Chapter 7: Penalty and barrier function methods

Keijo Ruotsalainen

Mathematics Division
Course of the lecture

- Penalty and barrier function techniques
  - Logarithmic penalty term
  - path following algorithms
- Barrier function method

Problem with inequality constraints

Minimize \( f(x) \)
subject to \( x \in X, \ g_j(x) \leq 0, \ j = 1, \ldots, m, \)
assuming that \( f(x), \ g_1(x), \ldots, \ g_m(x) \) are smooth, \( X = \mathbb{R}^n \) or \( \mathbb{R}_+^n \)
and that the set

\[
S = \{ x \in X : g_j(x) < 0, \ j = 1, \ldots, m \} \neq \emptyset.
\]
Barrier function method

- Introduce a \textbf{barrier function} $B(g(x)) \geq 0$, which is differentiable and approaches to $\infty$, as $g_j(x) \to 0^-:

- Examples:

  \[
  B(g(x)) = -\sum_{j=1}^{m} \ln(-g_j(x)), \quad B(g(x)) = -\sum_{j=1}^{m} \frac{1}{g_j(x)}.
  \]

- \textbf{Barrier function method}: For every $k = 0, 1, 2, 3, \ldots$ solve

  \[
  \min_{x \in S} \{ f(x) + \varepsilon^{(k)} B(g(x)) \} \Rightarrow x^{(k)},
  \]

  where the parameter sequence $\{ \varepsilon^{(k)} \}_{k \in \mathbb{N}}$ satisfies the condition:

  \[
  0 < \varepsilon^{(k+1)} < \varepsilon^{(k)} \text{ and } \varepsilon^{(k)} \to 0.
  \]
Introductory example

- Optimization problem: \( \min \{x - 1 \geq 0\} \times \)
- Optimization with logarithmic barrier function:
  \[
  \min_{x \geq 1} \{x - \epsilon \ln(x - 1)\} \rightarrow x^*(\epsilon) = 1 + \epsilon.
  \]
Example 1

Solve: \( \min_{1-x_1-x_2 \leq 0} x_1^2 + x_2^2 \).

- Logarithmic barrier function: \( B(g(x)) = -\ln(x_1 + x_2 - 1) \).
- Solve the problem \( \min_{x_1 + x_2 - 1 \geq 0} \{f(x) + \epsilon B(x)\} \).
- Let us forget the constraint for a moment: \( \min x_1^2 + x_2^2 - \epsilon \ln(x_1 + x_2 - 1) \).
- The optimality criteria (convexity):

\[
\nabla f(x^*) + \epsilon \nabla B(x^*) = \begin{bmatrix} 2x_1^* - \frac{\epsilon}{x_1^* + x_2^* - 1} \\ 2x_2^* - \frac{\epsilon}{x_1^* + x_2^* - 1} \end{bmatrix} = 0 \Rightarrow \begin{cases} x_1^* = x_2^* \\ 2x_1^* - \frac{\epsilon}{2x_1^* - 1} = 0 \end{cases}
\]

\[ \Rightarrow 4x_1^{*2} - 2x_1^2 - \epsilon = 0 \Rightarrow x_1^* = \frac{2 + 2\sqrt{1 + 4\epsilon}}{8} \in S, \left( \frac{2 - 2\sqrt{1 + 4\epsilon}}{8} \notin S \right) \]

Hence we obtain

\[ \lim_{\epsilon \to 0} x^* = \left( \frac{1}{2}, \frac{1}{2} \right). \]
**On the convergence**

**Theorem 7.1**

Let \( \{ x^{(k)} \} \) be a sequence constructed by the barrier function method. Then every accumulation point of the sequence is a solution of the optimization problem.

**Todistus.**

Let \( \lim_{k \to \infty} x^{(k)} = \bar{x}, \ x^{(k)} \in S \Rightarrow \bar{x} \in S \). If \( \bar{x} \) is not a global minimizer, then there exists \( x^* \in S \cap \{ g_j(x^*) < 0, \forall j \} \) such that \( f(x^*) < f(\bar{x}) \). By the definition of the sequence \( x^{(k)} \)

\[
    f(x^{(k)}) + \epsilon^{(k)}B(x^{(k)}) \leq f(x^*) + \epsilon^{(k)}B(x^*), \forall k
\]

\[\Rightarrow f(\bar{x}) + \lim_{k \to \infty} \epsilon^{(k)}B(x^{(k)}) \leq f(x^*) < f(\bar{x}).\]

In other words, \( \lim_{k \to \infty} \epsilon^{(k)}B(x^{(k)}) < 0 \). On the other hand \( \bar{x} \in S \) and hence \( \lim_{k \to \infty} \epsilon^{(k)}B(x^{(k)}) \geq 0 \) which is a contradiction.
Lagrange multipliers

- KKT-conditions: There exist $\mu_i \geq 0$, $i = 1, \ldots, m$ such that
  \[
  \nabla f(\mathbf{x}) + \sum_{i=1}^{m} \mu_i \nabla g_i(\mathbf{x}) = 0, \quad \mu_i g_i(\mathbf{x}) = 0, \quad i = 1, \ldots, m
  \]

- Barrier problem: $\min_{\{g(\mathbf{x}) < 0\}} \{f(\mathbf{x}) + \epsilon \sum_{i=1}^{m} \Phi(g_j(\mathbf{x}))\}$

- The solution $x(\epsilon)$ of the barrier problem has the optimality condition
  \[
  \nabla f(x(\epsilon)) + \sum_{i=1}^{m} \epsilon \Phi'(g_i(x(\epsilon))) \nabla g_i(x(\epsilon)) = 0.
  \]

- When $\epsilon \to 0^+$, then $x(\epsilon) \to \mathbf{x}$ and
  \[
  \epsilon \Phi'(g_i(x(\epsilon))) \to 0, \quad g_i(\mathbf{x}) < 0
  \]
  \[
  \epsilon \Phi'(g_i(x(\epsilon))) \to \mu_i, \quad g_i(\mathbf{x}) = 0
  \]
LP-problem and logarithmic barrier function

- LP-problem: \( \min \{Ax = b, x \geq 0\} \ c^T x \)
- Barrier problem:

\[
\min_{x \in S} F_\epsilon(x) = c^T x - \epsilon \sum_{i=1}^{n} \ln(x_i) \rightarrow x(\epsilon),
\]

where \( S = \{x | Ax = b, x > 0\} \neq \emptyset \) and bounded.

- When \( \epsilon \to 0 \), \( x(\epsilon) \) follows the central path.
- All the central paths start from the analytic center

\[
x_\infty : \min_{x \in S} - \sum_{i=1}^{n} \ln(x_i)
\]

and converge towards the optimal vertex.
Path following algorithm

- The Lagrangian: $c^T x - \epsilon \sum_{i=1}^{n} \ln(x_i) - v^T [Ax - b]$

- KKT-conditions: Find $x \in \mathbb{R}^n$, $v \in \mathbb{R}^m$ such that

$$c - \epsilon \text{diag}(x)^{-1} e - A^T v = 0$$

$$Ax = b, \ (x > 0),$$

where $e = [1, 1, \cdots, 1]^T \in \mathbb{R}^n$, $X = \text{diag}(x)$.

- alternative formulation:

$$Ax = b$$

$$A^T v + u = c$$

$$u = \epsilon X^{-1} e \text{ or } UXe = \epsilon e.$$
Path following algorithm

- For every $\epsilon > 0$ the barrier problem has a unique solution $x(\epsilon) > 0$.
- $\Rightarrow$ The dual variables $u(\epsilon) = \epsilon X(\epsilon)^{-1}e$ ja $v(\epsilon)$ are uniquely defined.
- $w(\epsilon) = (x(\epsilon), u(\epsilon), v(\epsilon)) \rightarrow w = (\overline{x}, \mu, \nu)$, which is the solution of the primal-dual problem (saddle point).
- Duality gap: $c^T x - b^T v = u^T x = n\epsilon$, because
  \[
  c^T x - b^T v = c^T x - (Ax)^T v = x^T c - x^T A^T v \\
  x^T (c - A^T v) = x^T u = x^T (\epsilon X^{-1} e) = n\epsilon \rightarrow 0
  \]
- The idea of the algorithm: Start with $\epsilon > 0$ and $\tilde{w}$ close to the vector $w(\epsilon)$. Update $\epsilon \mapsto \beta \epsilon$, $0 < \beta < 1$, and $\tilde{w}$ with one step of the Newton method.
On the convergence of Path Following-algorithm

- $w$ is “close enough” of the vector $w(\epsilon)$, when

$$Ax = b, \quad A^\top v + u = c, \quad \|XUe - \epsilon e\| \leq \theta \epsilon,$$

where $0 \leq \theta < 0.5$ ja $u^\top x = n\epsilon$.

- In this way constructed sequence converges towards the optimal primal-dual solution.

- Let $\epsilon > 0$, $\tilde{w} = (\tilde{x}, \tilde{u}, \tilde{v})$ be given such that (1) is satisfied.

- Update $\epsilon \mapsto \beta \epsilon$. Investigate the KKT-conditions $H(w) = 0$ for the new parameter value $\epsilon$.

- First order approximation to the KKT-conditions are

$$H(\tilde{w}) + J(\tilde{w})(w - \tilde{w}) = 0,$$

where $J(\tilde{w})$ is the Jacobian of $H$, i.e its derivative.
Path following-algorithm

- Newton step:

\[ J(\tilde{w})d_w = -H(\tilde{w}) \]
\[ \hat{w} = \tilde{w} + d_w \]

- The correction term \( d_w = (d_x, d_u, d_v) \) is solved from the system of equations:

\[ A d_x = 0 \] \hspace{1cm} (2)
\[ A^\top d_v + d_u = 0 \] \hspace{1cm} (3)
\[ \tilde{U} d_x + \tilde{X} d_u = \epsilon e - \tilde{X} \tilde{U} e \] \hspace{1cm} (4)

- The solution (dick head-form):

\[ d_v = -[A\tilde{U}^{-1}\tilde{X}A^\top]^{-1}A\tilde{U}^{-1}[\epsilon I - \tilde{X} \tilde{U}]e \]
\[ d_u = -A^\top d_v \]
\[ d_x = \tilde{U}^{-1}[\epsilon e - \tilde{X} \tilde{U} e - \tilde{X} d_u] \]
Convergence of the Path Following

Theorem 7.2

Let \( \tilde{w} = (\tilde{x}, \tilde{u}, \tilde{v}) \) be such that \( \tilde{x} > 0, \tilde{u} > 0 \) ja (1) on voimassa, kun \( \epsilon > 0 \). Define \( \tilde{\epsilon} = \beta \epsilon \), where \( \beta \) satisfies

\[
\beta = 1 - \frac{\delta}{\sqrt{n}}, \quad 0 < \delta < \sqrt{n}, \quad \frac{\theta^2 + \delta^2}{2(1 - \theta)} \leq \theta(1 - \frac{\delta}{\sqrt{n}}), \quad 0 \leq \theta \leq \frac{1}{2}.
\]

Then the point \( \hat{w} \) computed by the path following algorithm fulfills the condition (1) and \( \hat{x}, \hat{u} > 0 \), when \( \epsilon = \tilde{\epsilon} \). In other words, starting the iterations from the point that satisfies the condition (1) we obtain a sequence \( (\epsilon_k, w_k) \) whose limit point (\( \approx \) accumulation point) is the solution of the LP-problem (primal-dual solution).

The complexity: **Polynomial**. Works for quadratic and convex problems as well.
Penalty function method
Constrained problem:

\[
\begin{aligned}
\min & \quad f(x) \\
\text{subject to} & \quad g(x) \leq 0 \\
& \quad h(x) = 0
\end{aligned}
\]

Penalty function

\[
P(x) = \sum_{i=1}^{m} \Phi(g_i(x)) + \sum_{j=1}^{l} \Psi(h_j(x)).
\]

Assumptions:

\[
\begin{aligned}
\Phi'(y) & \geq 0 \\
\Phi(y) &= 0, \quad y \leq 0, \\
\Phi(y) &> 0, \quad y > 0 \\
\Psi(0) &= 0 \\
\Psi(y) &> 0, \quad y \neq 0 \\
\Psi'(y) &\text{jatkuva}
\end{aligned}
\]
Penalty problem

- Usually $\Phi(y) = \max\{0, y^2\}$, $\Psi(y) = y^2$
- The penalty problem:
  \[
  \theta(\mu) = \min_{x \in \mathbb{R}^n} f(x) + \mu P(x).
  \]
- Sakkofunktio-ongelmassa ratkaistaan
  \[
  \max_{\mu \geq 0} \theta(\mu).
  \]
- Voidaan osoittaa, että
  \[
  \inf\{f(x) : g(x) \leq 0, h(x) = 0\} = \lim_{\mu \to \infty} \theta(\mu).
  \]