

MVE165/MMG631

Linear and integer optimization with applications

Lecture 14

Overview of nonlinear programming
and summary of the course

Ann-Brith Strömberg

2018-05-15

Structural optimization

- Design of aircraft, ships, bridges, etc
- Decide on the material and the topology and thickness of a mechanical structure
- Minimize weight, maximize stiffness, constraints on deformation at certain loads, strength, fatigue limit, 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 or tyres, etc.

For each function evaluation a computationally expensive (time consuming) simulation may be needed

Maximize the volume of a cylinder

... while keeping, e.g., the surface area constant

Wind power generation

The energy content in the wind is $\propto v^3$ (in Ass 3a it is discretized and measured data is used)

An overview of nonlinear optimization

General notation for nonlinear programs

$$\begin{array}{ll} \text{minimize } \mathbf{x} \in \mathbb{R}^n & f(\mathbf{x}) \\ \text{subject to} & g_i(\mathbf{x}) \leq 0, \quad i \in \mathcal{L}, \\ & h_i(\mathbf{x}) = 0, \quad i \in \mathcal{E}. \end{array}$$

Some special cases

- Unconstrained problems ($\mathcal{L} = \mathcal{E} = \emptyset$):

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

- Convex programming: f convex, g_i convex, $i \in \mathcal{L}$, h_i linear, $i \in \mathcal{E}$.
- Linear constraints: g_i , $i \in \mathcal{L}$, and h_i , $i \in \mathcal{E}$

- Quadratic programming:

$$\boxed{f(\mathbf{x}) = \mathbf{c}^\top \mathbf{x} + \frac{1}{2} \mathbf{x}^\top \mathbf{Q} \mathbf{x}}$$

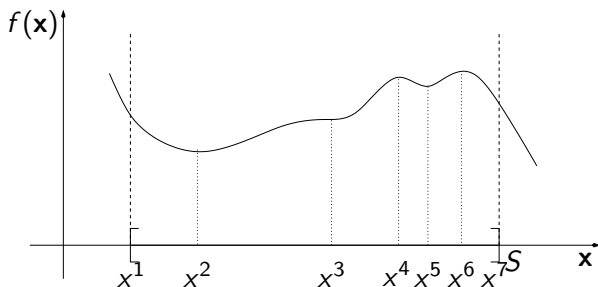
- Linear programming:

$$\boxed{f(\mathbf{x}) = \mathbf{c}^\top \mathbf{x}}$$

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?
- f may be differentiable or non-differentiable
E.g., the Lagrangean dual objective function
- *For convex problems*: Algorithms (typically) converge to an optimal solution
- Nonlinear problems can have *local* optima that are *not global* optima

Consider the problem to minimize $f(x)$ subject to $x \in S$



Possible extremal points are

- boundary points of $S = [x^1, x^7]$ (i.e., $\{x^1, x^7\}$)
- stationary points, where $f'(x) = 0$ (i.e., $\{x^2, \dots, x^6\}$)
- discontinuities in f or f'

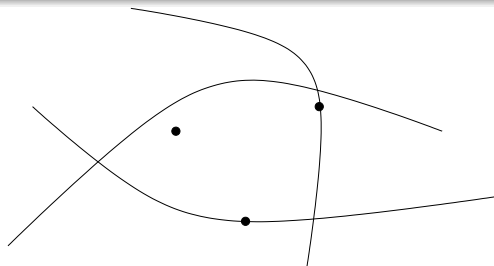
DRAW!

Boundary points

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

$$S = \{\mathbf{x} \in \mathbb{R}^n \mid g_i(\mathbf{x}) \leq 0, i \in \mathcal{L}\}$$

if $g_i(\bar{\mathbf{x}}) \leq 0, i \in \mathcal{L}$, and $g_i(\bar{\mathbf{x}}) = 0$ for at least one index $i \in \mathcal{L}$



Stationary points

$\bar{\mathbf{x}}$ is a *stationary* point to f if $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}^n$ (for $n = 1$: if $f'(\bar{\mathbf{x}}) = 0$)

Consider the nonlinear optimization problem to

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

Local minimum

- *In words:* A solution is a *local* minimum if it is *feasible* and no other feasible solution in a sufficiently *small neighbourhood* of the solution at hand has a lower objective value
- *Formally:* $\bar{\mathbf{x}}$ is a local minimum if $\bar{\mathbf{x}} \in S$ and $\exists \epsilon > 0$ such that $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$ for all $\mathbf{x} \in \{\mathbf{y} \in S : \|\mathbf{y} - \bar{\mathbf{x}}\| \leq \epsilon\}$ DRAW!!

Global minimum

- *In words:* A solution is a *global* minimum if it is *feasible* and no other feasible solution has a lower objective value
- *Formally:* $\bar{\mathbf{x}}$ is a global minimum if $\bar{\mathbf{x}} \in S$ and $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$ for all $\mathbf{x} \in S$

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

Definition 9.5: Convex optimization problem

If f and g_i , $i \in \mathcal{L}$, are convex functions, then

$$\text{minimize } f(\mathbf{x}) \text{ subject to } g_i(\mathbf{x}) \leq 0, i \in \mathcal{L}$$

is said to be a *convex* optimization problem

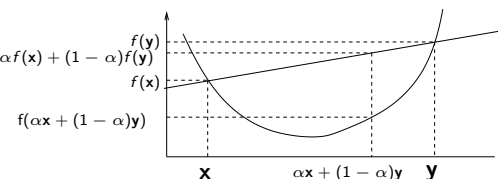
Theorem 9.1: Global optimum

Let \mathbf{x}^* be a *local* optimum of a *convex* optimization problem. Then \mathbf{x}^* is also a *global* optimum

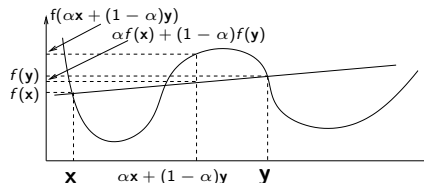
Convex functions

A function f is *convex* on S if, for any $\mathbf{x}, \mathbf{y} \in S$ it holds that $f(\alpha\mathbf{x} + (1 - \alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1 - \alpha)f(\mathbf{y})$ for all $0 \leq \alpha \leq 1$

A CONVEX FUNCTION



A NON-CONVEX FUNCTION



The 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

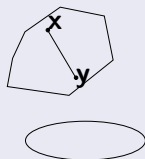
$$f(\alpha\mathbf{x} + [1 - \alpha]\mathbf{y}) < \alpha f(\mathbf{x}) + (1 - \alpha)f(\mathbf{y}) \text{ for all } 0 < \alpha < 1$$

Convex sets

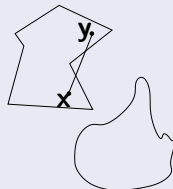
A set S is convex if, for any $\mathbf{x}, \mathbf{y} \in S$ it holds that $\alpha\mathbf{x} + (1 - \alpha)\mathbf{y} \in S$ for all $0 \leq \alpha \leq 1$

Examples

Convex sets



Non-convex sets



Consider a set S defined by the intersection of $m = |\mathcal{L}|$ inequalities, where the functions $g_i : \mathbb{R}^n \mapsto \mathbb{R}$, $i \in \mathcal{L}$, as

$$S = \{ \mathbf{x} \in \mathbb{R}^n \mid g_i(\mathbf{x}) \leq 0, i \in \mathcal{L} \}$$

Theorems 9.2 & 9.3

If all the functions g_i , $i \in \mathcal{L}$, are convex on \mathbb{R}^n , then the set S is convex

The Karush-Kuhn-Tucker conditions: necessary conditions for optimality

Let $S := \{ \mathbf{x} \in \mathbb{R}^n \mid g_i(\mathbf{x}) \leq 0, i \in \mathcal{L} \}$

- Assume that the following hold
 - the function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is differentiable;
 - the functions $g_i : \mathbb{R}^n \mapsto \mathbb{R}$, $i \in \mathcal{L}$, are convex and differentiable;
 - there exists a point $\bar{\mathbf{x}} \in S$ such that $g_i(\bar{\mathbf{x}}) < 0$, $i \in \mathcal{L}$
- If $\mathbf{x}^* \in S$ is a local minimum of f over S , then there exists a vector $\boldsymbol{\mu} \in \mathbb{R}^m$ (where $m = |\mathcal{L}|$) such that

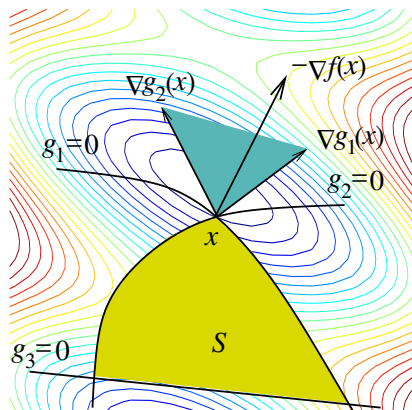
$$\nabla f(\mathbf{x}^*) + \sum_{i \in \mathcal{L}} \mu_i \nabla g_i(\mathbf{x}^*) = \mathbf{0}^n,$$

$$\mu_i g_i(\mathbf{x}^*) = 0, \quad i \in \mathcal{L},$$

$$\boldsymbol{\mu} \geq \mathbf{0}^m.$$

□

Geometry of the Karush-Kuhn-Tucker conditions



Figur: Geometric interpretation of the Karush-Kuhn-Tucker conditions. At a local minimum, the negative gradient of the objective function can be expressed as a non-negative linear combination of the gradients of the active constraints at this point

The Karush-Kuhn-Tucker conditions: sufficient for optimality under convexity

Assume that the functions $f, g_i : \mathbb{R}^n \mapsto \mathbb{R}$, $i \in \mathcal{L}$, are convex and differentiable, and let $S = \{ \mathbf{x} \in \mathbb{R}^n \mid g_i(\mathbf{x}) \leq 0, i \in \mathcal{L} \}$

If the conditions (where $m = |\mathcal{L}|$)

$$\nabla f(\mathbf{x}^*) + \sum_{i \in \mathcal{L}} \mu_i \nabla g_i(\mathbf{x}^*) = \mathbf{0}^n,$$

$$\mu_i g_i(\mathbf{x}^*) = 0, \quad i \in \mathcal{L},$$

$$\boldsymbol{\mu} \geq \mathbf{0}^m$$

hold, then $\mathbf{x}^* \in S$ is a global minimum of f over S . □

- The Karush-Kuhn-Tucker conditions can also be stated for optimization problems with equality constraints
- For unconstrained optimization KKT reads: $\nabla f(\mathbf{x}^*) = \mathbf{0}$
- For a quadratic program KKT forms a system of linear (in)equalities plus the complementarity constraints

The optimality conditions can be used to..

- verify an (local) optimal solution
- solve certain special cases of nonlinear programs (e.g. quadratic programs)
- algorithm construction
- derive properties of a solution to a non-linear program

Example

$$\begin{aligned} &\text{minimize} && f(\mathbf{x}) := 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2 \\ &\text{subject to} && x_1^2 + x_2^2 \leq 5 \\ &&& 3x_1 + x_2 \leq 6 \end{aligned}$$

Is $\mathbf{x}^0 = (1, 2)^\top$ a Karush-Kuhn-Tucker point?

- Is it an optimal solution?
- Derive: $\nabla f(\mathbf{x}) = (4x_1 + 2x_2 - 10, 2x_1 + 2x_2 - 10)^\top$,
 $\nabla g_1(\mathbf{x}) = (2x_1, 2x_2)^\top$, and $\nabla g_2(\mathbf{x}) = (3, 1)^\top$

$$\begin{aligned} 4x_1^0 + 2x_2^0 - 10 + 2x_1^0\mu_1 + 3\mu_2 &= 0 \\ 2x_1^0 + 2x_2^0 - 10 + 2x_2^0\mu_1 + \mu_2 &= 0 \\ \mu_1[(x_1^0)^2 + (x_2^0)^2 - 5] + \mu_2(3x_1^0 + x_2^0 - 6) &= 0 \\ \mu_1, \mu_2 &\geq 0 \end{aligned}$$

\iff

$$\begin{aligned} 2\mu_1 + 3\mu_2 &= 2 \\ 4\mu_1 + \mu_2 &= 4 \\ 0\mu_1 - \mu_2 &= 0 \\ \mu_1, \mu_2 &\geq 0 \end{aligned}$$

$$\Rightarrow \mu_2 = 0 \quad \Rightarrow \quad \mu_1 = 1 \geq 0$$

Example, continued

OK, the Karush-Kuhn-Tucker conditions hold

Is the solution optimal? Check convexity!

- $\nabla^2 f(\mathbf{x}) = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}$, $\nabla^2 g_1(\mathbf{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$, $\nabla^2 g_2(\mathbf{x}) = \mathbf{0}^{2 \times 2}$

$\Rightarrow f$, g_1 , and g_2 are convex

$\Rightarrow \mathbf{x}^0 = (1, 2)^\top$ is an optimal solution and $f(\mathbf{x}^0) = -20$

- 1 Choose a starting solution, $\mathbf{x}^0 \in \mathbb{R}^n$. 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**?

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):
$$f(\mathbf{x}^k + t\mathbf{d}^k) = f(\mathbf{x}^k) + t\nabla f(\mathbf{x}^k)^\top \mathbf{d}^k + \mathcal{O}(t^2)$$
- For sufficiently small $t > 0$:
$$f(\mathbf{x}^k + t\mathbf{d}^k) < f(\mathbf{x}^k) \quad \Rightarrow \quad \nabla f(\mathbf{x}^k)^\top \mathbf{d}^k < 0$$

\Rightarrow

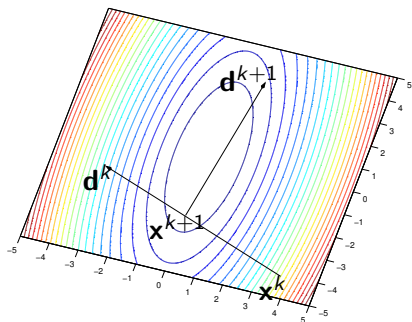
Definition

If $\nabla f(\mathbf{x}^k)^\top \mathbf{d}^k < 0$ then \mathbf{d}^k is a descent direction for f at \mathbf{x}^k
If $\nabla f(\mathbf{x}^k)^\top \mathbf{d}^k > 0$ then \mathbf{d}^k is an ascent direction for f at \mathbf{x}^k

We wish to minimize (maximize) f over \mathbb{R}^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, etc

- 1 Choose a starting solution, $\mathbf{x}^0 \in \mathbb{R}^n$. 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

- 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)

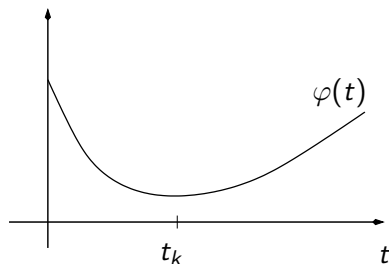
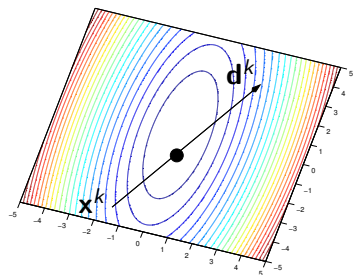
Numerical solution methods

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

General iterative search method for unconstrained optimization

- 1 Choose a starting solution, $\mathbf{x}^0 \in \mathbb{R}^n$. 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 ($\varepsilon_j > 0, j = 1, \dots, 4$)

- (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) $t_k < \varepsilon_4$

The criteria (a)–(d) are often combined

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

Constrained optimization: Penalty methods

Consider both inequality and equality constraints

$$\begin{aligned} & \text{minimize}_{\mathbf{x} \in \mathbb{R}^n} && f(\mathbf{x}) \\ & \text{subject to} && g_i(\mathbf{x}) \leq 0, \quad i \in \mathcal{L}, \\ & && h_i(\mathbf{x}) = 0, \quad i \in \mathcal{E}. \end{aligned} \tag{1}$$

Drop the constraints and add terms in the objective that *penalize infeasible solutions*

$$\text{minimize}_{\mathbf{x} \in \mathbb{R}^n} F_\mu(\mathbf{x}) := f(\mathbf{x}) + \mu \sum_{i \in \mathcal{L} \cup \mathcal{E}} \alpha_i(\mathbf{x}) \tag{2}$$

where $\mu > 0$ and $\alpha_i(\mathbf{x}) = \begin{cases} = 0 & \text{if } \mathbf{x} \text{ satisfies constraint } i \\ > 0 & \text{otherwise} \end{cases}$

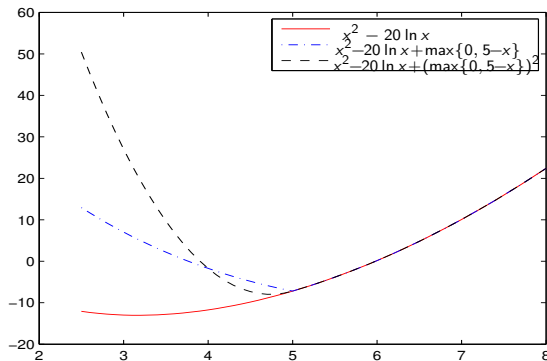
Common penalty functions (which of these are differentiable?)

$$i \in \mathcal{L}: \quad \alpha_i(\mathbf{x}) = \max\{0, g_i(\mathbf{x})\} \quad \text{or} \quad \alpha_i(\mathbf{x}) = (\max\{0, g_i(\mathbf{x})\})^2$$

$$i \in \mathcal{E}: \quad \alpha_i(\mathbf{x}) = |h_i(\mathbf{x})| \quad \text{or} \quad \alpha_i(\mathbf{x}) = |h_i(\mathbf{x})|^2$$

Squared and non-squared penalty functions

minimize $(x^2 - 20 \ln x)$ subject to $x \geq 5$



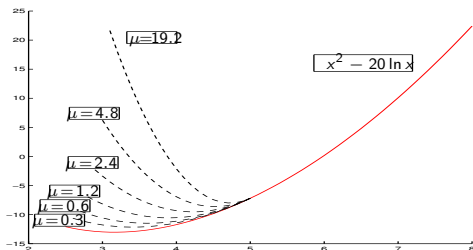
Figur: Squared and non-squared penalty function. g_i differentiable \implies squared penalty function differentiable

Squared penalty functions

- In practice: Start with a low value of $\mu > 0$ and increase the value as the computations proceed

- **Example:** $\text{minimize } (x^2 - 20 \ln x) \text{ subject to } x \geq 5$ (*)

\Rightarrow $\text{minimize } (x^2 - 20 \ln x + \mu(\max\{0, 5 - x\})^2)$ (**)

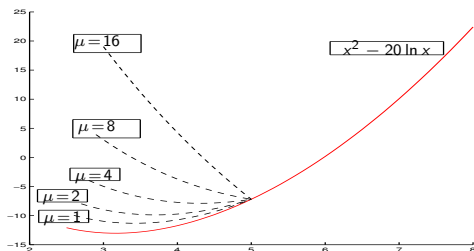


Figur: Squared penalty function: $\exists \mu < \infty$ such that an optimal solution for (**) is optimal (feasible) for (*)

Non-squared penalty functions

- In practice: Start with a low value of $\mu > 0$ and increase the value as the computations proceed
- **Example:** minimize $(x^2 - 20 \ln x)$ subject to $x \geq 5$ (+)

⇒ minimize $(x^2 - 20 \ln x + \mu \max\{0, 5 - x\})$ (++)



Figur: Non-squared penalty function: For $\mu \geq 6$ the optimal solution for (++) is optimal (and feasible) for (+)

Constrained optimization: Barrier methods

Consider only inequality constraints

$$\begin{aligned} & \text{minimize}_{\mathbf{x} \in \mathbb{R}^n} && f(\mathbf{x}) \\ & \text{subject to} && g_i(\mathbf{x}) \leq 0, \quad i \in \mathcal{L} \end{aligned} \quad (3)$$

- Drop the constraints and add terms in the objective that *prevents from approaching the boundary* of the feasible set

$$\text{minimize}_{\mathbf{x} \in \mathbb{R}^n} F_\mu(\mathbf{x}) := f(\mathbf{x}) + \mu \sum_{i \in \mathcal{L}} \alpha_i(\mathbf{x}) \quad (4)$$

where $\mu > 0$ and $\alpha_i(\mathbf{x}) \rightarrow +\infty$ as $g_i(\mathbf{x}) \rightarrow 0$ (as constraint i approaches being active)

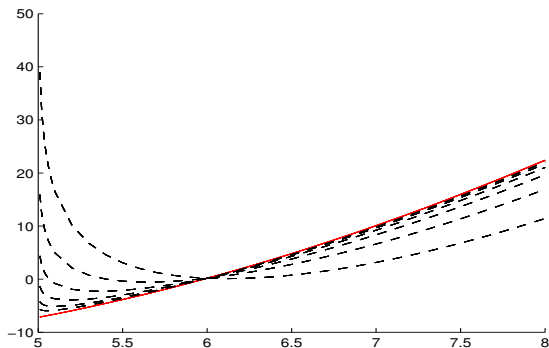
Common barrier functions

$$\alpha_i(\mathbf{x}) = -\ln[-g_i(\mathbf{x})] \quad \text{or} \quad \alpha_i(\mathbf{x}) = \frac{-1}{g_i(\mathbf{x})}$$

Logarithmic barrier functions

- Choose $\mu > 0$ and decrease it as the computations proceed
- **Example:** minimize $(x^2 - 20 \ln x)$ subject to $x \geq 5$

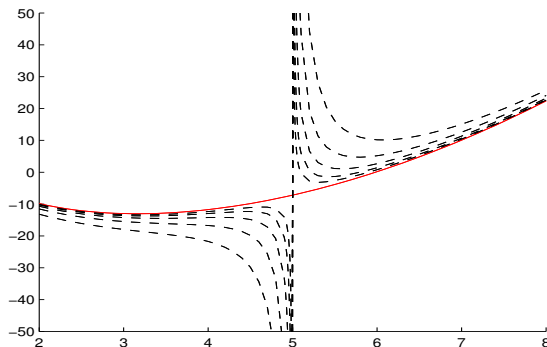
\Rightarrow minimize $_{x>5} (x^2 - 20 \ln x - \mu \ln(x - 5))$



Fractional barrier functions

- Choose $\mu > 0$ and decrease it as the computations proceed
- **Example:** minimize $(x^2 - 20 \ln x)$ subject to $x \geq 5$

\Rightarrow minimize $x > 5 \left(x^2 - 20 \ln x + \frac{\mu}{x-5} \right)$



Summary of the theoretical content of the course ...

... which may appear at the oral exam:

- Mathematical modelling of optimization problems; graphic solution
- *Linear programming*: BFSs; the simplex method; degeneracy; multiple optima; unbounded solution; infeasibility; starting solutions; LP duality; post-optimal and sensitivity analysis
- *Discrete and combinatorial optimization*: models of specific ILP problems; mathematical properties; complexity; algorithms; local/global optima; neighbourhoods; heuristics
- *Network flows*: Shortest paths; dynamic programming; LP models of network flows; maximum flows; minimum cost network flows; unimodularity; integrality property
- *Multi-objective optimization*: Pareto optimality; (non-)convexity; solution methods; objective space representation;
- *Non-linear optimization*: convexity; local/global optimality; mathematical properties; solution methods