E. Herberg, M. Marić, V. Stein
Heidelberg University



# Exercise 12 - Solution

Date issued: 1st July 2024
Date due: 9th July 2024

**Homework Problem 12.1** (CQs are invariant under Slack Transformation) 10 Points

We can reformulate the original nonlinear problem

$$\left.\begin{array}{rll} \text{Minimize} & f(x) & \text{where } x \in \mathbb{R}^n \\ \text{subject to} & g_i(x) \le 0 & \text{for } i = 1, \dots, n_{\text{ineq}} \\ \text{and} & h_j(x) = 0 & \text{for } j = 1, \dots, n_{\text{eq}} \end{array}\right\} \tag{5.1}$$

by introducing a so called **slack variable** $s \in \mathbb{R}^{n_{\text{ineq}}}$ to obtain the simple one-sided box-constrained problem

$$\left.\begin{array}{rll} \text{Minimize} & f(x) & \text{where } (x, s) \in \mathbb{R}^{n \times n_{\text{ineq}}} \\ \text{subject to} & g_i(x) + s = 0 & \text{for } i = 1, \dots, n_{\text{ineq}} \\ \text{and} & -s \le 0 & \\ \text{and} & h_j(x) = 0 & \text{for } j = 1, \dots, n_{\text{eq}} \end{array}\right\}. \tag{5.1$_\text{s}$}$$

(*i*) Derive the KKT-system of (5.1$_\text{s}$) and show that there is a one-to-one connection between the solutions of the KKT systems corresponding to (5.1) and (5.1$_\text{s}$).

(*ii*) Show that GCQ/ACQ/MFCQ/LICQ is satisfied at a feasible $(x, s)$ for (5.1$_\text{s}$) if the respective condition is satisfied at $x$ for (5.1).

For which CQs can you show equivalence?

**Solution.**

($i$) All quantities corresponding to the slacked system (KKT$_s$) will be denoted with a tilde, e. g., for (5.1$_s$), we set the constraints $\widetilde{g}\colon \mathbb{R}^{n\times n_{\text{ineq}}} \longrightarrow \mathbb{R}^{n_{\text{ineq}}}$ and $\widetilde{h}\colon \mathbb{R}^{n\times n_{\text{ineq}}} \longrightarrow \mathbb{R}^{n_{\text{eq}}+n_{\text{ineq}}}$ as

$$\begin{aligned} \widetilde{g}(x,s) &:= -s & \text{with} \quad & \widetilde{g}'(x,s) = \begin{bmatrix} 0 & -\mathrm{Id} \end{bmatrix} \\ \widetilde{h}(x,s) &:= \begin{bmatrix} h(x) \\ g(x)+s \end{bmatrix} & \text{with} \quad & \widetilde{h}'(x,s) = \begin{bmatrix} h'(x) & 0 \\ g'(x) & \mathrm{Id} \end{bmatrix} \end{aligned}$$

and denote the corresponding feasible set by $\widetilde{F} \subseteq \mathbb{R}^{n\times n_{\text{ineq}}}$.

The KKT conditions are

$$\begin{aligned} \nabla f(x) + g'(x)^\top \mu + h'(x)^\top \lambda &= 0, \\ \mu \geq 0, \quad g(x) \leq 0, \quad \mu^\top g(x) &= 0 \\ h(x) &= 0, \end{aligned} \tag{KKT}$$

and

$$\begin{aligned} \nabla f(x) + \widetilde{g}'(x,s)^\top \widetilde{\mu} + \widetilde{h}'(x,s)^\top \widetilde{\lambda} &= 0, \\ \widetilde{\mu} \geq 0, \quad \widetilde{g}(x,s) \leq 0, \quad \widetilde{\mu}^\top \widetilde{g}(x,s) &= 0 \\ \widetilde{h}(x,s) &= 0, \end{aligned} \tag{KKT$_s$}$$

for multipliers $\widetilde{\mu} \in \mathbb{R}^n_{\text{ineq}}$ and $\widetilde{\lambda} = (\widetilde{\lambda}_x, \widetilde{\lambda}_s) \in \mathbb{R}^{n_{\text{eq}}+n_{\text{ineq}}}$ respectively, where (KKT$_s$) expands to the system

$$\begin{aligned} \begin{bmatrix} \nabla_x f(x) \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ -\widetilde{\mu} \end{bmatrix} + \begin{bmatrix} h'(x)^\top \widetilde{\lambda}_x + g'(x)\widetilde{\lambda}_s \\ \widetilde{\lambda}_s \end{bmatrix} &= 0, \\ \widetilde{\mu} \geq 0, \quad -s \leq 0, \quad \widetilde{\mu}^\top s &= 0 \\ g(x) = -s, \; h(x) &= 0, \end{aligned} \tag{KKT$_s^*$}$$

showing the one to one correspondence of the KKT solutions, since we can always replace $g(x)$ with $s$ and identify the multipliers $\mu = \widetilde{\mu} = \widetilde{\lambda}_s$ and $\widetilde{\lambda}_x = \lambda$. (2 Points)

($ii$) First off, note that if we define the map

$$\Phi\colon \mathbb{R}^n \longrightarrow \mathbb{R}^{n+n_{\text{ineq}}}, \quad \Phi(x) := \begin{pmatrix} x \\ -g(x) \end{pmatrix} \quad \text{with} \quad \Phi'(x) = \begin{bmatrix} \mathrm{Id} \\ -g'(x) \end{bmatrix}$$

then clearly $\Phi(F) = \widetilde{F}$ and since $\Phi$ is injective, it is invertible on its image, i. e., we can define

$$\Phi^{-1}(x,s) = x$$

as the (right) inverse on the image (this is not a full inverse because $\Phi$ is clearly not surjective!). Additionally, we immediately see that $\mathcal{A}(x) = \widetilde{\mathcal{A}(\Phi(x))}$.

(a) For LICQ, we observe that

$$\begin{bmatrix} \widetilde{g}_i{}'(x,s)|_{i\in\widetilde{\mathcal{A}(x,s)}} \\ \widetilde{h}_j{}'(x,s)|_{j=1,\dots,n_{\text{eq}}} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 0 & -\mathrm{Id} \end{bmatrix}_{i\in\widetilde{\mathcal{A}(x,s)}} \\ \begin{bmatrix} h'(x) & 0 \\ g'(x) & \mathrm{Id} \end{bmatrix} \end{bmatrix},$$

which has full row-rank if and only if

$$\begin{bmatrix} g_i'(x)|_{i\in\mathcal{A}(x)} \\ h_j'(x)|_{j=1,\dots,n_{\text{eq}}} \end{bmatrix}$$

does, so LICQ holds at $x$ for (5.1) if and only if it holds at $(x,s) = \Phi(x)$ for ($5.1_{\text{s}}$).

(b) For MFCQ, we observe that

$$\widetilde{h}{}'(x,s) = \begin{bmatrix} h'(x) & 0 \\ g'(x) & \mathrm{Id} \end{bmatrix}$$

has full (row) rank if and only if $h'(x)$ does. Additionally, existence of a $d = (d_x, d_s) \in \mathbb{R}^{n+n_{\text{ineq}}}$ such that

$$\begin{aligned} \widetilde{g}_i'(x,s)d = -d_{s,i} \quad & < 0,\ i \in \widetilde{\mathcal{A}(x,s)} \\ \widetilde{h}{}'(x)d = \begin{bmatrix} h'(x)d_x \\ g'(x)d_x + d_s \end{bmatrix} \quad & = 0 \end{aligned}$$

is equivalent to the existence of a $d_x$ such that

$$\begin{aligned} g_i'(x)d_x &< 0\ i \in \mathcal{A}(x) \\ h'(x)d_x &= 0, \end{aligned}$$

so MFCQ holds at $x$ for (5.1) if and only if it holds at $(x,s) = \Phi(x)$ for ($5.1_{\text{s}}$).

(c) For ACQ, we can find a fairly compact closed form of the tangent- and linearizing cones for the slacked problem. Starting with the tangent cone $\mathcal{T}_{(x,s)}(\widetilde{F})$, we know that by definition this contains all directions $(d_x, d_s) \in \mathbb{R}^{n+n_{\text{ineq}}}$ such that there exists a positive null sequence $t^{(k)}$ and a sequence $(x^{(k)}, s^{(k)}) \in \widetilde{F} = \Phi(F)$ such that

$$\frac{\Phi(x^{(k)}) - \Phi(x)}{t_k} = \frac{(x^{(k)}, s^{(k)}) - (x,s)}{t^{(k)}} \longrightarrow (d_x, d_s)$$

so the $d_x$ components are exactly the directions in $\mathcal{T}_F(x)$. For the $s$ and $s^{(k)}$ we know that $s = -g(x)$ and $s^{(k)} = -g(x^{(k)})$, so that, using the mean value theorem for the continuous differentiable $g$, we obtain that

$$d_s \xleftarrow{k\to\infty} \frac{s^{(k)} - s}{t^{(k)}} = -\frac{g(x^{(k)}) - g(x)}{t^{(k)}} = -\frac{g'(\xi^{(k)})(x^{(k)} - x)}{t^{(k)}} \xrightarrow{k\to\infty} -g'(x)d_x,$$

which shows that the $d_s$ components of $d \in \mathcal{T}_{(x,s)}(\widetilde{F})$ are exactly the corresponding $-g'(x)$-transformed tangent directions $d_x \in \mathcal{T}_F(x)$. Over all, we obtain that $\mathcal{T}_{(x,s)}(\widetilde{F}) = \Phi'(x)\mathcal{T}_F(x)$ for all $x$ in $F$.

For the linearizing cone we know that for $(x, s) = \Phi(x)$ (by definition)

$$
\begin{aligned}
\mathcal{T}^{\text{lin}}_{(x,s)}(\widetilde{F}) &:= \left\{ d = (d_x, d_s) \in \mathbb{R}^{n+n_{\text{ineq}}} \;\middle|\; \begin{aligned} &\widetilde{g}_i{}'(x)\, d \le 0 && \text{for all } i \in \widetilde{\mathcal{A}(x,s)} \\ &\widetilde{h}_j{}'(x)\, d = 0 && \text{for all } j = 1, \dots, n_{\text{eq}} + n_{\text{ineq}} \end{aligned} \right\} \\
&= \{(d_x, d_s) \in \mathbb{R}^{n+n_{\text{ineq}}} \mid d_{s,i} \ge 0,\ i \in \widetilde{\mathcal{A}(x,s)},\ g'(x)d_x + d_s = 0,\ h'(x)d_x = 0\} \\
&= \{(d_x, -g'(x)d_x) \mid \text{for } d \in \mathbb{R}^n \text{ with } g'(x)d_x \le 0,\ i \in \mathcal{A}(x),\ h'(x)d_x = 0\} \\
&= \Phi'(x)\mathcal{T}^{\text{lin}}_F(x).
\end{aligned}
$$

So for $(x, s) = \Phi(x)$ we have the transformations

$$
\begin{aligned}
\mathcal{T}_{(x,s)}(\widetilde{F}) &= \Phi'(x)\mathcal{T}_F(x) \\
\mathcal{T}^{\text{lin}}_{(x,s)}(\widetilde{F}) &= \Phi'(x)\mathcal{T}^{\text{lin}}_F(x).
\end{aligned}
\tag{0.1}
$$

This shows that if $\mathcal{T}_F(x) = \mathcal{T}^{\text{lin}}_F(x)$, i. e., ACQ is satisfied at $x$ for (5.1), then ACQ holds at $(x, s)$ for ($5.1_\text{s}$).

In fact, because

$$
\Phi'(x) = \begin{bmatrix} \text{Id} \\ -g'(x) \end{bmatrix},
$$

we know that $\Phi'(x)$ has the left inverse

$$
\Phi^{-L} = \begin{bmatrix} \text{Id} & 0 \end{bmatrix}
$$

(the projection/restriction to the first $n$ components) and therefore (0.1) implies

$$
\begin{aligned}
\Phi'(x)^{-L}\mathcal{T}_{(x,s)}(\widetilde{F}) &= \mathcal{T}_F(x) \\
\Phi'(x)^{-L}\mathcal{T}^{\text{lin}}_{(x,s)}(\widetilde{F}) &= \mathcal{T}^{\text{lin}}_F(x),
\end{aligned}
\tag{0.2}
$$

which shows the reverse implication, so if ACQ is satisfied at $(x, s)$ for ($5.1_\text{s}$), then ACQ is also satisfied at $x$ for (5.1).

(d) For GCQ, we can reuse the transformation property (0.1) that we derived for the ACQ investigation. For any $x$ with corresponding $(x, s) = \Phi(x)$ we have that

$$\begin{aligned}(\mathcal{T}_{(x,s)}(\widetilde{F}))^\circ &= (\Phi'(x)\mathcal{T}_F(x))^\circ \\ &= \{f = (f_x, f_s) \in \mathbb{R}^{n+n_{\text{ineq}}} \mid f^\top \Phi'(x) d \le 0 \text{ for all } d \in \mathcal{T}_F(x)\} \\ &= \{f = (f_x, f_s) \in \mathbb{R}^{n+n_{\text{ineq}}} \mid \left(\Phi'(x)^\top f\right)^\top d \le 0 \text{ for all } d \in \mathcal{T}_F(x)\} \\ &= \{f = (f_x, f_s) \in \mathbb{R}^{n+n_{\text{ineq}}} \mid \Phi'(x)^\top f \in \mathcal{T}_F(x)^\circ\}\end{aligned}$$

and analogously

$$(\mathcal{T}^{\text{lin}}_{(x,s)}(\widetilde{F}))^\circ = \{f = (f_x, f_s) \in \mathbb{R}^{n+n_{\text{ineq}}} \mid \Phi'(x)^\top f \in \mathcal{T}^{\text{lin}}_F(x)^\circ\}.$$

Therefore, when $\mathcal{T}_F(x)^\circ = \mathcal{T}^{\text{lin}}_F(x)^\circ$ (GCQ is satisfied at $x$ for (5.1)), then of course $\mathcal{T}_{(x,s)}(\widetilde{F})^\circ = \mathcal{T}^{\text{lin}}_{(x,s)}(\widetilde{F})^\circ$ (GCQ holds at $\Phi(x)$ for ($5.1_s$)).

The reverse implication I don't expect to hold but I have yet to construct an example showing that.

(8 Points)

**Homework Problem 12.2** (Generalized derivatives) 5 Points

$(i)$ Compute the Bouligand- and Clarke generalized derivatives for $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = |x|$ at every $x \in \mathbb{R}$.

$(ii)$ Show that if $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz continuous on some neighborhood of $x \in \mathbb{R}^n$, then the Bouligand generalized derivative $\partial_B f(x)$ and the Clarke generalized derivative $\partial f(x)$ are nonempty and compact. In addition, $\partial f(x)$ is convex.

**Solution.**

$(i)$ The absolute value function is $C^\infty$ everywhere except for at the origin with $f'(x) = \operatorname{sgn}(x)$, $x \neq 0$, which is continuous in $x$.

Accordingly, for all $x \neq 0$,

$$\partial_B f(x) = \partial f(x) = \{f'(x)\} = \{\operatorname{sgn}(x)\}.$$

Since $D_F = \mathbb{R} \setminus \{0\}$, $f'(D_F) = \{-1, 1\}$ so that $\partial_B f(0) \subseteq \{-1, 1\}$ and the sequences $x^{\pm(k)} := \pm\frac{1}{k}$, $k \in \mathbb{N}$ with $f'(x^{\pm(k)}) = \pm 1$ show that in fact $\partial_B f(0) = \{-1, 1\}$ so that $\partial f(0) = [-1, 1]$. (2 Points)

$(ii)$ If $f$ is Lipschitz continuous with modulous $L > 0$ on some neighborhood $U(x)$ of $x$, then it is differentiable almost everywhere in that neighborhood (Rademacher's theorem).

For each $y \in D_F \cap U(x)$, the derivative $f'(y)$ satisfies $\|f'(y)\| \leq L$ because

$$\|f'(y)\| := \sup_{d \neq 0} \frac{\|f'(y)d\|}{\|d\|} = \sup_{d \neq 0} \lim_{t \searrow 0} \frac{\|f(y + td) - f(y)\|}{t\|d\|} \leq \sup_{d \neq 0} \lim_{t \searrow 0} \frac{Lt\|d\|}{t\|d\|} = L.$$

Additionally, there exists a sequence $x^{(k)} \in D_F$ with $x^{(k)} \to x$. Since $f'(x^{(k)})$ is bounded, there exists a convergent subsequence. The limit point of this subsequence is in the Bouligand generalized derivative, so it is nonempty. Neither is Clarke's generalized derivative, which is a superset.

Additionally, because of the boundedness of the derivative $f'(\cdot)$ by $L$ on $D_F \cap U(x)$, we of course have that $\|M\| \leq L$ for all $M \in \partial f(x)$, so both generalized derivatives are bounded. To show compactness, we only need to additionally show closedness of the generalized derivatives.

For $\partial_B f(x)$, let $M^{(k)} \in \partial_B f(x)$ such that $M^{(k)} \to M$ with sequences $x^{(k,l)} \in D_F$ such that $x^{(k,l)} \xrightarrow{l \to \infty} x$ and $f'(x^{(k,l)}) \xrightarrow{l \to \infty} M^{(k)}$. Then set any index $l^0(k)$ such that

$$\begin{aligned} \|x^{(k,l)} - x\| &\leq \frac{1}{k} \\ \|f'(x^{(k,l)}) - M^{(k)}\| &\leq \frac{1}{k} \end{aligned}$$

for all $l \geq l^0(k)$. Then the diagonal sequence $x^{(k,l^0(k))}$ obviously still converges to $x$ and

$$\|f'(x^{(k,l^0(k))}) - M\| \leq \|f'(x^{(k,l^0(k))}) - M^{(k)}\| + \|M^{(k)} - M\| \leq \frac{1}{k} + \|M^{(k)} - M\| \xrightarrow{k \to \infty} 0$$

shows that its derivatives converge to $M$, so $\partial_B f(x)$ is always compact.

In $\mathbb{R}^n$, the convex hull of a compact set is still compact Rockafellar, 1970, Thm. 17.2, which shows compactness of $\partial f(x)$.

Convexity of $\partial f(x) := \operatorname{conv} \partial_B f(x)$ is clear from definition.

(3 Points)

**Homework Problem 12.3** (Semismooth NCP functions) 6 Points

Show that

$$\Phi_{\min}(a,b) := \min\{a,b\} \qquad \text{“min” function,} \tag{11.8a}$$

$$\Phi_{\text{FB}}(a,b) := \sqrt{a^2+b^2}-a-b \qquad \text{Fischer-Burmeister function (Fischer, 1992)} \tag{11.8b}$$

as functions from $\mathbb{R}^2 \to \mathbb{R}$

$(i)$ are NCP functions (Definition 11.4).

$(ii)$ are semismooth everywhere (Definition 11.7).

**Solution.**

$(i)$ For $\Phi_{\min}(a,b)$, this is an easy observation, as $\min(a,b) = 0$ if and only if $a$ or $b$ are $0$ and the other value is $\geq 0$, so the zero levelset of $\min(a,b)$ on $\mathbb{R}^2$ is exactly the solution set of the complementarity condition.

For $\Phi_{\text{FB}}(a,b) = \sqrt{a^2+b^2}-a-b$ we have that

$$\begin{aligned}
\Phi_{\text{FB}}(a,b) := \sqrt{a^2+b^2}-a-b = 0 &\Leftrightarrow \sqrt{a^2+b^2} = a+b \\
&\Leftrightarrow a^2+b^2 = (a+b)^2 = a^2+b^2+2ab \text{ and } a+b \geq 0 \\
&\Leftrightarrow ab = 0 \text{ and } a+b \geq 0 \\
&\Leftrightarrow ab = 0 \text{ and } a \geq 0, b \geq 0.
\end{aligned}$$

$(ii)$ Let's first prove that when $f$ continuously differentiable around $x$, then $f$ is semismooth in $x$. This also implies that $f \in C^1$ is semismooth everywhere as claimed in the lecture notes.

Given the assumptions above, $f$ is locally Lipschitz around $x$. Now let $d^{(k)} \to d$, $t^{(k)} \searrow 0$ and $M^{(k)} \in \partial f(x + t^{(k)} d^{(k)})$. Then $x + t^{(k)} d^{(k)}$ will be inside the local neighborhood of continuous differentiability around $x$ from some index $k_0$ on, so

$$M^{(k)} = f'(\underbrace{x + t^{(k)} d^{(k)}}_{\to x}) \to f'(x),$$

i. e., the limit of $M^{(k)}d^{(k)}$ is $f'(x)d$ and therefore exists, which is semismoothness by definition. Accordingly, for the remainder of the exercise, we only need to consider the points of nondifferentiability of the NCP functions (they are continuously differentiable in a neighborhood of any other point).

For $\Phi_{\min}(a,b)$, this is exactly the set where $a = b$. The Clarke generalized derivative is

$$\partial\Phi_{\min}(a,b) = \begin{cases} \{(0,1)\} & \text{for } (a,b) \in H^+, \\ \{(\alpha, 1-\alpha) \mid \alpha \in [0,1]\} & \text{for } (a,b) \in H, \\ \{(1,0)\} & \text{for } (a,b) \in H^- \end{cases}$$

for

$$\begin{aligned} H &:= \{(a,b) \in \mathbb{R}^2 \mid a = b\} \\ H^+ &:= \{(a,b) \in \mathbb{R}^2 \mid a > b\} \\ H^- &:= \{(a,b) \in \mathbb{R}^2 \mid a < b\}, \end{aligned}$$

(see Example 11.6 of the lecture notes). Now let $d^{(k)} \to d$, $t^{(k)} \searrow 0$ and $M^{(k)} \in \partial\Phi_{\min}(x+t^{(k)}d^{(k)})$. For $d \in H^+$ or $d \in H^-$, the generalized derivatives $M^{(k)}$ are either $(1,0)$ or $(0,1)$, respectively, from an index $k_0$ on, meaning that the limits obviously exist. The interesting case is therefore, when $d_1 = d_2$, i. e. $d \in H$. In this case, there is a sequence $\alpha^{(k)} \in [0,1]$ with

$$M^{(k)}d^{(k)} = (\alpha^{(k)}, 1-\alpha^{(k)}) \begin{pmatrix} d_1^{(k)} \\ d_2^{(k)} \end{pmatrix} = \alpha^{(k)}d_1^{(k)} + (1-\alpha^{(k)})d_2^{(k)} \to d_1 = d_2$$

i. e. the limit exists. **Note:** The case where $x + t^{(k)}d^{(k)} \in H$ is covered by this argument.

For $\Phi_{\text{FB}}(a,b)$, the nondifferentiability is at $x = (0,0)$, where we first need to compute the Bouligand generalized derivative. For $(a^{(k)}, b^{(k)}) \to 0$ in $D_F$, we have that

$$\Phi'_{FB}(a^{(k)}, b^{(k)}) = \frac{1}{\|(a^{(k)}, b^{(k)})^\intercal\|_2} \begin{pmatrix} a^{(k)} \\ b^{(k)} \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

so the Bouligand generalized derivative is the shifted sphere

$$\partial_B\Phi_{\text{FB}(0,0)} = \{x - (1,1)^\intercal \mid x \in \mathbb{R}^2,\ \|x\|_2 = 1\}$$

and the Clarke generalized derivative is a shifted, closed euclidean 2-Ball:

$$\partial\Phi_{\text{FB}(0,0)} = \operatorname{cl} B_1^{\text{Id}}((0,0)^\intercal) - \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Now let $d^{(k)} \to d$, $t^{(k)} \searrow 0$ and $M^{(k)} \in \partial f(t^{(k)} d^{(k)})$. If $d \neq 0$, then $d^{(k)} \neq 0$ for the tail of the series and therefore (due to continuous differentiability) we know that

$$M^{(k)} = \frac{1}{\|d^{(k)}\|_2} \begin{pmatrix} d_1^{(k)} \\ d_2^{(k)} \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} \to \frac{1}{\|d\|_2} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

For $d = 0$, boundedness of the generalized derivatives shows that $M^{(k)} d^{(k)} \to 0$.

(6 Points)

**Homework Problem 12.4** (Reduced reformulation of the semismooth Newton step) 3 Points

Show that the semismooth Newton step (in abbreviated notation), cf. Equation (11.15):

$$\begin{bmatrix} H & -\mathrm{Id} & B^\mathsf{T} \\ D_\mathcal{A} & D_\mathcal{I} & 0 \\ B & 0 & 0 \end{bmatrix} \begin{pmatrix} d \\ \mu \\ \lambda \end{pmatrix} = \begin{pmatrix} b \\ D_\mathcal{A} \ell \\ c \end{pmatrix}$$

can be transferred by using selection matrices

$$\begin{aligned} Z_\mathcal{A} &:= \text{rows of Id} \in \mathbb{R}^{n \times n} \text{ pertaining to active indices} \\ Z_\mathcal{I} &:= \text{rows of Id} \in \mathbb{R}^{n \times n} \text{ pertaining to inactive indices} \end{aligned}$$

and subvectors $d_\mathcal{A} = Z_\mathcal{A} d$, $d_\mathcal{I} = Z_\mathcal{I} d$, $\mu_\mathcal{A} = Z_\mathcal{A} \mu$, $\mu_\mathcal{I} = Z_\mathcal{I} \mu$ into the equivalent reduced problem, cf. Equation (11.16):

$$\begin{bmatrix} Z_\mathcal{I} H Z_\mathcal{I}^\mathsf{T} & Z_\mathcal{I} B^\mathsf{T} \\ B Z_\mathcal{I}^\mathsf{T} & 0 \end{bmatrix} \begin{pmatrix} d_\mathcal{I} \\ \lambda \end{pmatrix} = \begin{pmatrix} Z_\mathcal{I} (b - H D_\mathcal{A} \ell) \\ c - B D_\mathcal{A} \ell \end{pmatrix}.$$

**Solution.**

In the lecture we already saw that from the second block row in Equation (11.15) we have $d_\mathcal{A} = Z_\mathcal{A} \ell$, $\mu_\mathcal{I} = 0$. We will explicitly insert these values into the system of equations and therefore the second block row is always fulfilled.

The first block row in Equation (11.15) can be manipulated as follows:

$$\begin{aligned} & H d - \mu + B^\mathsf{T} \lambda = b \\ \Leftrightarrow \quad & H Z_\mathcal{I}^\mathsf{T} d_\mathcal{I} - Z_\mathcal{A}^\mathsf{T} \mu_\mathcal{A} + B^\mathsf{T} \lambda = b - H D_\mathcal{A} \ell \quad \text{since } d = Z_\mathcal{A}^\mathsf{T} d_\mathcal{A} + Z_\mathcal{I}^\mathsf{T} d_\mathcal{I} \text{ and } \mu_\mathcal{I} = 0. \end{aligned}$$

We split this block row into two components by multiplication from the left with $Z_{\mathcal{I}}$ and $Z_{\mathcal{A}}$, respectively. We first multiply by $Z_{\mathcal{I}}$ to obtain

$$\Rightarrow \quad Z_{\mathcal{I}} H Z_{\mathcal{I}}^{\mathsf{T}} d_{\mathcal{I}} + Z_{\mathcal{I}} B^{\mathsf{T}} \lambda = Z_{\mathcal{I}} (b - H D_{\mathcal{A}} \ell).$$

This is the first equation in Equation (11.16). Multiplication with $Z_{\mathcal{A}}$ instead yields

$$\Rightarrow \quad Z_{\mathcal{A}} H Z_{\mathcal{I}}^{\mathsf{T}} d_{\mathcal{I}} - \mu_{\mathcal{A}} + Z_{\mathcal{A}} B^{\mathsf{T}} \lambda = Z_{\mathcal{A}} (b - H D_{\mathcal{A}} \ell),$$

which we resolve for $\mu_{\mathcal{A}}$ to obtain

$$\mu_{\mathcal{A}} = Z_{\mathcal{A}} H Z_{\mathcal{I}}^{\mathsf{T}} d_{\mathcal{I}} + Z_{\mathcal{A}} B^{\mathsf{T}} \lambda - Z_{\mathcal{A}} (b - H D_{\mathcal{A}} \ell) = Z_{\mathcal{A}} (H d + B^{\mathsf{T}} \lambda - b)$$

as claimed. The third block row in Equation (11.15) becomes

$$\begin{aligned} & B d = c \\ \Leftrightarrow \quad & B Z_{\mathcal{I}}^{\mathsf{T}} d_{\mathcal{I}} = c - B D_{\mathcal{A}} \ell, \end{aligned}$$

which is the second equation in Equation (11.16).

(3 Points)

Please submit your solutions as a single pdf and an archive of programs via moodle.

## References

Fischer, A. (1992). "A special Newton-type optimization method". *Optimization. A Journal of Mathematical Programming and Operations Research* 24.3-4, pp. 269–284. DOI: 10.1080/02331939208843795.

Rockafellar, R. T. (1970). *Convex Analysis*. Vol. 28. Princeton Mathematical Series. Princeton, New Jersey: Princeton University Press. URL: https://www.jstor.org/stable/j.ctt14bs1ff.