MVE165/MMG630, Applied Optimization
Lecture 13
Overview of nonlinear programming

Ann-Brith Strömberg

2012–05–03

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
- Wind power generation: The energy content in the wind $\propto v^3$ (but Ass3b uses discretized measured data)

An overview of nonlinear optimization

General notation of nonlinear programs

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

Some special cases

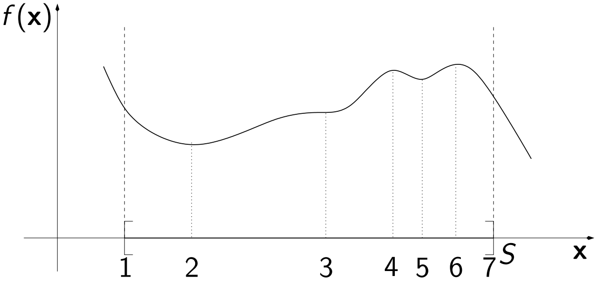
- Unconstrained problems ($\mathcal{L} = \mathcal{E} = \emptyset$):
  minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in \Re^n$
- Convex programming: f convex, g_i convex, $i \in \mathcal{L}$, h_i linear, $i \in \mathcal{E}$.
- Linear constraints: g_i, $i \in \mathcal{L}$, and h_i, $i \in \mathcal{E}$
  - Quadratic programming: $f(\mathbf{x}) = \mathbf{c}^T\mathbf{x} + \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x}$
  - Linear programming: $f(\mathbf{x}) = \mathbf{c}^T\mathbf{x}$

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; Ass3a)
- For **convex** problems: Algorithms converge to an optimal solution
- Nonlinear problems can have *local* optima that are *not global* optima

Possible extremal points for

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



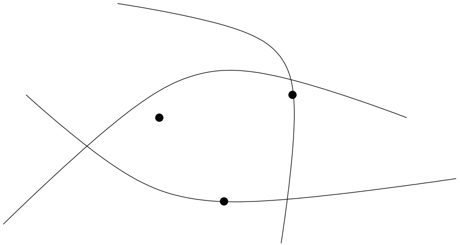
- boundary points of S
- stationary points, where $f'(\mathbf{x}) = 0$
- discontinuities in f or f' DRAW!

Boundary and stationary points (Ch. 10.0)

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

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

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



- $\bar{\mathbf{x}}$ 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

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

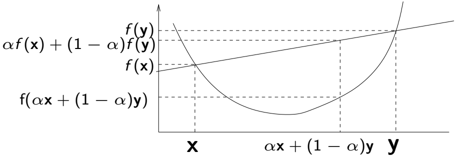
- The concept of **convexity** is essential
- Functions: convex (minimization), concave (maximization)
- Sets: convex (minimization and maximization)
- The minimization (maximization) of a convex (concave) function over a convex set is referred to as a convex optimization problem
- (Def. 9.5) If f and 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
- (Thm. 9.1) Let $\mathbf{x}^*$ be a *local* optimum for a convex optimization problem. Then $\mathbf{x}^*$ is also a *global* optimum

Convex functions

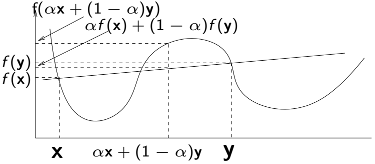
- A function f is *convex* on S if, for any $\mathbf{x}, \mathbf{y} \in S$ it holds that

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

A CONVEX FUNCTION



A NON-CONVEX FUNCTION

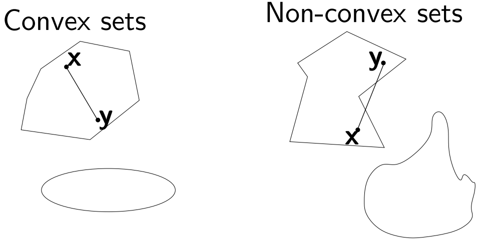


- f is *strictly convex* on S if, for any $\mathbf{x}, \mathbf{y} \in S$ such that $\mathbf{x} \neq \mathbf{y}$ it holds that

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

Convex 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:



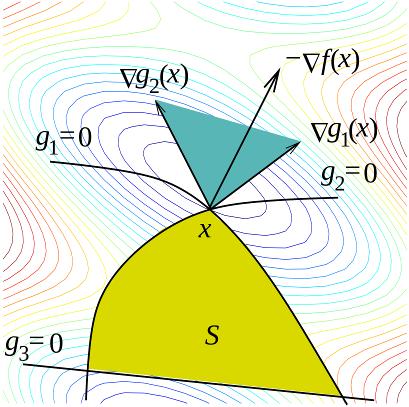
- Consider a set S defined by the intersection of $m = |\mathcal{L}|$ inequalities, where the functions $g_i : \Re^n \mapsto \Re$, $i \in \mathcal{L}$:
$$S = \{\, \mathbf{x} \in \Re^n \mid g_i(\mathbf{x}) \leq 0, \;\; i \in \mathcal{L} \,\}$$
- (Thms. 9.2 & 9.3) If all the functions $g_i(\mathbf{x})$ $i \in \mathcal{L}$, are convex on $\Re^n$, then S is a convex set

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

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

$$\begin{aligned}
\nabla f(\mathbf{x}^*) + \sum_{i \in \mathcal{L}} \mu_i \nabla g_i(\mathbf{x}^*) &= \mathbf{0}^n \\
\mu_i g_i(\mathbf{x}^*) &= 0, \qquad i \in \mathcal{L} \\
g_i(\mathbf{x}^*) &\le 0, \qquad i \in \mathcal{L} \\
\boldsymbol{\mu} &\ge \mathbf{0}^m
\end{aligned}$$

Geometry of the Karush-Kuhn-Tucker conditions



Figur: Geometric interpretation of the Karush-Kuhn-Tucker conditions. At a local minimum, minus the gradient of the objective 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 $f, g_i : \Re^n \mapsto \Re$, $i \in \mathcal{L}$, are convex and differentiable.
- If the conditions (where $m = |\mathcal{L}|$)

$$\begin{aligned} \nabla f(\mathbf{x}^*) + \sum_{i \in \mathcal{L}} \mu_i \nabla g_i(\mathbf{x}^*) &= \mathbf{0}^n \\ \mu_i g_i(\mathbf{x}^*) &= 0, \qquad i \in \mathcal{L} \\ \boldsymbol{\mu} &\geq \mathbf{0}^m \end{aligned}$$

  hold, then $\mathbf{x}^* \in S$ is a global minimum of f over $S = \{ \mathbf{x} \in \Re^n \mid g_i(\mathbf{x}) \leq 0, i \in \mathcal{L} \}$.
- 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)
- algorithm construction
- derive properties of a solution to a non-linear program

Example

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

- Is $\mathbf{x}^0 = (1,2)^{\mathrm{T}}$ a Karush-Kuhn-Tucker point?
- Is it an optimal solution?
- $\nabla f(\mathbf{x}) = (4x_1 + 2x_2 - 10, 2x_1 + 2x_2 - 10)^{\mathrm{T}}$, $\nabla g_1(\mathbf{x}) = (2x_1, 2x_2)^{\mathrm{T}}$, $\nabla g_2(\mathbf{x}) = (3,1)^{\mathrm{T}}$

$$\Rightarrow$$

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

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

Example, continued

- 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

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

General iterative search method for unconstrained optimization (Ch. 2.5.1)

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)

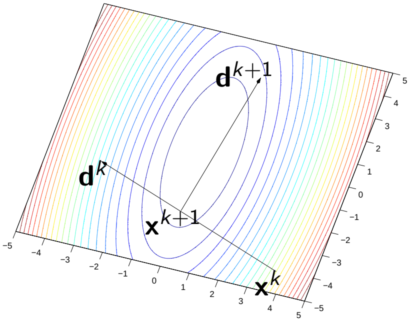
- Goal: $f(\mathbf{x}^{k+1}) < f(\mathbf{x}^k)$ (minimization)
- How does f change locally in a direction $\mathbf{d}^k$ at $\mathbf{x}^k$?
- Taylor expansion (Ch. 9.2):
  $f(\mathbf{x}^k + t\mathbf{d}^k) = f(\mathbf{x}^k) + t\nabla f(\mathbf{x}^k)^{\mathrm{T}}\mathbf{d}^k + \mathcal{O}(t^2)$
- For sufficiently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

General iterative search method for unconstrained optimization (Ch. 2.5.1)

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:

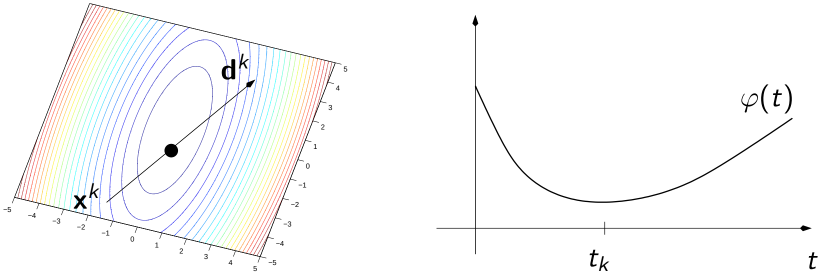
$$\text{minimize}_{\ t \geq 0} \varphi(t) := f(\mathbf{x}^k + t \cdot \mathbf{d}^k)$$

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

General iterative search method for unconstrained optimization

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 ($\varepsilon_j > 0$, $j = 1, \dots, 4$)
  (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) $t_k < \varepsilon_4$

  These 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:

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

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

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

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

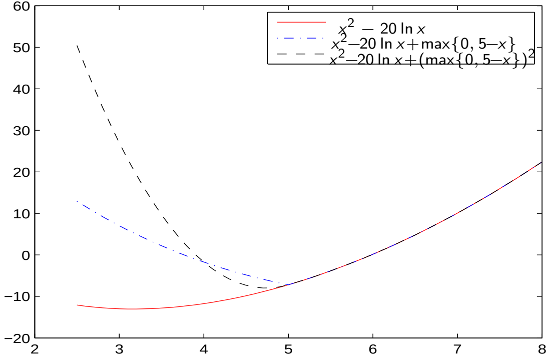
- Common penalty functions (which are differentiable?):

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

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

Squared and non-squared penalty functions

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

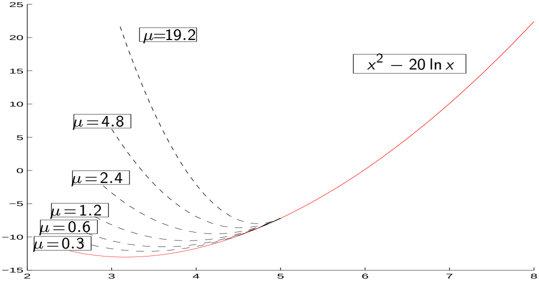


Figur: Squared and non-squared penalty function. g_i differentiable $\implies$ squared penalty function differentiable

Squared penalty functions

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

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

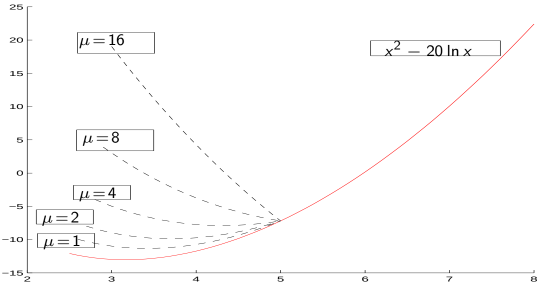


Figur: Squared penalty function: $\nexists \mu < \infty$ such that an optimal solution for $(**)$ is optimal (feasible) for $(*)$

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 $x \geq 5$ $\qquad (+)$

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



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:

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

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

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

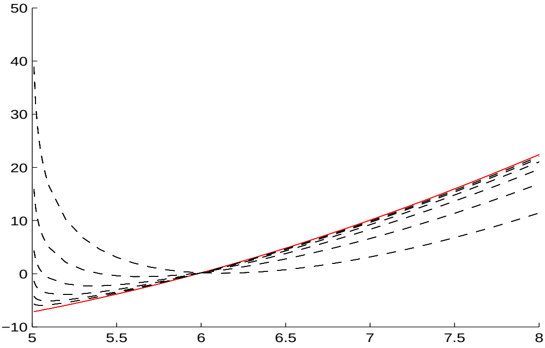
  where $\mu > 0$ and $\alpha_i(\mathbf{x}) \to +\infty$ as $g_i(\mathbf{x}) \to 0$ (as constraint i approaches being active)

- Common barrier functions:
  - $\alpha_i(\mathbf{x}) = -\ln[-g_i(\mathbf{x})]$ or $\alpha_i(\mathbf{x}) = \frac{-1}{g_i(\mathbf{x})}$

Logarithmic barrier functions

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

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

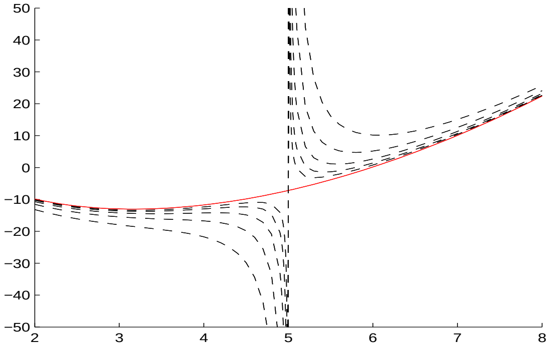


Figur: Logarithmic barrier function: $\mu \in \{10, 5, 2.5, 1.25, 0.625, 0.3125\}$

Fractional barrier functions

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

$\Rightarrow$ minimize $_{x>5}$ $\; x^2 - 20 \ln x + \frac{\mu}{x-5}$



Figur: Fractional barrier function: $\mu \in \{10, 5, 2.5, 1.25, 0.625\}$