3.1 Optimality conditions in unconstrained optimization

Recall the definitions of global, local minimizer.

“Geometry of minimization”

Consider for $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in C^1$ a point $\bar{x} \in \mathcal{D}_\alpha := \{ x \mid f(x) = \alpha \}$ with $\nabla f(\bar{x}) \neq 0$. Then:

- In a neighborhood of $\bar{x}$ the solution set $\mathcal{D}_\alpha$ is a $C^1$-manifold of dimension $n - 1$ and at $\bar{x}$ we have
  \[
  \nabla f(\bar{x}) \perp \mathcal{D}_\alpha
  \]
  i.e., $\nabla f(\bar{x})$ is perpendicular to the level set $\mathcal{D}_\alpha$. $\nabla f(\bar{x})$ points into the direction where $f(x)$ has increasing values.
Example: the ‘humpback function’

\[
\min x_1^2(4 - 2.1x_1^2 + \frac{1}{3}x_1^4) + x_1x_2 + x_2^2(-4 + 4x_2^2).
\]

Two global minima: (0.089 -0.717) and (-0.0898 0.717), and four strict local minima.
Th. 3.1 (Necessary conditions) [KRT, Th. 2.5]

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a $C^1$-function ($C^2$). If the point $\overline{x} \in \mathbb{R}^n$ is a local minimizer of the function $f$ then $\nabla f(\overline{x}) = 0$ (and $\nabla^2 f(\overline{x})$ is psd).

Example. The conditions $\nabla f(\overline{x}) = 0$, $\nabla^2 f(\overline{x})$ psd, are necessary but not sufficient minimality conditions. Take $f(x) = x^3$ or $\pm x^4$ as examples.

Th. 3.2 (Sufficient condition) [KRT, Cor. 2.8]

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a $C^2$-function. If $\nabla f(\overline{x}) = 0$ and $\nabla^2 f(\overline{x})$ is positive definite then the point $\overline{x}$ is a strict local minimizer of the function $f$. 
Global minimizer of convex functions

For convex functions the situation is “easier”

Lemma 3.3 \([\text{KRT,L.2.4}]\)

Let \(f : C \to \mathbb{R}, C \subset \mathbb{R}^n\) convex, be a convex function. Then a (strict) local minimizer of \(f\) is a (strict) global minimizer.

Ex.3.1 Prove Lemma 3.3.

Th.3.4 \([\text{KRT,Th.2.6}]\)

Let \(f : \mathbb{R}^n \to \mathbb{R}\) be a convex \(C^1\)-function. The point \(\bar{x} \in \mathbb{R}^n\) is a global minimizer of \(f\) if and only if \(\nabla f(\bar{x}) = 0\).
Example  (convex quadratic function)  With positive definite $Q$, $c \in \mathbb{R}^n$ consider for $x \in \mathbb{R}^n$ the quadratic function

$$f(x) := \frac{1}{2} x^T Q x + c^T x.$$  

Now, $f$ is convex and therefore $\bar{x}$ is a global minimizer iff

$$\nabla f(\bar{x}) = 0 \quad \text{or} \quad Q\bar{x} + c = 0$$

So the (unique) global minimizer is given by $\bar{x} = -Q^{-1}c$. 
We consider the convex optimization problem:

\[(CO) \quad \begin{align*}
\min & \quad f(x) \\
\text{s.t.} & \quad g_j(x) \leq 0, \quad j \in J \\
& \quad x \in C,
\end{align*} \]

where \( J = \{1, \ldots, m\} \),

- \( C \subseteq \mathbb{R}^n \) is a \textit{convex (closed) set};
- \( f, g_1, \ldots, g_m \) are convex functions on \( C \) (or on an open set containing \( C \)).
- \( \mathcal{F} \) denotes the set of feasible points (feasible set):

\[ \mathcal{F} = \{x \in C \mid g_j(x) \leq 0, \quad j = 1, \ldots, m\}. \]

and \( v(CO) \) denotes the min value of \( (CO) \).
Ex.3.2 Show that the feasible set $\mathcal{F}$ of (CO) is a convex set. Also the set of (global) minimizers is convex (possibly empty).

Ex.3.3 (General characterization of minimizers) Consider with convex $\mathcal{F} \subset \mathbb{R}^n$ and convex $C^1$-function $f$ the problem: $(P) \quad \min f(x) \quad \text{s.t. } x \in \mathcal{F}.

Show that $\overline{x} \in \mathcal{F}$ is a (global) minimizer of $(P)$ if and only if $\nabla f(\overline{x})^T (x - \overline{x}) \geq 0 \ \forall x \in \mathcal{F}$ holds.

Definition 3.5 With respect to (CO), an index $j \in J$ is called active in $\overline{x} \in \mathcal{F}$ if $g_j(\overline{x}) = 0$.

The set of all active indices in $\overline{x}$, also called active index set, is denoted by $J_{\overline{x}}$.

Rem. !! In optimality conditions for $\overline{x} \in \mathcal{F}$ only constraints $g_j \leq 0$ with $j \in J_{\overline{x}}$ will play a (real) role.
Let us assume: \((CO)\) with \(\mathcal{C} = \mathbb{R}^n\), \(f, g_j \in \mathcal{C}^1\) 
(in the rest of this subsection 3.2)

**Def.3.6**

The feasible point \(\bar{x} \in \mathcal{F}\) satisfies the Karush-Kuhn-Tucker (KKT) conditions if there exist *multipliers* \(\bar{y}_j, j \in J\) such that: \(\bar{y}_j \geq 0, \ \forall j \in J\) and

\[
\nabla f(\bar{x}) = - \sum_{j \in J} \bar{y}_j \nabla g_j(\bar{x}), \quad \bar{y}_j g_j(\bar{x}) = 0, \ \forall j \in J.
\]

The KKT-conditions can equivalently be given as

\[
\nabla f(x) = - \sum_{j \in J_x} \bar{y}_j \nabla g_j(x), \quad \bar{y}_j \geq 0, \quad \forall j \in J_x.
\]

**Th.3.7**  \([KRT,\text{Th.2.16}]\) *(Sufficient condition)*

If \(\bar{x} \in \mathcal{F}\) satisfies the KKT conditions then \(\bar{x}\) is a minimizer of \((CO)\) (with \(\mathcal{C} = \mathbb{R}^n\); \(f, g_j \in \mathcal{C}^1\)).
The KKT conditions are not necessary for minimality. Take as example:

$$\min \{ x \in \mathbb{R} : x^2 \leq 0 \}.$$  

The minimum occurs at $x = 0$ (which is the only feasible point!). But there is no $y$ such that

$$(KKT) \quad 1 = -y \cdot 0, \quad y \cdot 0 = 0, \quad y \geq 0.$$  

(The problem here is that the constraint $x^2 \leq 0$ does not allow a strictly feasible $\bar{x}$, i.e., $\bar{x}$ with $\bar{x}^2 < 0$.)

We will show later on that under mild extra conditions, e.g., the so-called Slater conditions, the KKT condition is necessary for optimality.
Example: (Application of the KKT-conditions)

(minimizing a linear function over a sphere)

Let \( a, c \in \mathbb{R}^n \), \( c \neq 0 \). Consider the problem

\[
\min_{x \in \mathbb{R}^n} \left\{ \sum_{i=1}^{n} c_i x_i \mid \sum_{i=1}^{n} (x_i - a_i)^2 \leq 1 \right\}.
\]

The objective function is linear (hence convex), and there is one constraint function \( g_1(x) (m = 1) \), which is convex. The KKT conditions give

\[
\begin{bmatrix}
  c_1 \\
  \vdots \\
  c_n
\end{bmatrix}
= -2\lambda
\begin{bmatrix}
  x_1 - a_1 \\
  \vdots \\
  x_n - a_n
\end{bmatrix},
\]

\[
\lambda \left( \sum_{i=1}^{n} (x_i - a_i)^2 - 1 \right) = 0, \quad \lambda \geq 0.
\]
\[ c \neq 0 \text{ implies } \lambda > 0 \text{ and } \sum_{i=1}^{n}(x_i - a_i)^2 = 1. \] We conclude that \( x - a = -\alpha c, \) for \( \alpha = 1/(2\lambda) > 0. \) Since \( \|x - a\| = 1, \) we have \( \alpha \|c\| = 1. \) So, the minimizing vector is

\[
x = a - \alpha c = a - \frac{c}{\|c\|}.
\]

Hence, the minimal value of the objective function is:

\[
c^T x = c^T a - \frac{c^T c}{\|c\|} = c^T a - \frac{\|c\|^2}{\|c\|} = c^T a - \|c\|.
\]
3.3 Convex Farkas Lemma

Remark  The so-called convex Farkas Lemma is the basis for strong duality theory in convex optimization. The proof of this lemma is based on the following Separation theorem:

Th.3.8,  [KRT, Th.2.23]  (Separation Theorem)

Let $\emptyset \neq U \subseteq \mathbb{R}^n$ be convex and $w \notin U$. Then there exists a separating hyperplane

$$H = \{x \mid a^T x = \alpha\}, \quad 0 \neq a \in \mathbb{R}^n, \quad \alpha \in \mathbb{R}$$

such that

- $a^T w \geq \alpha \geq a^T x \quad \forall x \in U$
- and $\alpha > a^T u_0$ for some $u_0 \in U$. 
Proof: Sketch in 3 steps. Case \( w \notin \partial U \): (\( \partial U \), boundary).
Then by [FKS,Th.10.1] it holds for some \( a \neq 0, \alpha \):

\[
a^T w > \alpha \geq a^T x \quad \forall x \in U
\]

Case \( w \in \partial U \): Then by [FKS,Th.10.2] it holds for some \( a \neq 0, \alpha \):

\[
a^T w = \alpha \geq a^T x \quad \forall x \in U
\]

It remains to show: (in this case \( w \in \partial U \))

\[
(\star) \quad \alpha > a^T u_0 \text{ for some } u_0 \in U.
\]

We reduce the construction to \( \text{aff } U \) see Extrapoofs.pdf.
Remark: An essential role in strong duality and for necessary optimality conditions plays the so-called Slater condition.

Recall With convex $f, g_j, j \in J := \{1, \ldots, m\}$ and convex $C \subset \mathbb{R}^n$

$$\text{(CO)} \quad \min f(x) \quad \text{s.t.} \quad g_j(x) \leq 0, \quad \forall j \in J, \quad x \in C$$

Def.3.9 [KRT,Def.2.18] $x^0 \in C^0$ is called a Slater point of (CO) if

$$g_j(x^0) < 0, \quad \forall j \text{ where } g_j \text{ is nonlinear},$$

$$g_j(x^0) \leq 0, \quad \forall j \text{ where } g_j \text{ is (affine) linear}.$$

We say that (CO) satisfies the Slater condition (constraint qualification) if a Slater point exists.

Definition Some constraints $g_j(x) \leq 0$ may take the value zero at all feasible points. Each such constraint can be replaced by the equality constraint $g_j(x) = 0$, without changing the feasible region. These constraints are called singular while the others are called regular.
Under the Slater condition, we define the index sets $J_s$ of singular- and $J_r$ of regular constraints:

$$J_s := \{ j \in J \mid g_j(x) = 0, \forall x \in \mathcal{F} \},$$

$$J_r := J \setminus J_s = \{ j \in J \mid g_j(x) < 0 \text{ for some } x \in \mathcal{F} \}.$$

**Remark:** Note, that if (CO) satisfies the Slater condition, then all singular constraints must be linear. Note also that an affine linear constraint $h(x) := a^T x + b = 0$ can equivalently be expressed by the two (singular) constraints: $g_1(x) := a^T x + b \leq 0$ and $g_2(x) := -(a^T x + b) \leq 0$.

**Def.3.10** [KRT,Def.2.20] A Slater point $x^* \in C^0$ is called an *Ideal Slater point* of (CO) if

$$g_j(x^*) < 0 \quad \text{for all } j \in J_r,$$

$$g_j(x^*) = 0 \quad \text{for all } j \in J_s.$$
Lemma 3.11, \([KRT,L.2.21]\)

If \((CO)\) has a Slater point then it also has an ideal Slater point \(x^* \in \mathcal{F}\).

**Ex.3.4** Show that an ideal Slater point is in the relative interior of \(\mathcal{F}\).

**Example.** (Slater regularity:) Consider the feasible set of:

\[
\begin{align*}
\text{min} & \quad f(x) \\
\text{s.t.} & \quad x_1^2 + x_2^2 \leq 4 \\
& \quad x_1 - x_2 \geq 2 \\
& \quad x_2 \geq -1 \\
& \quad C = \mathbb{R}^2.
\end{align*}
\]

The point \((1, -1)\) is a Slater point, but \textbf{not} an ideal Slater point.

The point \(\left(\frac{3}{2}, -\frac{3}{4}\right)\) is an ideal Slater point.
Observation: A number $a$ is a lower bound for the optimal value of $(CO)$, i.e., $a \leq v(CO)$, if and only if the system

$$\begin{align*}
    f(x) &< a \\
    g_j(x) &\leq 0, \quad j = 1, \ldots, m \\
    x &\in C
\end{align*}$$

has no solution.

**Th.3.12 [KRT,L.2.25] (Farkas Lemma)**

Let $(CO)$ satisfy the Slater condition. Then the inequality system (1) has no solution if and only there exists a vector $y = (y_1, \ldots, y_m)$ such that

$$\begin{align*}
    y &\geq 0, \quad \text{and} \quad f(x) + \sum_{j=1}^{m} y_j g_j(x) \geq a, \quad \forall x \in C.
\end{align*}$$

The systems (1) and (2) are called **alternative systems**, because exactly one of them has a solution.
Proof of the (convex) Farkas lemma

We need to show that precisely one of the systems (1) or (2) has a solution.

- First we show (easy part) that not both systems can have a solution:
  Otherwise for some \( x \in C \) and \( y \geq 0 \), the relation

  \[
  a \leq f(x) + \sum_{j=1}^{m} y_j g_j(x) \leq f(x) < a,
  \]

  would hold, a contradiction.

- Then we show that at least one of the two systems has a solution.
  To do so, we show that if (1) has no solution, then (2) has a solution.
Proof: Define the set $\mathcal{U} \subset \mathbb{R}^{m+1}$ as follows.

$$\mathcal{U} = \{ u = (u_0; \ldots; u_m) \mid \exists x \in \mathcal{C} \text{ such that}$$

$$\begin{cases}
    f(x) < a + u_0 \\
    g_j(x) \leq u_j \text{ if } j \in J_r \\
    g_j(x) = u_j \text{ if } j \in J_s
\end{cases}$$

Since (1) is infeasible, $0 \notin \mathcal{U}$. One easily checks that $\mathcal{U}$ is a nonempty convex set (using that the singular functions are linear). So there exists a hyperplane that separates 0 from $\mathcal{U}$, i.e., there exists a nonzero vector $y = (y_0, y_1, \ldots, y_m)$ such that

$$y^T u \geq 0 = y^T 0 \text{ for all } u \in \mathcal{U}$$

$$y^T \bar{u} > 0 \text{ for some } \bar{u} \in \mathcal{U}. \quad (\ast)$$

The rest of the proof is divided into four parts.
I. Prove that $y_0 \geq 0$ and $y_j \geq 0$ for all $j \in J_r$.

II. Establish that

$$y_0(f(x) - a) + \sum_{j=1}^{m} y_j g_j(x) \geq 0, \quad \forall x \in C.$$ 

holds

III. Prove that $y_0$ must be positive.

IV. Show by induction that we can assume $y_j > 0, \quad \forall j \in J_s$. 

Ex.3.5 ("Linear Farkas Lemma") Let \( A \in \mathbb{R}^{m \times n} \) and \( b \in \mathbb{R}^m \).

Exactly one of the following alternative systems (I) or (II) is solvable:

(I) \[ A^T x \geq 0, \quad x \geq 0, \quad b^T x < 0, \]

or

(II) \[ Ay \leq b, \quad y \geq 0. \]

**Hint for the proof:** Apply Farkas’ lemma to the special case:

\( f(x) = b^T x, \)

\[ g_j(x) = -(A^T x)_j = - \left( a^j \right)^T x, \quad j = 1, \cdots, n \]

where \( a^j \) denotes column \( j \) of \( A \), and \( C \) is the nonnegative orthant, \( C := \{ z \in \mathbb{R}^m \mid z \geq 0 \} \). Make use of:

\[ z^T x = \sum_{i} z_i x_i \geq 0 \forall x \geq 0 \iff z \geq 0 \quad (\star) \]