Prof. Dr. Karsten Urban
M. Sc. Lewin Ernst

Winter term 2024/2025
Sheet 07, 30 Points

Numerical Optimization
Solution to exercise sheet

review on 04.12.2024 during the exercise class

1. *(Slater Condition)*
Let $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^p$ and $h : \mathbb{R}^n \to \mathbb{R}^m$ and let f and g be continuously differentiable. Consider a *convex* optimization problem of the form

$$\begin{cases} \min\limits_{x\in\mathbb{R}^n} & f(x) \\ \text{s.t.} & h(x) = 0, \\ & g(x) \leq 0, \end{cases} \tag{1}$$

where $x^* \in \mathbb{R}^n$ denotes a solution. We say that Problem (1) satisfies the *regularity condition of Slater*, if

$$\mathring{\mathcal{F}} := \{x \in \mathbb{R}^n : g(x) < 0,\ h(x) = 0\} \neq \emptyset.$$

Prove: If the Problem (1) satisfies the regularity condition of Slater, then there exist Lagrange multipliers $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ such that $(x^*, \lambda^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ is a KKT point of (1).

Hint:

- Show that the Slater Condition implies the Abadie Constraint Qualification for arbitrary $\hat{x} \in \mathcal{F}$. Then by Theorem 3.1.17 the KKT conditions hold. In order to do so, show
  i) $\mathcal{T}_{\text{lin}}(\mathcal{F}, \hat{x}) \subseteq \overline{\mathring{\mathcal{T}}_{\text{lin}}(\mathcal{F}, \hat{x})}$, where
  $$\mathring{\mathcal{T}}_{\text{lin}}(\mathcal{F}, \hat{x}) := \{d \in \mathbb{R}^n : \nabla g_j(\hat{x})^\top d < 0, j \in \mathcal{A}(\hat{x}); \nabla h(\hat{x})^\top d = 0\}$$
  and
  ii) $\mathring{\mathcal{T}}_{\text{lin}}(\mathcal{F}, \hat{x}) \subseteq \mathcal{T}(\mathcal{F}, \hat{x})$.
  From that it follows
  $$\mathcal{T}_{\text{lin}}(\mathcal{F}, \hat{x}) \subseteq \overline{\mathring{\mathcal{T}}_{\text{lin}}(\mathcal{F}, \hat{x})} \subseteq \overline{\mathcal{T}(\mathcal{F}, \hat{x})} = \mathcal{T}(\mathcal{F}, \hat{x})$$
  and by Lemma 3.1.10 we have the other inclusion as well. Therefore we have ACQ.

(14 Points)

Solution: It is sufficient to show that the Slater Condition implies the Abadie Constraint Qualification, since the claim then follows by Theorem 3.1.17. To show the Abadie Constraint Qualification we have to show

$$\mathcal{T}_{\text{lin}}(\mathcal{F}, \hat{x}) = \mathcal{T}(\mathcal{F}, \hat{x}),$$

where $\hat{x} \in \mathcal{F}$. By Lemma 3.1.10, we already have one set inclusion, hence we only need to show

$$\mathcal{T}_{\text{lin}}(\mathcal{F}, \hat{x}) \subseteq \mathcal{T}(\mathcal{F}, \hat{x}).$$

Recall that by Definition 3.1.9 it is

$$\mathcal{T}_{\text{lin}}(\mathcal{F},\hat{x}) = \{d \in \mathbb{R}^n : \nabla g_j(\hat{x})^\top d \leq 0, j \in \mathcal{A}(\hat{x}); \nabla h(\hat{x})^\top d = 0\}.$$

First, we define the set

$$\mathring{\mathcal{T}}_{\text{lin}}(\mathcal{F},\hat{x}) := \{d \in \mathbb{R}^n : \nabla g_j(\hat{x})^\top d < 0, j \in \mathcal{A}(\hat{x}); \nabla h(\hat{x})^\top d = 0\}.$$

We show the following inclusions:

i) $\mathcal{T}_{\text{lin}}(\mathcal{F},\hat{x}) \subseteq \overline{\mathring{\mathcal{T}}_{\text{lin}}(\mathcal{F},\hat{x})}$

ii) $\mathring{\mathcal{T}}_{\text{lin}}(\mathcal{F},\hat{x}) \subseteq \mathcal{T}(\mathcal{F},\hat{x})$

Proof of i). Let $d \in \mathcal{T}_{\text{lin}}(\mathcal{F},\hat{x})$. We are going to construct a sequence $(d_\epsilon)_{\epsilon>0} \subset \mathring{\mathcal{T}}_{\text{lin}}(\mathcal{F},\hat{x})$ convering to d as $\epsilon \to 0$, such that we have $d \in \overline{\mathring{\mathcal{T}}_{\text{lin}}(\mathcal{F},\hat{x})}$.
Lets define this sequence. Since the regularity condition of Slater is satisfied,

$$d_\varepsilon := d + \varepsilon(y - \hat{x}) \qquad \text{for } \varepsilon > 0$$

is well-defined for some

$$y \in \mathring{\mathcal{F}} = \{x \in \mathbb{R}^n : g(x) < 0, h(x) = 0\} \neq \emptyset.$$

For $j \in \mathcal{A}(\hat{x})$ we have

$$\begin{aligned}
\nabla g_j(\hat{x})^\top d_\varepsilon &= \nabla g_j(\hat{x})^\top (d + \varepsilon(y - \hat{x})) \\
&= \underbrace{\nabla g_j(\hat{x})^\top d}_{\leq 0} + \nabla g_j(\hat{x})^\top (\varepsilon(y - \hat{x})) \qquad \text{convexity (Theorem 2.1.8)} \\
&\leq \varepsilon(\underbrace{g_j(y)}_{<0} - \underbrace{g_j(\hat{x})}_{=0}) < 0.
\end{aligned}$$

Clearly it also holds

$$\begin{aligned}
\nabla h_z(\hat{x})^\top d_\varepsilon &= \nabla h_z(\hat{x})^\top (d + \varepsilon(y - \hat{x})) \\
&= \underbrace{\nabla h_z(\hat{x})^\top d}_{=0} + \varepsilon \nabla h_z(\hat{x})^\top (y - \hat{x}) \\
&= \varepsilon(\nabla h_z(\hat{x})^\top y - \nabla h_z(\hat{x})^\top \hat{x}),
\end{aligned}$$

for $z \in \{1, \ldots, m\}$. Since h is affine linear, we have $h_z(y) = \nabla h_z(\hat{x})^\top y + h_z(0)$. As $y \in \mathring{\mathcal{F}}$, it holds $h_z(y) = \nabla h_z(\hat{x})^\top y + h_z(0) = 0$ which means $\nabla h_z(\hat{x}) y^\top = -h_z(0)$. Similarly we have $\nabla h_z(\hat{x})^\top \hat{x} = -h_z(0)$, thus we obtain

$$= \varepsilon(-h_z(0) + h_z(0)) = 0 \quad \Rightarrow \nabla h(\hat{x})^\top d_\varepsilon = 0.$$

Therefore we have $d_\varepsilon \in \mathring{\mathcal{T}}_{\text{lin}}(\mathcal{F},\hat{x})$ by definition. Taking the limit we have found a sequence that converges to d, since $\lim_{\varepsilon \to 0^+} d_\varepsilon = d$, hence $d \in \overline{\mathring{\mathcal{T}}_{\text{lin}}(\mathcal{F},\hat{x})}$, therefore we have shown

$$\mathcal{T}_{\text{lin}}(\mathcal{F},\hat{x}) \subseteq \overline{\mathring{\mathcal{T}}_{\text{lin}}(\mathcal{F},\hat{x})}.$$

$\square$

Proof of ii). First recall that the tangential cone is given by

$$T(\mathcal{F},\hat{x}) = \{\bar{d} \in \mathbb{R}^n : \exists (\eta^{(l)})_{l\in\mathbb{N}} \subset \mathbb{R}^+, (x^{(l)})_{l\in\mathbb{N}} \subset \mathcal{F} : \lim_{l\to\infty} x^{(l)} = \hat{x}, \lim_{l\to\infty} \eta^{(l)}(x^{(l)} - \hat{x}) = \bar{d}\}.$$

Let $d \in \overset{\circ}{\mathcal{T}}_{\text{lin}}(\mathcal{F},\hat{x})$. Define the sequences by $x^{(l)} := \hat{x} + \frac{1}{l}d$ and $\eta^{(l)} := l$, then we have

$$\lim_{l\to\infty} x^{(l)} = \hat{x} \quad \text{and} \quad \lim_{l\to\infty} \eta^{(l)}(x^{(l)} - \hat{x}) = d.$$

Clearly it is $(\eta^{(l)})_{l\in\mathbb{N}} \subset \mathbb{R}^+$, thus it only remains to show that $(x^{(l)})_{l\in\mathbb{N}} \subset \mathcal{F}$. Observe that for $j \in \mathcal{A}(\hat{x})$, we can apply the mean value theorem to the continuous differentiable function g_j at $x^{(l)}$ and $\hat{x}$. This gives us some $y^{(l)} \in \overline{x^{(l)}\hat{x}}$ such that

$$\begin{aligned} & g_j(x^{(l)}) - g_j(\hat{x}) = \nabla g_j(y^{(l)})^\top (x^{(l)} - \hat{x}) \\ \Longrightarrow \qquad & g_j(x^{(l)}) = g_j(\hat{x}) + \nabla g_j(y^{(l)})^\top \left(\frac{1}{l}d\right). \end{aligned}$$

Since $y^{(l)} \in \overline{x^{(l)}\hat{x}}$ and $\lim_{l\to\infty} x^{(l)} = \hat{x}$, we also have $\lim_{l\to\infty} y^{(l)} = \hat{x}$ by the sandwich theorem. Thus we obtain for large l

$$g_j(x^{(l)}) = \underbrace{g_j(\hat{x})}_{=0, \text{ as } j\in\mathcal{A}(\hat{x})} + \frac{1}{l} \underbrace{\nabla g_j(y^{(l)})^\top(d)}_{\longrightarrow \nabla g_j(\hat{x})^\top d < 0, \text{ by choice of } d} \leq 0.$$

Similarly, for $j \in \{1,\dots,m\} \setminus \mathcal{A}(\hat{x})$ we have

$$g_j(x^{(l)}) = \underbrace{g_j(\hat{x})}_{<0} + \underbrace{\frac{1}{l}\nabla g_j(y^{(l)})^\top(d)}_{\longrightarrow \frac{1}{l}\nabla g_j(\hat{x})^\top d \longrightarrow 0} \leq 0.$$

Finally, we see that that

$$h(x^{(l)}) = \nabla h(\hat{x})^\top \left(\hat{x} + \frac{1}{l}d\right) + h(0) = \underbrace{\nabla h(\hat{x})^\top \hat{x} + h(0)}_{=h(\hat{x})=0} + \frac{1}{l} \underbrace{\nabla h(\hat{x})^\top d}_{=0 \text{ by choice of } d} = 0.$$

Therefore, when we choose l large enough $x^{(l)}$ is feasible and the inclusion is shown. ☐

2. *(Linear independence constraint qualification (LICQ Condition))*
   Let $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^p$ and $h : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable functions. An optimization problem of the form

$$\begin{cases} \min\limits_{x\in\mathbb{R}} & f(x) \\ \text{s.t.} & h(x) = 0 \\ & g(x) \leq 0 \end{cases} \tag{2}$$

satisfies the *LICQ condition* in a feasible point $\hat{x} \in \mathcal{F} \subset \mathbb{R}^n$, if $\nabla g_i(\hat{x}) \in \mathbb{R}^n$ and $\nabla h_j(\hat{x}) \in \mathbb{R}^n$ are linear independent for all $i \in \mathcal{A}(\hat{x})$ and for all $j = 1,\dots,m$.

Prove: If a local solution $x^* \in \mathcal{F}$ satisfies the LICQ condition, than the Lagrange multipliers $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ at a KKT point $(x^*,\lambda^*,\mu^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ of (2) are unique.

(6 Points)

Solution: We assume there exit two KKT points (x^*, λ^*, μ^*) and $(x^*, \overline{\lambda}, \overline{\mu})$ of the optimization problem (2) with $(\lambda^*, \mu^*) \neq (\overline{\lambda}, \overline{\mu})$. This means the KKT conditions of Theorem 3.1.17 (first order optimality conditions) hold for these points. Further, let the LICQ condition be fulfilled at x^*.

We insert both KKT points in the "multiplier rule". For (x^*, λ^*, μ^*) we get

$$\nabla f(x^*) + \sum_{i \in \mathcal{A}(x^*)} \mu_i^* \nabla g_i(x^*) + \sum_{j=1}^{m} \lambda_j^* \nabla h_j(x^*) = 0. \tag{3}$$

Analogously, for $(x^*, \overline{\lambda}, \overline{\mu})$ it is

$$\nabla f(x^*) + \sum_{i \in \mathcal{A}(x^*)} \overline{\mu}_i \nabla g_i(x^*) + \sum_{j=1}^{m} \overline{\lambda}_j \nabla h_j(x^*) = 0. \tag{4}$$

Next, we subtract (4) from (3) to get

$$\sum_{i \in \mathcal{A}(x^*)} (\mu_i^* - \overline{\mu}_i) \nabla g_i(x^*) + \sum_{j=1}^{m} (\lambda_j^* - \overline{\lambda}_j) \nabla h_j(x^*) = 0. \tag{5}$$

For satisfying the LICQ condition it is required that $\nabla g_i(x^*)$ and $\nabla h_j(x^*)$ are linear independent for all $i \in \mathcal{A}(x^*)$ and for all $j = 1, ..., m$, which is:

$$\sum_{i \in \mathcal{A}(x^*)} a_i \nabla g_i(x^*) + \sum_{j=1}^{m} b_j \nabla h_j(x^*) = 0 \quad \Rightarrow a_i = 0,\ i \in \mathcal{A}(x^*),\ \text{and } b_j = 0,\ j = 1, ..., m. \tag{6}$$

It follows $(\lambda^*, \mu^*) = (\overline{\lambda}, \overline{\mu})$.

3. *(Another condition and the relation of the CQs, Mangasarian-Fromovitz constraint qualification (MFCQ))*

   Let $\hat{x} \in \mathcal{F}$. We say that the MFCQ holds at $\hat{x}$ if the gradients

   $$\nabla h_j(\hat{x}), \quad j = 1, \ldots, m$$

   are linear independent and there exists a vector $d \in \mathbb{R}^n$ such that

   $$\nabla g_i(\hat{x})^T d < 0, \quad i \in \mathcal{A}(x), \quad \nabla h(\hat{x})^T d = 0.$$

   One can show: If $x \in \mathcal{F}$ fulfills MFCQ, then ACQ holds. Moreover we have:

   Theorem 1. *Let $x \in \mathcal{F}$ be given. Then the following implications hold*

   $$\begin{array}{ccccccc} \mathit{LICQ}(x) & \Rightarrow & \mathit{MFCQ}(x) & \Rightarrow & \mathit{ACQ}(x) & \Rightarrow & \mathit{GCQ}(x) \\ & & & & \Uparrow \textit{Convex problems} & & \\ & & & & \mathit{Slater} & & \end{array}$$

   Remark: Note that the Slater condition implies ACQ(x) for all $x \in \mathcal{F}$.

   Prove: LICQ(x) $\Rightarrow$ MFCQ(x).

   (4 Points)

Solution: Obviously, the gradients $\nabla h_j(\hat{x})$ for $j = 1, \dots, m$ are linear independent due to LICQ. It remains to find a suitable vector $d \in \mathbb{R}^n$ such that

$$\nabla g_i(\hat{x})^T d < 0, \quad i \in \mathcal{A}(x), \quad \nabla h(\hat{x})^T d = 0.$$

hold. Lets define the matrix

$$\begin{pmatrix} \nabla g_i(\hat{x})^T & i \in \mathcal{A}(\hat{x}) \\ \nabla h_j(\hat{x})^T & j = 1, \dots, m \end{pmatrix} \in \mathbb{R}^{(|\mathcal{A}(\hat{x})|+m)\times n}$$

which has full rank by LICQ. Hence we can add rows to obtain a non-singular matrix $A(\hat{x}) \in \mathbb{R}^{n\times n}$. Then, the linear system

$$A(\hat{x})d = \begin{pmatrix} -e \\ 0 \end{pmatrix},$$

where $e = (1, \dots, 1)^T \in \mathbb{R}^{|\mathcal{A}(\hat{x})|}$ has a solution $\hat{d} \in \mathbb{R}^n$ which fulfills the requirements of MFCQ.

4. *(LICQ, MFCQ and Slater)*
We consider the following optimization problem with $p = 4$ constraints

$$\min_{x\in\mathbb{R}^2} \; x_1^2 + x_2^2 \tag{7}$$
$$\text{s.t.} \quad \begin{aligned} x_1^2 + 4x_2^2 &\le 4, \\ (x_1 - 2)^2 + x_2^2 &\le 5, \\ x_1,\ x_2 &\ge 0. \end{aligned}$$

(a) Check, if $\hat{x} := (0, 1)^T \in \mathbb{R}^2$ fulfills LICQ and MFCQ.

(b) Prove that the minimum $\bar{x}$ of (7) fulfills the KKT conditions.

(3 + 3 = 6 Points)

Solution:

(a) The constraints and the Jacobian matrix of the constraints are

$$g(x) = \begin{pmatrix} x_1^2 + 4x_2^2 - 4 \\ (x_1 - 2)^2 + x_2^2 - 5 \\ -x_1 \\ -x_2 \end{pmatrix}, \quad g(\hat{x}) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \quad g'(x) = \begin{pmatrix} 2x_1 & 8x_2 \\ 2x_1 - 4 & 2x_2 \\ -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad g'(\hat{x}) = \begin{pmatrix} 0 & 8 \\ -4 & 2 \\ -1 & 0 \\ 0 & -1 \end{pmatrix},$$

where the first three constraints are active at $\hat{x}$, i.e. $\mathcal{A}(\hat{x}) = \{1, 2, 3\}$.

LICQ: Because $|\mathcal{A}(\hat{x})| = 3$ constraints are active, but there are only $n = 2$ variables, the gradients of the contraints can not be linear independent. Therefore the LICQ can not hold.

MFCQ: We do not have equality constraints so that for MFCQ we need to have $v \in \mathbb{R}^2$ such that

$$\begin{pmatrix} 0 & 8 \\ -4 & 2 \\ -1 & 0 \end{pmatrix} v < \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

The vector $v = (1, -1)^T$ fulfills this, which is readily seen

$$\begin{pmatrix} 0 & 8 \\ -4 & 2 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -8 \\ -6 \\ -1 \end{pmatrix}.$$

(b) The problem (7) is convex (only linear and quadradic inequality constraints), therefore we want test if the Slater condition is fulfilled. We consider the point $x = (1,\, 0.5)^T$, for which $x \in \overset{\circ}{\mathcal{F}}$, i.e. $\overset{\circ}{\mathcal{F}} \neq \emptyset$, due to

$$g(x) = \begin{pmatrix} x_1^2 + 4x_2^2 - 4 \\ (x_1 - 2)^2 + x_2^2 - 5 \\ -x_1 \\ -x_2 \end{pmatrix} = \begin{pmatrix} -2 \\ -3.75 \\ -1 \\ -0.5 \end{pmatrix} < \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Therefore there are Lagrange parameters $\bar{\lambda}$ and $\bar{\mu}$, such that $(\bar{x}, \bar{\lambda}, \bar{\mu})$ fulfills the KKT-conditions.

Remark: To check if a CQ is fulfilled in x the function f is irrelevant. We could have taken any other convex function f and nothing would have changed. The objective function f only becomes relevant if we try to find a minimum.