4TE3/6TE3

Algorithms for

Continuous Optimization
Tamás TERLAKY
Computing and Software
McMaster University

Hamilton, January 2004

terlaky@mcmaster.ca Tel: 27780
The general NLO problem

\[(NLO) \ \min \ f(x) \]
\[\text{s.t.} \quad h_i(x) = 0, \quad i \in I = \{1, \ldots, p\} \]
\[g_j(x) \leq 0, \quad j \in J = \{1, \ldots, m\} \]
\[x \in \mathbb{C}.\]

where \(x \in \mathbb{R}^n, \mathbb{C} \subseteq \mathbb{R}^n\) is a certain set and
\(f, h_1, \ldots, h_p, g_1, \ldots, g_m\) are functions defined on \(\mathbb{C}\).

Set of feasible solutions:
\[\mathcal{F} = \{x \in \mathbb{C} | h_i(x) = 0, \forall i \text{ and } g_j(x) \leq 0, \forall j\}.\]

**Definition 1** Let two points \(x^1, x^2 \in \mathbb{R}^n\) and \(0 \leq \lambda \leq 1\) be given. Then the point
\[x = \lambda x^1 + (1 - \lambda)x^2\]
is a convex combination of the two points \(x^1, x^2\).

**Definition 2** The set \(\mathbb{C} \subseteq \mathbb{R}^n\) is called convex, if all convex combinations of any two points \(x^1, x^2 \in \mathbb{C}\) are again in \(\mathbb{C}\).

Convex and nonconvex sets in the plane.
Convex functions

**Definition 2** A function $f : \mathcal{C} \to \mathbb{R}$ defined on a convex set $\mathcal{C}$ is called convex if for all $x^1, x^2 \in \mathcal{C}$ and $0 \leq \lambda \leq 1$ one has

$$f(\lambda x^1 + (1 - \lambda)x^2) \leq \lambda f(x^1) + (1 - \lambda)f(x^2).$$

**Definition 3** The epigraph of a function $f : \mathcal{C} \to \mathbb{R}$ is the $(n + 1)$-dimensional set

$$\{(x, \tau) : f(x) \leq \tau, x \in \mathcal{C}, \tau \in \mathbb{R}\}.$$ 

The epigraph of a convex function $f$.

**Lemma 1** A function $f$ is convex iff $\text{epi}(f)$ is convex. 

$\mathbb{R}^n_+$ is convex.

$f : \mathbb{R}^n \to \mathbb{R}$ is called a quadratic function if there is a square matrix $Q \in \mathbb{R}^{n \times n}$, a vector $c \in \mathbb{R}^n$ and a number $\gamma \in \mathbb{R}$ such that

$$f(x) = \frac{1}{2}x^TQx + c^Tx + \gamma \text{ for all } x \in \mathbb{R}^n;$$

if $Q = 0$ then $f$ is called affine (linear).
Examples

Convex functions:

- $\exp(x)$
- $-\log(x)$
- $x \log(x)$
- $-\sqrt{x}$ when $x \geq 0$
- $(x_1 - 2)^2 + (x_2 + 1)^2 - 2$
- $\cos(x)$ on the interval $[\frac{\pi}{2}, \frac{3\pi}{2}]$

Non-Convex functions:

- $\log(x)$
- $-(x_1 - 2)^2 + (x_2 + 1)^2 - 2$
- $\sin(x)$
- $x \sin(x)$
- $(x - 2)^4 - 10 \times (x - 2)^2$
Pre-knowledge:

You are supposed to be familiar with.

- Calculus
- Linear algebra
- Positive semidefinite matrix
- Gradient of $n$-dimensional functions
- Hessian of $n$-dimensional functions
- Taylor series of $n$-dimensional functions
- Lagrange multiplicator rule
Classification of NLO problems

Linear Optimization (LO): $f, h_1, \ldots, h_p, g_1, \ldots, g_m$ are affine (linear) and the set $C$ either equals to $\mathbb{R}^n$ or to $\mathbb{R}_+^n$.

Unconstrained Optimization: The index sets $I$ and $J$ are empty and $C = \mathbb{R}^n$.

Convex Optimization (CO): $f, g_1, \ldots, g_m$ are convex, $h_1, \ldots, h_p$ are affine and $C$ is convex.

Smooth Convex Optimization: It is a CO problem and all the functions are at least twice continuously differentiable (further smoothness requirements later).

Quadratic Optimization (QO): The objective $f$ is quadratic $h_1, \ldots, h_p, g_1, \ldots, g_m$ are affine (linear) and either $C = \mathbb{R}^n$ or $C = \mathbb{R}_+^n$.

Quadratically Constrained QO (QCQO): The same as QO, but $g_1, \ldots, g_m$ are quadratic.

Convex Quadratic Optimization (CQO): The same as QO, but the objective $f$ is convex.

Convex Quadratically Constrained QO (CQCQO): As QCQO, but the objective function $f$ and the quadratic functions $g_1, \ldots, g_m$ are convex.
Solve the problems

Solve the following problem:

\[
\min \; \exp(x) + \frac{1}{2}x^2
\]

Show that the function \(\exp(x) + \frac{1}{2}x^2\) is convex.

Consider the following problem:

\[
\begin{align*}
\min & \quad \sqrt{(x_1 - x_4)^2 + (x_2 - x_5)^2 + (x_3 - x_6)^2} \\
& \quad x_1^2 + x_2^2 + x_3^2 \leq 5 \\
& \quad (x_4 - 3)^2 + x_5^2 \leq 1 \\
& \quad 4 \leq x_6 \leq 8
\end{align*}
\]

Check that this is a convex problem.

Find the optimal solution of this problem.
Generic Algorithm

\[
\begin{align*}
\min & \quad f(x) \\
\text{s.t.} & \quad x \in C.
\end{align*}
\]

Input:
\(\epsilon > 0\) is the accuracy parameter;
\(x^0\) is a given (relative interior) feasible point;

Step 0: \(x := x^0, k = 0;\)

Step 1: Find search direction \(s^k\) s.t. \(\delta f(x^k, s^k) < 0\)
(This should be a descending feasible direction in the constrained case).

Step 1a: If no such direction exists STOP, optimum found.

Step 2: Line search : find \(\lambda^k = \min_{\lambda} f(x^k + \lambda s^k);\)

Step 3: \(x^{k+1} = x^k + \lambda^k s^k, \quad k = k + 1;\)

Step 4: If stopping criteria satisfied STOP, else GOTO Step 1.
Consider the problem
\[ \text{minimize } f(x), \]
where \( x \in \mathcal{F} \subseteq \mathbb{R}^n \) a convex set and \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is a differentiable function. First we define local and global minima of the above problem.

**Definition 4** Let a function \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) be given.

A point \( \bar{x} \in \mathbb{R}^n \) is a local minimum of the function \( f \) if there is an \( \epsilon > 0 \) such that \( f(\bar{x}) \leq f(x) \) for all \( x \in \mathbb{R}^n \) when \( \|\bar{x} - x\| \leq \epsilon. \)

A point \( \bar{x} \in \mathbb{R}^n \) is a strict local minimum of the function \( f \) if there is an \( \epsilon > 0 \) such that \( f(\bar{x}) < f(x) \) for all \( x \in \mathbb{R}^n \) when \( \bar{x} \neq x \) and \( \|\bar{x} - x\| \leq \epsilon. \)

A point \( \bar{x} \in \mathbb{R}^n \) is a global minimum of the function \( f \) if \( f(\bar{x}) \leq f(x) \) for all \( x \in \mathbb{R}^n. \)

A point \( \bar{x} \in \mathbb{R}^n \) is a strict global minimum of the function \( f \) if \( f(\bar{x}) < f(x) \) for all \( x \in \mathbb{R}^n \) when \( \bar{x} \neq x. \)

**Lemma 2** Any (strict) local minimum \( \bar{x} \) of a convex function \( f \) over the set \( \mathcal{F} \) is a (strict) global minimum of \( f \) as well.
Examples:

Local and Global minimizers

Find the local/global minimum of the functions if exist

- $\exp(x)$
- $-\log(x)$
- $x \log(x)$
- $-\sqrt{x}$ when $x \geq 0$
- $(x_1 - 2)^2 + (x_2 + 1)^2 - 2$
- $\cos(x)$ on the interval $[\frac{\pi}{2}, \frac{3\pi}{2}]$
- $\log(x)$
- $-(x_1 - 2)^2 + (x_2 + 1)^2 - 2$
- $\sin(x)$
- $x \sin(x)$
- $(x - 2)^4 - 10 \cdot (x - 2)^2$
Definition 5 Let $\alpha_1, \alpha_2, \ldots, \alpha_k, \ldots \to \alpha$ be a convergent sequence. We say that the order of convergence of this sequence is $p^*$ if

$$p^* = \sup \left\{ p : \limsup_{k \to \infty} \frac{|\alpha_{k+1} - \alpha|}{|\alpha_k - \alpha|^p} < \infty \right\}.$$ 

The larger $p^*$ is, the faster the convergence. Let

$$\beta = \limsup_{k \to \infty} \frac{|\alpha_{k+1} - \alpha|}{|\alpha_k - \alpha|}.$$ 

If $p^* = 1$ and $0 < \beta < 1$ we are speaking about linear (or geometric rate of) convergence. If $p^* = 1$ and $\beta = 0$ the convergence rate is super-linear, while if $\beta = 1$ the convergence rate is sub-linear. If $p^* = 2$ then the convergence is quadratic.

Example 1: The sequence $\alpha_k = a^k$, where $0 < a < 1$ converges linearly to zero while $\beta = a$.
Example 2: The sequence $\alpha_k = a^{(2^k)}$, where $0 < a < 1$ converges quadratically to zero.
Example 3: The sequence $\alpha_k = \frac{1}{k}$ converges sub-linearly to zero.
Example 4: The sequence $\alpha_k = \left(\frac{1}{k}\right)^k$ converges super-linearly to zero.

Exercise: Find a sequence that converges to its limit with order 4.
A sequence of candidate solutions converge to an optimal point:

\[ x^0, x^1, x^2, \ldots, x^k, \ldots \rightarrow \bar{x} \in \mathbb{R}^n \]

We can define \( \alpha_k \) in various ways:

- \( \alpha_k = \|x^k - \bar{x}\| \)
  Problem: depends on scaling.
  **Do normalize!**

- \( \alpha_k = \frac{\|x^k - \bar{x}\|}{\|x^k\|} \)
  Problem: \( x^k \) might be zero.
  **Never divide by zero!**

- \( \alpha_k = \frac{\|x^k - \bar{x}\|}{1 + \|x^k\|} \)
  Problem: \( \|x^k - \bar{x}\| \) might be small while \( |f(x^k) - f(x)| \) big.
  **Use objective or combined merit function.**

- \( \alpha_k = \frac{|f(x^k) - f(\bar{x})|}{1 + |f(x^k)|} \), or \( \alpha_k = \frac{\|x^k - \bar{x}\|}{1 + \|x^k\|} + \frac{|f(x^k) - f(\bar{x})|}{1 + |f(x^k)|} \)
  Problem: \( \bar{x} \) is not known!
  **Monitor progress of the algorithm.**

- \( \alpha_k = \frac{|f(x^{k+1}) - f(x^k)|}{1 + |f(x^k)|} \), or
  \( \alpha_k = \frac{\|x^{k+1} - x^k\|}{1 + \|x^k\|} + \frac{|f(x^{k+1}) - f(x^k)|}{1 + |f(x^k)|} . \)
Optimization versus feasibility

Solving nonlinear equations/inequalities

\[
\min \{ f(x) \mid x \in \mathcal{F} \} \\
\mathcal{F} = \{ x \in \mathcal{C} \mid h_i(x) = 0, \forall \ i \ \text{and} \ g_j(x) \leq 0, \forall \ j \}.
\]

- Feasibility (solving nonlinear equations/inequalities) is a part of the optimization problem.
- Solving any nonlinear equations, inequalities can be regarded as an optimization problem.
  - We consider the objective function to be the zero function.
  - Solve: \( \{ h_i(x) = 0, \forall \ i, \ g_j(x) \leq 0, \forall \ j \} \) can be cast as

\[
\min \sum_{i=1}^{p} (h_i(x))^2 + \sum_{j=1}^{m} (g_j(x) + z_j^2)^2
\]

or

\[
\min \sum_{i=1}^{p} |h_i(x)| + \sum_{j=1}^{m} \max\{0, g_j(x)\}
\]

- Optimization problem can be cast as solving a series of feasibility problems for various \( \vartheta \) values:

\[
f(x) = \vartheta; \ h_i(x) = 0, \forall \ i, \ g_j(x) \leq 0, \forall \ j
\]

- We will see later that duality theory will provide a more powerful approach.

Nonlinear equations can be solved by using Newton’s method. Let \( F : \mathbb{R}^n \rightarrow \mathbb{R}^n \) be continuously differentiable.

\[
F(x) = 0 \\
F(x^k) + JF(x)(x - x^k) = 0 \\
x^{k+1} = x^k - (JF(x))^{-1} F(x^k)
\]
Direct search (Black box) methods

Pattern search methods

Cyclic coordinate search (Hooke and Jeeves)

\[ \min \{ f(x) \mid x \in \mathbb{R}^n \} \]

**Input:** \( \epsilon > 0 \) is the accuracy parameter;
\( n \) linearly independent search direction \( d^1, d^2, \ldots, d^n \);
\( 0 < \alpha < 1 \) damping factor;
\( x^0 \) is a given initial point;

**Step 0:** \( k = 0; y^0 = x^0 \);

**Step 1:** \( y^0 = x^k \), CALL PS;

**Step 2:** IF \( y^n = y^0 \) THEN \( d^j = \alpha d^j \); GOTO Step 6.

**Step 2a:** \( z = y^n \)

**Step 3:** Here \( f(z) < f(x^k) \);
Let \( y^0 := y^n + (y^n - x^k) \); CALL PS;

**Step 4:** IF \( f(y^n) < f(z) \) THEN \( z = y^n \);

**Step 5:** Here \( f(z) < f(x^k) \); Let \( x^{k+1} = z \); \( k = k+1 \)

**Step 6:** IF \( \|d_1\| < \epsilon \) STOP, ELSE GOTO Step 1.

**SUBROUTINE (PS):** Pattern search:

FOR \( j = 1 : n \) DO
IF \( f(y^{j-1} + d^j) < f(y^{j-1}) \) THEN \( y^j = y^{j-1} + d^j \);
ELSEIF \( f(y^{j-1} - d^j) < f(y^{j-1}) \)
THEN \( y^j = y^{j-1} - d^j \);
ELSE \( y^j = y^{j-1} \);
END
Direct search methods II

Simplex method (Nelder and Mead) I

\[
\min \{ f(x) \mid x \in \mathbb{R}^n \}
\]

**Input:** \( \epsilon > 0 \) is the accuracy parameter;
\( n + 1 \) points \( x^0, x^1, \ldots, x^n \) in general position;
\( 0 < \alpha < 1 \) damping factor;
\( 0 < \beta < 1 \) contraction factor;
\( 1 < \gamma \) extension factor;

**Step 1:** SORT the points in the order of ascending function value;
\[
f(x^0) \leq f(x^1) \leq \cdots \leq f(x^n)
\]

**Step 2:** LET \( \bar{x} := \frac{1}{n} \sum_{j=0}^{n-1} x^j; \ x^r := \bar{x} + (\bar{x} - x^n); \)

**Step 2a:** IF \( f(x^r) < f(x^0) \)
THEN \( x^e := \bar{x} + \gamma(\bar{x} - x^n) \)
IF \( f(x^e) \leq f(x^r) \)
THEN DROP \( x^n \) ADD \( x^e \) GOTO 1.
ELSE DROP \( x^n \) ADD \( x^r \) GOTO 1.
Direct search methods III

Simplex method (Nelder and Mead)

Step 2b: IF $f(x^0) \leq f(x^r) < f(x^{n-1})$
THEN DROP $x^n$ ADD $x^r$ GOTO 1.

Step 2c: IF $f(x^{n-1}) \leq f(x^r) < f(x^n)$
THEN $x^c := \bar{x} + \beta(\bar{x} - x^n)$
IF $f(x^c) \leq f(x^{n-1})$
THEN DROP $x^n$ ADD $x^c$ GOTO 1.

Step 2d: IF $f(x^r) \geq f(x^n)$
THEN $x^c := \bar{x} - \beta(\bar{x} - x^n)$
IF $f(x^c) \leq f(x^{n-1})$
THEN DROP $x^n$ ADD $x^c$ GOTO 1.

Step 3: IF $\max\{\|x^i - x^j\| : 1 \leq i, j\} < \epsilon$
OR $f(x^n) - f(x^0) < \epsilon$ THEN STOP

Step 4: Contract the simplex to $x^0$;
$x^j := x^0 + \alpha(x^j - x^0), j = 1, 2, \ldots, n$; GOTO 1.