4TE3/6TE3

Algorithms for

Continuous Optimization

Tamás TERLAKY

Computing and Software

McMaster University

Hamilton, ON January 2004

terlaky@mcmaster.ca Tel: 27780

The general NLO problem

$$\begin{array}{lll} (NLO) & \text{min} & f(x) \\ & \text{s.t.} & h_i(x) = 0, \quad i \in I = \{1, \cdots, p\} \\ & & g_j(x) \leq 0, \quad j \in J = \{1, \cdots, m\} \\ & & x \in \mathcal{C}. \end{array}$$

where $x \in \mathbb{R}^n$, $\mathcal{C} \subseteq \mathbb{R}^n$ is a certain set and
$f, h_1, \cdots, h_p, g_1, \cdots, g_m$ are functions defined on $\mathcal{C}$.
Set of feasible solutions:

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

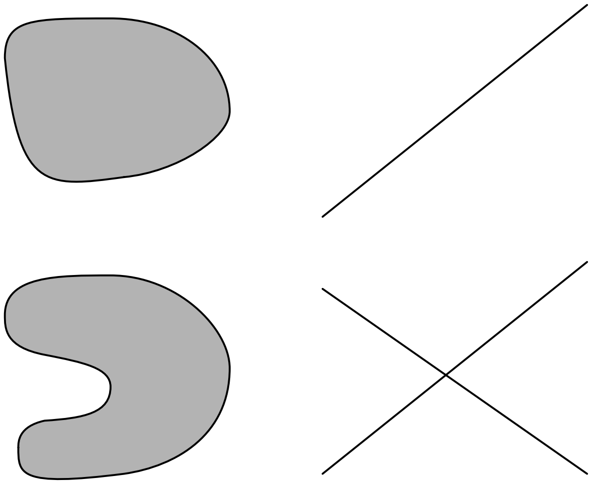
Definition 1 *Let two points* $x^1, x^2 \in \mathbb{R}^n$ *and* $0 \leq \lambda \leq 1$ *be given. Then the point*

$$x = \lambda x^1 + (1 - \lambda) x^2$$

is a convex combination *of the two points* x^1, x^2*.*

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

Convex and nonconvex sets in the plane.



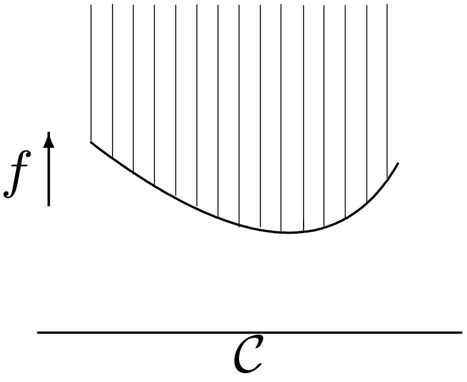
Convex functions

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

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

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

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



The epigraph of a convex function f.

For any set $\mathcal{S} \subset \mathbb{R}^n$ we can define a convex set, its convex hull in the following way.

Definition 4 *Let* $\mathcal{S} \subset \mathbb{R}^n$ *be an arbitrary set. The set*

$$\text{conv}(\mathcal{S}) := \{x \,|\, x = \lambda x^1 + (1-\lambda)x^2, \ x^1, x^2 \in \mathcal{S} \cup \text{conv}(\mathcal{S}), \ \lambda \in (0,1)\}$$

is called the convex hull *of the set* $\mathcal{S}$.

Affine hull

If $\mathcal{L}$ is a (linear) subspace, $a \in \mathbb{R}^n$ then $a + \mathcal{L}$ is called an *affine subspace* of $\mathbb{R}^n$. By definition, the dimension of $a + \mathcal{L}$ is the dimension of $\mathcal{L}$.

Definition 5 *The smallest affine space $a + \mathcal{L}$ containing a convex set $\mathcal{C} \subseteq \mathbb{R}^n$ is the so-called* affine hull *of $\mathcal{C}$ and denoted by* aff($\mathcal{C}$)*. The* dimension *of $\mathcal{C}$ is defined as the dimension of* aff($\mathcal{C}$)*.*

Definition 6 *Let two points $x^1, x^2 \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$ be given. Then the point*

$$x = \lambda x^1 + (1 - \lambda)x^2$$

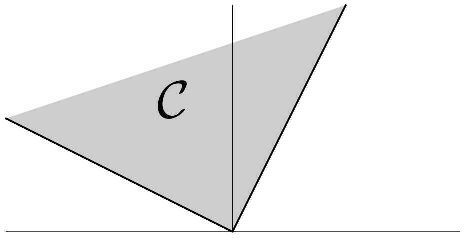
is an affine combination *of the two points x^1, x^2.*

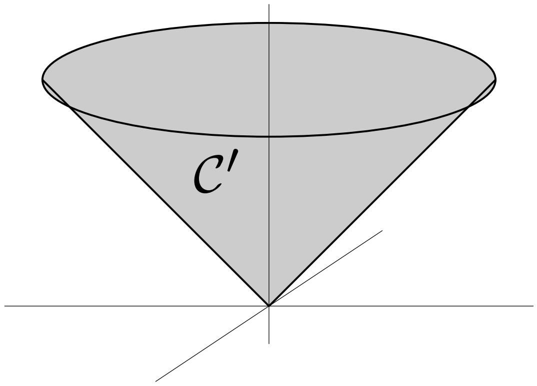
Convex Cones

Definition 7 *The set $\mathcal{C} \subset \mathbb{R}^n$ is a* convex cone *if it is a convex set and for all $x \in \mathcal{C}$ and $0 \le \lambda$ one has $\lambda x \in \mathcal{C}$.*

- The set $\mathcal{C} = \{(x_1, x_2) \in \mathbb{R}^2 |\ x_2 \ge 2x_1,\ x_2 \ge -\frac{1}{2}x_1\}$ is a convex cone in $\mathbb{R}^2$.

- The following set is a convex cone in $\mathbb{R}^3$: $\mathcal{C}' = \{(x_1, x_2, x_3) \in \mathbb{R}^3 |\ x_1^2 + x_2^2 \le x_3^2,\ x_3 \ge 0\}$.





Definition 8 *A convex cone is called* pointed *if it does not contain any subspace except the origin.*

A pointed closed convex cone could be defined equivalently as a convex cone that does not contain any line.

Lemma 1 *A convex cone $\mathcal{C}$ is pointed if and only if the origin 0 is an extremal point of $\mathcal{C}$.*

Recession Cone

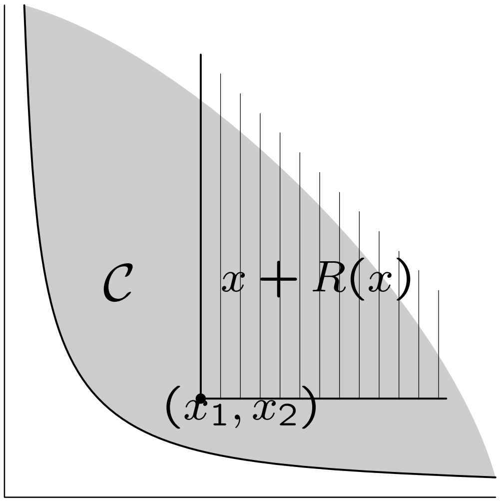
Lemma 2 *Let us assume that the convex set $\mathcal{C}$ is closed and not bounded. Then*

- (i) *for each $x \in \mathcal{C}$ there is a vector $z \in \mathrm{I\!R}^n$ such that $x+\lambda z \in \mathcal{C}$ for all $\lambda \geq 0$, i.e. the set $R(x) = \{z \mid x + \lambda z \in \mathcal{C},\ \lambda \geq 0\}$ is not empty;*
- (ii) *the set $R(x)$ is a closed convex cone (the so-called* recession cone *at x);*
- (iii) *the cone $R(x) = \mathcal{R}$ is independent of x, thus it is* 'the' *recession cone of the convex set $\mathcal{C}$;*
- (iv) *$\mathcal{R}$ is a pointed cone if and only if $\mathcal{C}$ has at least one extremal point.*

Corollary 1 *The nonempty closed convex set $\mathcal{C}$ is bounded if and only if its recession cone $\mathcal{R}$ consists of the zero vector alone.*

Let $\mathcal{C}$ be the epigraph of $f(x) = \frac{1}{x}$.
Then every point on $x_2 = \frac{1}{x_1}$ is an extreme point of $\mathcal{C}$. For $x = (x_1, x_2)$ the recession cone is given by

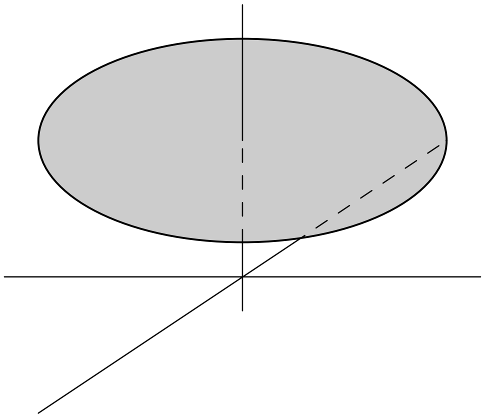
$$R(x) = \{z \in \mathrm{I\!R}^2 \mid z_1, z_2 \geq 0\}.$$



Relative Interior

Definition 9 *Let a convex set $\mathcal{C}$ be given. The point $x \in \mathcal{C}$ is in the* relative interior *of $\mathcal{C}$ if for all $\bar{x} \in \mathcal{C}$ there exists $\tilde{x} \in \mathcal{C}$ and $0 < \lambda < 1$ such that $x = \lambda\bar{x} + (1-\lambda)\tilde{x}$. The set of relative interior points of the set $\mathcal{C}$ will be denoted by $\mathcal{C}^0$.*

Let $\mathcal{C} = \{x \in \mathbb{R}^3 |\ x_1^2 + x_2^2 \leq 1, x_3 = 1\}$ and $\mathcal{L} = \{x \in \mathbb{R}^3 |\ x_3 = 0\}$, then $\mathcal{C} \subset \text{aff}(\mathcal{C}) = (0,0,1) + \mathcal{L}$. Hence, $\dim(\mathcal{C}) = 2$ and $\mathcal{C}^0 = \{x \in \mathbb{R}^3 |\ x_1^2 + x_2^2 < 1, x_3 = 1\}$.



Lemma 3 *Let $\mathcal{C} \subset \mathbb{R}^n$ be a convex set. Then for each $x \in \mathcal{C}^0$, $y \in \bar{\mathcal{C}}$ and $0 < \lambda \leq 1$ we have*

$$z = \lambda x + (1-\lambda)y \in \mathcal{C}^0 \subseteq \mathcal{C}.$$

Corollary 2 *The relative interior $\mathcal{C}^0$ of a convex set $\mathcal{C} \subset \mathbb{R}^n$ is convex.*

Lemma 4 *Let $\mathcal{C}$ be a convex set. Then $(\mathcal{C}^0)^0 = \mathcal{C}^0$. Moreover, if $\mathcal{C}$ is nonempty then its relative interior $\mathcal{C}^0$ is nonempty as well.*

Convex Functions

Lemma 5 *Let f be a convex function defined on the convex set $\mathcal{C}$. Then f is continuous on the relative interior $\mathcal{C}^0$ of $\mathcal{C}$.*

Lemma 6 (Jensen inequality) *Let f be a convex function defined on a convex set $\mathcal{C} \subseteq \mathbb{R}^n$. Let the points $x^1, \cdots, x^k \in \mathcal{C}$ and $\lambda^1, \cdots, \lambda^k \geq 0$ with $\sum_{i=1}^k \lambda^i = 1$ be given. Then*

$$f(\sum_{i=1}^{k} \lambda^i x^i) \leq \sum_{i=1}^{k} \lambda^i f(x^i).$$

Lemma 7 *Let $f^1, \cdots, f^k$ be a convex functions defined on a convex set $\mathcal{C} \subseteq \mathbb{R}^n$. Then*

- *for all $\lambda^1, \cdots, \lambda^k \geq 0$ the function*
  $$f(x) = \sum_{i=1}^{k} \lambda^i f^i(x)$$
  is convex;
- *the function*
  $$f(x) = \max_{1 \leq i \leq k} f^i(x)$$
  is convex.

Convex Functions cntd.

Definition 10 *The function* $h : \mathbb{R} \to \mathbb{R}$ *is called*

- *monotonically non-decreasing if for all* $t_1 < t_2 \in \mathbb{R}$ *one has* $h(t_1) \le h(t_2)$*;*
- *strictly monotonically increasing if for all* $t_1 < t_2 \in \mathbb{R}$ *one has* $h(t_1) < h(t_2)$*.*

Lemma 8 *Let* f *be a convex function on the convex set* $\mathcal{C} \subseteq \mathbb{R}^n$ *and* $h : \mathbb{R} \to \mathbb{R}$ *be a convex monotonically non-decreasing function. Then the composite function* $h(f(x)) : \mathcal{C} \to \mathbb{R}$ *is convex.*

Definition 11 *Let a convex function* $f : \mathcal{C} \to \mathbb{R}$ *defined on the convex set* $\mathcal{C}$ *be given. Let* $\alpha \in \mathbb{R}$ *be an arbitrary number. The set* $\mathcal{D}_\alpha = \{x \in \mathcal{C} \mid f(x) \le \alpha\}$ *is called a* level set *of the function* f*.*

Lemma 9 *If* f *is a convex function on* $\mathcal{C}$ *then for all* $\alpha \in \mathbb{R}$ *the level set* $\mathcal{D}_\alpha$ *is a (possibly empty) convex set.*

Convex Functions cntd.

Definition 12 *Let $x \in \mathbb{R}^n$ and a direction (vector) $s \in \mathbb{R}^n$ be given. The directional derivative $\delta f(x,s)$ of the function f, at point x, in the direction s, is defined as*

$$\delta f(x,s) = \lim_{\lambda \to 0} \frac{f(x+\lambda s) - f(x)}{\lambda}$$

if the above limit exists.

Lemma 10 *If the function f is continuously differentiable then for all $s \in \mathbb{R}^n$ we have*

$$\delta f(x,s) = \nabla f(x)^T s.$$

The Hesse matrix:

$$(\nabla^2 f(x))_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \quad \text{for all} \quad i,j = 1,\cdots,n.$$

Lemma 11 *Let f be a function defined on a convex set $\mathcal{C} \subseteq \mathbb{R}^n$. The function f is convex if and only if the function $\phi(\lambda) = f(x+\lambda s)$ is convex on the interval $[0,1]$ for all $x \in \mathcal{C}$ and $x + s \in \mathcal{C}$.*

Lemma 12 *Let f be a continuously differentiable function on the open convex set $\mathcal{C} \subseteq \mathbb{R}^n$. Then the following statements are equivalent.*

1. *The function f is convex on $\mathcal{C}$.*
2. *For any two vectors $x, \bar{x} \in \mathcal{C}$ one has*

$$\nabla f(x)^T(\bar{x} - x) \le f(\bar{x}) - f(x) \le \nabla f(\bar{x})^T(\bar{x} - x).$$

3. *For any $x, x+s \in \mathcal{C}$ the function $\phi(\lambda) = f(x+\lambda s)$ is continuously differentiable on the open interval $(0,1)$ and $\phi'(\lambda) = s^T \nabla f(x+\lambda s)$, which is a monotonically non-decreasing function.*

Proof of the Lemma

First we prove that *1* implies *2*. Let $0 \le \lambda \le 1$ and $x, \bar{x} \in \mathcal{C}$. Then the convexity of f imply

$$f(\lambda\bar{x} + (1-\lambda)x \le \lambda f(\bar{x}) + (1-\lambda)f(x).$$

This can be rewritten as

$$\frac{f(x + \lambda(\bar{x} - x)) - f(x)}{\lambda} \le f(\bar{x}) - f(x).$$

Taking the limit as $\lambda \to 0$ and applying Lemma 10 the left-hand-side inequality of *2* follows. As one interchanges the role x and $\bar{x}$, the right-hand-side inequality is obtained analogously.

Now we prove that *2* implies *3*. Let $x, x+s \in \mathcal{C}$ and $0 \le \lambda^1, \lambda^2 \le 1$. When we apply the inequalities of *2* with the points $x+\lambda^1 s$ and $x+\lambda^2 s$ the following relations are obtained.

$$(\lambda^2-\lambda^1)\nabla f(x+\lambda^1 s)^T s \le f(x+\lambda^2 s) - f(x+\lambda^1 s) \le (\lambda^2-\lambda^1)\nabla f(x+\lambda^2 s)^T s,$$

hence

$$(\lambda^2 - \lambda^1)\phi'(\lambda^1) \le \phi(\lambda^2) - \phi(\lambda^1) \le (\lambda^2 - \lambda^1)\phi'(\lambda^2).$$

Assuming $\lambda^1 < \lambda^2$ we have

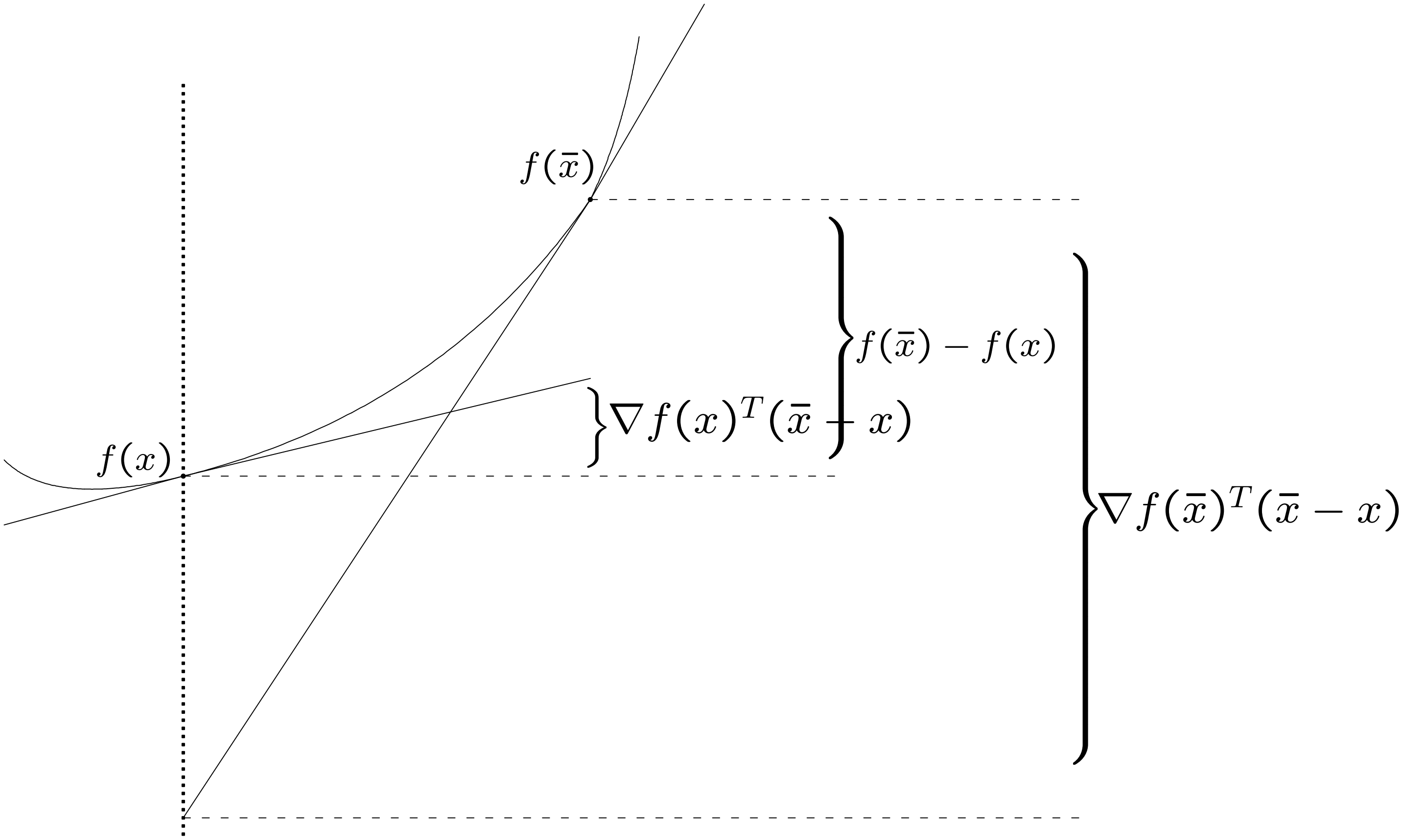
$$\phi'(\lambda^1) \le \frac{\phi(\lambda^2) - \phi(\lambda^1)}{\lambda^2 - \lambda^1} \le \phi'(\lambda^2)$$

proving that the function $\phi'(\lambda)$ is monotonically non-decreasing.

Finally we prove that *3* implies *1*. We only have to prove that $\phi(\lambda)$ is convex if $\phi'(\lambda)$ is monotonically non-decreasing. Let us take $0 < \lambda^1 < \lambda^2 < 1$ where $\phi'(\lambda^1) < \phi'(\lambda^2)$. Then for $0 \le \alpha \le 1$ we may write

$$\begin{aligned}
(1-\alpha)\phi(\lambda^1) \quad &+ \quad \alpha\phi(\lambda^2) - \phi((1-\alpha)\lambda^1 + \alpha\lambda^2) \\
&= \quad \alpha[\phi(\lambda^2) - \phi(\lambda^1)] - [\phi((1-\alpha)\lambda^1 + \alpha\lambda^2) - \phi(\lambda^1)] \\
&= \quad \alpha(\lambda^2 - \lambda^1)\left(\int_0^1 \phi'(\lambda^1 + t(\lambda^2 - \lambda^1))\mathrm{d}t \right. \\
&\qquad \left. - \int_0^1 \phi'(\lambda^1 + t\alpha(\lambda^2 - \lambda^1))\mathrm{d}t\right) \\
&\ge \quad 0.
\end{aligned}$$

Illustration of the Lemma



Lemma 13 *Let f be a twice continuously differentiable function on the open convex set $\mathcal{C} \subseteq \mathbb{R}^n$. The function f is convex if and only if its Hesse matrix $\nabla^2 f(x)$ is PSD for all $x \in \mathcal{C}$. Furthermore $\phi''(\lambda) = s^T \nabla^2 f(x + \lambda s)s$.*

Optimality Conditions

Unconstrained minimization

Consider the problem

$$\text{minimize} \quad f(x),$$

where $x \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable function. First we define local and global minima of the above problem.

Definition 13 *Let a function $f : \mathbb{R}^n \to \mathbb{R}$ 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 $\|\bar{x} - x\| \leq \epsilon$.*

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 $x \in \mathbb{R}^n$.*

Lemma 14 *Any (strict) local minimum of a convex function f is a (strict) global minimum of f as well.*

Optimality conditions

Lemma 15 *Let f be continuously differentiable. If the point $\bar{x} \in \mathbb{R}^n$ is a minimum of the function f then $\nabla f(\bar{x}) = 0$.*

Lemma 16 *Let f be a continuously differentiable convex function. The point $\bar{x} \in \mathbb{R}^n$ is a minimum of the function f if and only if $\nabla f(\bar{x}) = 0$.*

Lemma 17 *Let f be a twice continuously differentiable function. If at a point $\bar{x} \in \mathbb{R}^n$ it holds that $\nabla f(\bar{x}) = 0$ and $\nabla^2 f(x)$ is positive semidefinite in an $\epsilon-$neighborhood ($\epsilon > 0$) of $\bar{x}$ then the point $\bar{x}$ is a local minimum of the function f.*

Corollary 3 *Let f be a twice continuously differentiable function. If at $\bar{x} \in \mathbb{R}^n$ the gradient $\nabla f(\bar{x}) = 0$ and the Hessian $\nabla^2 f(\bar{x})$ is positive definite then the point $\bar{x}$ is a strict local minimum of the function f.*

Constrained Optimization

Theorem 1 *Let us consider the optimization problem* $\min\{\, f(x) \,:\, x \in \mathcal{C}\}$ *where* $\mathcal{C}$ *is a relatively open convex set and* f *is a convex differentiable function. The point* $\bar{x}$ *is an optimal solution of this problem if and only if* $\nabla f(\bar{x})^T s = 0$ *for all* $s \in \mathcal{L}$*, where* $\mathcal{L}$ *denotes the linear subspace with* $\mathrm{aff}(\mathcal{C}) = x + \mathcal{L}$ *for any* $x \in \mathcal{C}$*. Here* $\mathrm{aff}(\mathcal{C})$ *denotes the affine hull of* $\mathcal{C}$*.*

Proof: If $\bar{x}$ is a minimum, one has

$$f(\bar{x}) \le f(\bar{x} + \lambda s) \text{ for all } s \text{ as } \bar{x} + \lambda s \in \mathcal{C}.$$

Here $s \in \mathcal{L}$ and all the vectors $s \in \mathcal{L}$ can be defined this way since by assumption $\mathcal{C}$ is a relatively open set. By bringing $f(\bar{x})$ to the right hand side and dividing by λ we have

$$0 \le \frac{f(\bar{x} + \lambda s) - f(\bar{x})}{\lambda}.$$

Taking the limit as $\lambda \to 0$ results in

$$0 \le \delta f(\bar{x}, s) = \nabla f(\bar{x})^T s \quad \text{for all} \quad s \in \mathcal{L}.$$

As $s \in \mathcal{L}$ is arbitrary we conclude that $\nabla f(\bar{x})^T s = 0$ for all $s \in \mathcal{L}$.

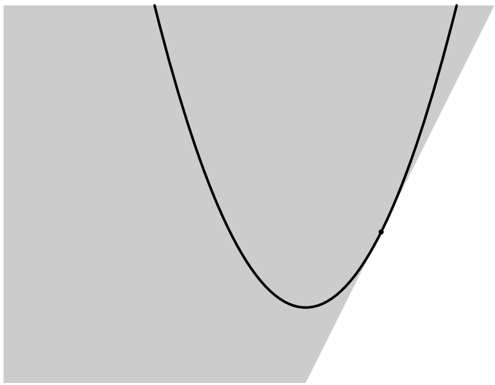
On the other hand, if f is a convex function and $\nabla f(\bar{x})^T s = 0$ for all $s \in \mathcal{L}$ then

$$f(x) - f(\bar{x}) \ge \nabla f(\bar{x})^T (x - \bar{x}) = 0$$

since $s = (x - \bar{x}) \in \mathcal{L}$, hence the theorem is proved. □

Feasible Directions

Definition 14 *The vector* $s \in \mathbb{R}^n$ *is called a feasible direction at a point* $x \in \mathcal{F}$ *if there is a* $\lambda_0 > 0$ *such that* $x + \lambda s \in \mathcal{F}$ *for all* $0 \le \lambda \le \lambda_0$*. The set of feasible directions at the feasible point* $x \in \mathcal{F}$ *is denoted by* $\mathcal{FD}(x)$



Lemma 18 *For any convex set* $\mathcal{F}$ *and for any* $x \in \mathcal{F}$ *the set of feasible directions* $\mathcal{FD}(x)$ *is a convex cone.*

Theorem 2 *The feasible point* $\bar{x} \in \mathcal{F}$ *is an optimal solution of the convex optimization problem (CO) if and only if for all* $s \in \mathcal{FD}(\bar{x})$ *one has* $\delta f(\bar{x}, s) \ge 0$*.*