \geq 0\ \forall d$)
- Example: Check convexity for $f(\mathbf{x}) = 2x^2 - 2xy + y^2 + 3x - y$
- $\nabla f(\mathbf{x}) = \begin{pmatrix} 4x - 2y + 3 \\ -2x + 2y - 1 \end{pmatrix} \quad H_f(\mathbf{x}) = \begin{pmatrix} 4 & -2 \\ -2 & 2 \end{pmatrix}$
- Eigenvalues for $H_f(\mathbf{x})$: $\det(H_f(\mathbf{x}) - \lambda I) = 0 \Leftrightarrow$
  $\begin{vmatrix} 4-\lambda & -2 \\ -2 & 2-\lambda \end{vmatrix} = (4-\lambda)(2-\lambda) - 4 = 0 \Leftrightarrow$
  $\lambda^2 - 6\lambda + 4 = 0 \Rightarrow \lambda_1 = 3 + \sqrt{5} > 0,\ \lambda_2 = 3 - \sqrt{5} > 0 \Rightarrow$
  $H_f(\mathbf{x})$ is positive definite $\Rightarrow$ f is strictly convex

Convex functions – Examples

- Check (strict?) convexity of the function $f(x,y) = x^3 + y^3$ on $\Re^2$ (on $\Re^2_+$)

- Check whether (where) the function $f(x,y) = \ln x - y^2 + cxy$ is convex, concave, or neither (assume that the constant $c > 0$)

Convex functions

- (Th. 9.4) A non-negative linear combination of convex functions is convex:

$$\left.\begin{array}{ll} f_i \text{ convex}, & i = 1, \ldots, m \\ \lambda_i \geq 0, & i = 1, \ldots, m \end{array}\right\} \Rightarrow f = \sum_{i=1}^{m} \lambda_i f_i \text{ is convex}$$

- Draw!!

- The pointwise maximum of convex functions is convex:

$$f_i(\mathbf{x}),\ i = 1, \ldots, m, \text{ convex} \quad \Rightarrow \quad f(\mathbf{x}) = \max_{i=1,\ldots,m} f_i(\mathbf{x}) \text{ convex}$$

- Draw!!

More about convex functions

- If $g : \Re \mapsto \Re$ is convex and non-decreasing and $h : \Re^n \mapsto \Re$ is convex, then the composite function $f = g(h) : \Re^n \mapsto \Re$ is convex

- Example: $g(y) = y \ln y$, $h(\mathbf{x}) = x_1^2 + x_2^2$
  - $g'(y) = 1 + \ln y > 0$ for $y > e$ ($\Rightarrow g$ nondecreasing),
  - $g''(y) = \frac{1}{y} > 0$ for $y > 0$ ($\Rightarrow g$ convex)
  - $\nabla h(\mathbf{x}) = (2x_1, 2x_2)^{\mathrm{T}}$, $H_h(\mathbf{x}) = \nabla^2 h(\mathbf{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ ($\Rightarrow h$ convex)

  $\Rightarrow$ $f(\mathbf{x}) = g(h(\mathbf{x})) = (x_1^2 + x_2^2)\ln(x_1^2 + x_2^2)$ is convex for $\mathbf{x} \in \Re^2$ such that $x_1^2 + x_2^2 > e$

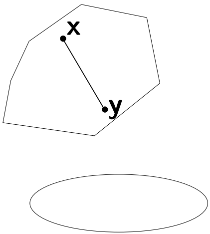
Convex sets

- A set S is convex if, for any elements $\mathbf{x}, \mathbf{y} \in S$ it holds that

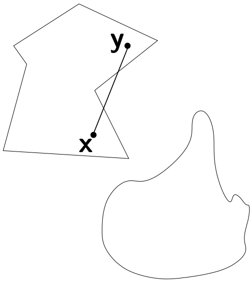
$$\alpha \mathbf{x} + (1 - \alpha)\mathbf{y} \in S \text{ for all } 0 \leq \alpha \leq 1$$

- Examples:

Convex sets



Non-convex sets



Convex sets

- Consider a set S defined by the intersection of m inequalities:
  $$S = \{ \mathbf{x} \in \Re^n \mid g_i(\mathbf{x}) \leq 0, \;\; i = 1, \dots, m \}$$
  where the functions $g_i : \Re^n \mapsto \Re$
- (Th. 9.2 & 9.3) If all the functions $g_i(\mathbf{x})$ $i = 1, \dots, m$, are convex on $\Re^n$, then S is a convex set
- Example:
  $g_1(\mathbf{x}) = x_1^2 + 3x_2^2 - 1$, $g_2(\mathbf{x}) = x_1 + x_2$, $g_3(\mathbf{x}) = x_1^2 - x_2$
  $S = \left\{ \mathbf{x} \in \Re^2 \mid g_i(\mathbf{x}) \leq 0, \;\; i = 1, 2, 3 \right\} \Rightarrow$
  $H_{g_1}(\mathbf{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 6 \end{pmatrix} \Rightarrow g_1$ strictly convex,
  $H_{g_2}(\mathbf{x}) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \Rightarrow g_2$ convex (& concave!),
  $H_{g_3}(\mathbf{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \Rightarrow g_3$ convex
  $\Rightarrow$ The set S is convex DRAW!!

Global optima of convex programs

- (Def. 9.5) If f and g_i, $i = 1, \ldots, m$, are convex functions, then
$$\text{minimize } f(\mathbf{x}) \text{ subject to } g_i(\mathbf{x}) \leq 0, \ \ i = 1, \ldots, m$$
is said to be a *convex* optimization problem
- (Th. 9.1) Let $\mathbf{x}^*$ be a *local* optimum for a convex optimization problem. Then $\mathbf{x}^*$ is also a *global* optimum
- If f is strictly convex and g_i, $i = 1, \ldots, m$, are convex, then there exists *at most* one optimal solution (a unique global optimum)
- (Th. 10.2) Necessary and sufficient condition for optimality in *unconstrained* minimization (maximization):
Suppose that $f : \Re^n \mapsto \Re$ is convex (concave) and continuously differentiable on $\Re^n$. A point $\mathbf{x}^* \in \Re^n$ is a global minimum for f if and only if $\nabla f(\mathbf{x}^*) = \mathbf{0}$

Solution methods for unconstrained optimization (Ch. 2.5.1)

- General iterative search method:
  1. Choose a starting solution, $\mathbf{x}^0 \in \Re^n$. Let $k = 0$
  2. Determine a search direction $\mathbf{d}^k$
  3. If a termination criterion is fulfilled $\Rightarrow$ Stop!
  4. Determine a step length, t_k, by solving:

$$\text{minimize}_{\,t \geq 0} \varphi(t) := f(\mathbf{x}^k + t \cdot \mathbf{d}^k)$$

  5. New iteration point, $\mathbf{x}^{k+1} = \mathbf{x}^k + t_k \cdot \mathbf{d}^k$
  6. Let $k := k + 1$ and return to step 2

- How choose *search directions* $\mathbf{d}^k$, *step lengths* t_k, and *termination criteria*?

Improving search directions (Ch. 10)

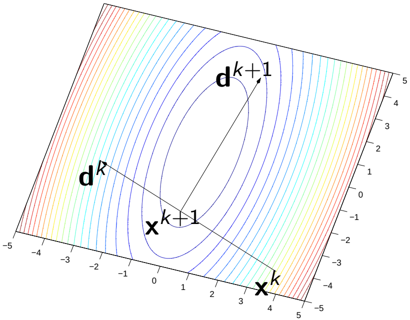
- Goal: $f(\mathbf{x}^{k+1}) < f(\mathbf{x}^k)$ (minimization)
- How does f change locally in a direction $\mathbf{d}^k$ at $\mathbf{x}^k$?
- Taylor expansion (Ch. 9.2):
  $f(\mathbf{x}^k + t\mathbf{d}^k) = f(\mathbf{x}^k) + t\nabla f(\mathbf{x}^k)^{\mathrm{T}}\mathbf{d}^k + \mathcal{O}(t^2)$
- For sufficiently small $t > 0$:
  $f(\mathbf{x}^k + t\mathbf{d}^k) < f(\mathbf{x}^k) \Rightarrow \nabla f(\mathbf{x}^k)^{\mathrm{T}}\mathbf{d}^k < 0$

$\Rightarrow$ **Definition:**
If $\nabla f(\mathbf{x}^k)^{\mathrm{T}}\mathbf{d}^k < 0$ then $\mathbf{d}^k$ is a descent direction for f at $\mathbf{x}^k$
If $\nabla f(\mathbf{x}^k)^{\mathrm{T}}\mathbf{d}^k > 0$ then $\mathbf{d}^k$ is an ascent direction for f at $\mathbf{x}^k$

- We wish to minimize (maximize) f over $\Re^n$:

$\Rightarrow$ Choose $\mathbf{d}^k$ as a descent (an ascent) direction from $\mathbf{x}^k$

An improving step



Figur: At $\mathbf{x}^k$, the descent direction $\mathbf{d}^k$ is generated. A step t_k is taken in this direction, producing $\mathbf{x}^{k+1}$. At this point, a new descent direction $\mathbf{d}^{k+1}$ is generated, and so on.

Solution methods for unconstrained optimization (Ch. 2.5.1)

- General iterative search method:

  1. Choose a starting solution, $\mathbf{x}^0 \in \Re^n$. Let $k = 0$
  2. Determine a search direction $\mathbf{d}^k$
  3. If a termination criterion is fulfilled $\Rightarrow$ Stop!
  4. Determine a step length, t_k, by solving:

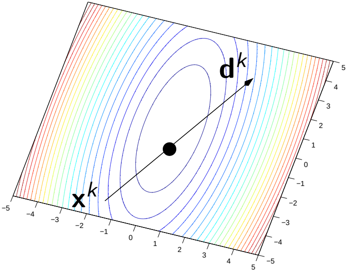
  $$\text{minimize}_{t \geq 0} \varphi(t) := f(\mathbf{x}^k + t \cdot \mathbf{d}^k)$$

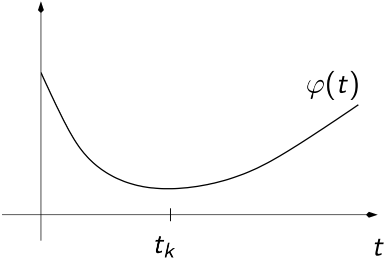
  5. New iteration point, $\mathbf{x}^{k+1} = \mathbf{x}^k + t_k \cdot \mathbf{d}^k$
  6. Let $k := k + 1$ and return to step 2

Step length—line search (minimization) (Ch. 10.4)

- Solve $\min_{t \geq 0} \varphi(t) := f(\mathbf{x}^k + t \cdot \mathbf{d}^k)$ where $\mathbf{d}^k$ is a descent direction from $\mathbf{x}^k$
- A minimization problem in one variable $\Rightarrow$ Solution t_k
- Analytic solution: $\varphi'(t_k) = 0$ (seldom possible to derive)
- Numerical solution methods:
  - The golden section method (reduce the interval of uncertainty)
  - The bi-section method (reduce the interval of uncertainty)
  - Newton-Raphson's method
  - Armijo's method
- In practice: Do not solve exactly, but to a sufficient improvement of the function value: $f(\mathbf{x}^k + t_k \mathbf{d}^k) \leq f(\mathbf{x}^k) - \varepsilon$ for some $\varepsilon > 0$

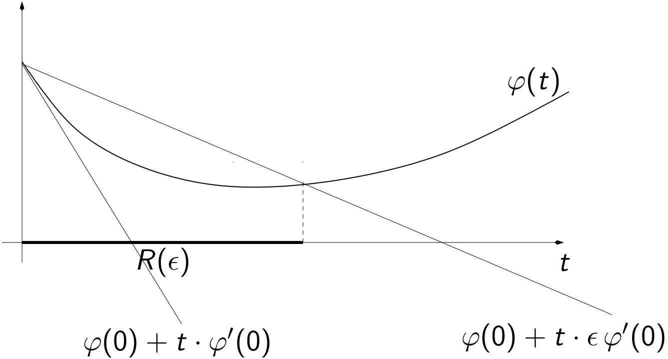
Line search





Figur: A line search in a descent direction.
t_k solves $\min_{t \geq 0} \varphi(t) := f(\mathbf{x}^k + t \cdot \mathbf{d}^k)$

Line search—the Armijo step length rule

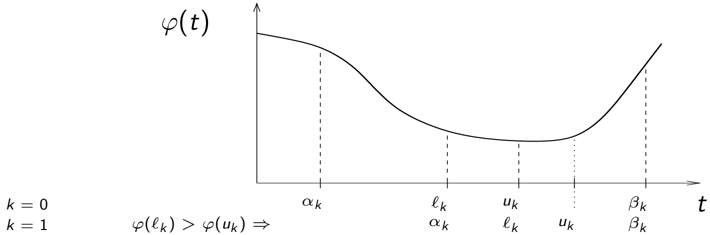


Figur: The interval $R(\epsilon)$ accepted by the Armijo step length rule.
$\epsilon =$ the fraction of decrease required, $0 < \epsilon < 1$
$R(\epsilon) = \{\, t \geq 0 \mid \varphi(t) \leq \varphi(0) + t \cdot \epsilon\, \varphi'(0) \,\}$ Note that $\varphi'(0) < 0$

Line search—the Golden section method

Based on decreasing an interval containing t^* (the function may contain no more than one local minimum in the interval considered)

1. Let α_k and β_k be lower and upper bound on t^*: $\alpha_k \leq t^* \leq \beta_k$
2. Choose $\ell_k = \beta_k - \gamma(\beta_k - \alpha_k)$, $u_k = \alpha_k + \gamma(\beta_k - \alpha_k)$ where $\gamma \approx 0.618$ (the (inverted) golden ratio)
3. Evaluate $\varphi(\ell_k)$, $\varphi(u_k)$ and replace α_k or β_k by ℓ_k or u_k
4. Terminate or let $k := k + 1$ and return to 2.



$\Rightarrow$ whichever of $[\alpha_k, u_k]$ or $[\ell_k, \beta_k]$ provides the next interval, its size will be γ times the current size

Solution methods for unconstrained optimization

- General iterative search method:

  1. Choose a starting solution, $\mathbf{x}^0 \in \Re^n$. Let $k = 0$
  2. Determine a search direction $\mathbf{d}^k$
  3. If a termination criterion is fulfilled $\Rightarrow$ Stop!
  4. Determine a step length, t_k, by solving:

$$\text{minimize }_{t \geq 0} \varphi(t) := f(\mathbf{x}^k + t \cdot \mathbf{d}^k)$$

  5. New iteration point, $\mathbf{x}^{k+1} = \mathbf{x}^k + t_k \cdot \mathbf{d}^k$
  6. Let $k := k + 1$ and return to step 2

Termination criteria

- Needed since $\nabla f(\mathbf{x}^k) = \mathbf{0}$ will never be fulfilled exactly
- Typical choices ($\varepsilon_j > 0$, $j = 1, \ldots, 4$)
  (a) $\|\nabla f(\mathbf{x}^k)\| < \varepsilon_1$
  (b) $|f(\mathbf{x}^{k+1}) - f(\mathbf{x}^k)| < \varepsilon_2$
  (c) $\|\mathbf{x}^{k+1} - \mathbf{x}^k\| < \varepsilon_3$
  (d) $t_k < \varepsilon_4$

  These are often combined
- The search method only guarantees a stationary solution, whose properties are determined by the properties of f (convexity, ...)