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

Exercise 8 - Solution

Date issued: 3rd June 2024
Date due: 11th June 2024

Homework Problem 8.1 (Two-Loop Recursion for Inverse BFGS Update) 6 Points

Show that Algorithm 4.53 in fact computes the action of the inverse BFGS updated matrix $B^{(k)}_{\text{BFGS}}$.

Solution.

We reuse the notation $\rho^{(k)} := 1/(y^{(k)})^\intercal s^{(k)}$ and $V^{(k)} := \mathrm{Id} - \rho^{(k)} y^{(k)} (s^{(k)})^\intercal$ from the lecture notes and note that the inverse BFGS update can then be written as

$$B^{(k+1)}_{\text{BFGS}} = (V^{(k)})^\intercal B^{(k)}_{\text{BFGS}} V^{(k)} + \rho^{(k)} s^{(k)} (s^{(k)})^\intercal.$$

For a fixed $k \in \mathbb{N}$ and an input vector $r \in \mathbb{R}^n$, the two-loop recursion in Algorithm 4.53 computes the quantities $q^{(i)}$ by setting (in the following the loop index i facilitates the move from the counter i to $i-1$ or $i+1$, respectively):

$$\begin{aligned}
i &= k &&: \quad q^{(k)} &&:= r \\
i &= k-1 &&: \quad q^{(k-1)} &&:= V^{(k-1)} q^{(k)} \\
&\vdots \\
i & &&: \quad q^{(i)} &&:= V^{(i)} q^{(i+1)} \\
&\vdots \\
i &= 0 &&: \quad q^{(0)} &&:= V^{(0)} q^{(1)}
\end{aligned}$$

in the backward loop storing the $\alpha^{(i)}$ but always rewriting the $q^{(i)}$ to the storage of r. In the forward loop, the values $z^{(i)}$ are computed as

$$\begin{aligned}
i = 0: \quad & z^{(0)} && := B_{\text{BFGS}}^{(0)} q^{(0)} \\
i = 1: \quad & z^{(1)} && := z^{(0)} + (\alpha^{(0)} - \beta^{(0)}) s^{(0)} \\
& \vdots \\
i+1 \quad : \quad & z^{(i+1)} && := z^{(i)} + (\alpha^{(i)} - \beta^{(i)}) s^{(i)} \\
& \vdots \\
i = k: \quad & z^{(k)} && := z^{(k)} + (\alpha^{(k)} - \beta^{(k)}) s^{(k)}.
\end{aligned}$$

Where we can inductively show that $z^{(i)} = B_{\text{BFGS}}^{(i)} q^{(i)}$, because

$$\begin{aligned}
z^{(i+1)} &:= z^{(i)} + (\alpha^{(i)} - \beta^{(i)}) s^{(i)} \\
&= z^{(i)} - \underbrace{\beta^{(i)}}_{\rho^{(i)} (y^{(i)})^\intercal z^{(i)}} s^{(i)} + \alpha^{(i)} s^{(i)} \\
&= (V^{(i)})^\intercal z^{(i)} + \alpha^{(i)} s^{(i)} \\
&= (V^{(i)})^\intercal B_{\text{BFGS}}^{(i)} q^{(i)} + \alpha^{(i)} s^{(i)} \\
&= (V^{(i)})^\intercal B_{\text{BFGS}}^{(i)} V^{(i)} q^{(i+1)} + \rho^{(i)} (s^{(i)})^\intercal q^{(i+1)} s^{(i)} \\
&= B_{\text{BFGS}}^{(i+1)} q^{(i+1)}
\end{aligned}$$

so after k iterations, we end up with $B_{\text{BFGS}}^{(k)} q^{(k)} = B_{\text{BFGS}}^{(k)} r$ as expected. (6 Points)

Homework Problem 8.2 (Examples for Tangent-, Linearizing and Normal Cones) 5 Points

Consider the feasible set

$$F := \left\{ x \in \mathbb{R}^n \,\middle|\, g_i(x) \le 0 \text{ for all } i = 1, \ldots, n_{\text{ineq}},\ h_j(x) = 0 \text{ for all } j = 1, \ldots, n_{\text{eq}} \right\} \tag{5.2}$$

without any equality restrictions h and with the inequality constraints $g\colon \mathbb{R}^3 \to \mathbb{R}^4$ defined by

$$g(x) = \begin{pmatrix} (x_1 - 1)^2 + x_2^2 - 1 \\ (x_1 - 3)^2 + x_2^2 - 1 \\ x_3 + 1 \\ -x_3 - 2 \end{pmatrix} \qquad \text{at} \qquad x^* = (2,\ 0,\ -1)^\intercal \in F.$$

Find the set of active indices $\mathcal{A}(x^*)$, an explicit representation of F, the tangent cone $\mathcal{T}_F(x^*)$, the **normal cone** $\mathcal{T}_F(x^*)^\circ$ and the linearizing cone $\mathcal{T}_F^{\text{lin}}(x^*)$ and sketch F and the cones.

Solution.

We have that

$$g(x^*) = \begin{pmatrix} (2-1)^2 - 1 \\ (2-3)^2 - 1 \\ -1 + 1 \\ -(-1) - 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ -1 \end{pmatrix} \leq 0$$

and therefore that in fact $x \in F$ with the set of active inequality constraint indices $\mathcal{A}(x^*) = \{1, 2, 3\}$.

The feasible set is the intersection of two cylinders of cross-section radius 1 with $(1, 0, x_3)$- and $(3, 0, x_3)$-axes, respectively, and two x_3-halfspaces, i. e., the line segment

$$F = \{(2, 0, x_3) \in \mathbb{R}^3 \mid x_3 \in [-2, -1]\}.$$

This makes it easy to compute the tangent cone, because all sequences in F are on a line, so the tangent cone ends up being the ray

$$\begin{aligned} \mathcal{T}_F(x^*) &:= \left\{ d \in \mathbb{R}^n \,\middle|\, \text{there exist sequences } x^{(k)} \in F \text{ and } t^{(k)} \searrow 0 \text{ such that } d = \lim_{k\to\infty} \frac{x^{(k)} - x^*}{t^{(k)}} \right\} \\ &= \{d \in \mathbb{R}^3 \mid d_1 = d_2 = 0,\ d_3 \leq 0\}. \end{aligned}$$

Accordingly the normal cone

$$\begin{aligned} \mathcal{T}_F(x^*)^\circ &:= \{s \in \mathbb{R}^n \mid s^\top x \leq 0 \text{ for all } x \in \mathcal{T}_F(x^*)\} \\ &= \left\{ s \in \mathbb{R}^3 \,\middle|\, s^\top \begin{pmatrix} 0 \\ 0 \\ x_3 \end{pmatrix} \leq 0 \text{ for all } x_3 \leq 0 \right\} \\ &= \{s \in \mathbb{R}^3 \mid s_3 \geq 0\} \end{aligned}$$

is the closed halfspace at 0 defined by the normal vector to be any nonzero element from the tangent ray.

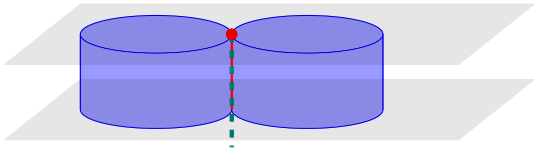

Figure 0.1: Constraining cylinders (blue), limiting hyperplanes of constraints on x_3 (gray), feasible set (red), point x^*, tangent cone (teal, dashed) shifted to x^* and limiting hyperplane of the half space normal cone (coincides with upper limiting hyperplane).

To find $\mathcal{T}_F^{\text{lin}}(x^*)$, we compute the derivatives of the active inequality constraints, which are

$$
\begin{aligned}
g_1'(x) &= \begin{pmatrix} 2(x_1-1) \\ 2x_2 \\ 0 \end{pmatrix} &&\Rightarrow & g_1'(x^*) &= \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} \\
g_2'(x) &= \begin{pmatrix} 2(x_1-3) \\ 2x_2 \\ 0 \end{pmatrix} &&\Rightarrow & g_2'(x^*) &= \begin{pmatrix} -2 \\ 0 \\ 0 \end{pmatrix} \\
g_3'(x) &= \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} &&\Rightarrow & g_3'(x^*) &= \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\mathcal{T}_F^{\text{lin}}(x^*) &:= \left\{ d \in \mathbb{R}^n \,\middle|\, \begin{matrix} g_i'(x^*)\, d \le 0 & \text{for all } i \in \mathcal{A}(x^*) \\ h_j'(x^*)\, d = 0 & \text{for all } j = 1, \ldots, n_{\text{eq}} \end{matrix} \right\} \\
&= \{ d \in \mathbb{R}^3 \mid d_1 = 0,\ d_3 \le 0 \} \\
&\supsetneq \mathcal{T}_F(x^*)
\end{aligned}
$$

(5 Points)

Homework Problem 8.3 (Linearizing Cone Depends on Description of Feasible Set) 2 Points

Consider the sets

$$
F^{(1)} := \left\{ x \in \mathbb{R}^2 \,\middle|\, \begin{pmatrix} -x_1 - 1 \\ x_1 - 1 \end{pmatrix} \le 0,\ x_2 = 0 \right\}, \quad \text{and} \quad F^{(2)} := \left\{ x \in \mathbb{R}^2 \,\middle|\, \begin{pmatrix} x_2 - (x_1+1)^3 \\ x_1 - 1 \end{pmatrix} \le 0,\ x_2 = 0 \right\}.
$$

Find an explicit description of the sets $F^{(1/2)}$ and compare the linearizing cones $\mathcal{T}_{F^{(1/2)}}^{\text{lin}}(x)$ at $x^* = (-1, 0)$.

Solution.

Due to the equality constraint, both sets in fact coincide with the line segment

$$F^{(1)} = F^{(2)} = \{x \in \mathbb{R}^2 \mid x_1 \in [-1, 1],\ x_2 = 0\}.$$

This automatically implies that their tangent- and normal cones (rays and closed halfspaces, respectively) coincide as well, as those are independent of the description of the sets. As long as the first inequality is inactive, the active inequalities conincide, so the linearizing cones of the two set descriptions coincide there as well.

At $x^* = (-1, 0)$ the active constraints differ. Due to $\mathcal{A}(x^*) = \{1\}$ in both cases, we compute

$$\begin{aligned}
g_1^{(1)\prime}(x) &= \begin{pmatrix}-1\\0\end{pmatrix} &&\Rightarrow\quad g_1^{(1)\prime}(x^*) = \begin{pmatrix}-1\\0\end{pmatrix}\\
g_1^{(2)\prime}(x) &= \begin{pmatrix}-3(x_1+1)^2\\1\end{pmatrix} &&\Rightarrow\quad g_1^{(2)\prime}(x^*) = \begin{pmatrix}0\\1\end{pmatrix}\\
h'(x) &= \begin{pmatrix}0\\1\end{pmatrix} &&\Rightarrow\quad h'(x^*) = \begin{pmatrix}0\\1\end{pmatrix}
\end{aligned}$$

and therefore obtain that

$$\mathcal{T}_{F^{(1)}}^{\text{lin}}(x^*) = \{d \in \mathbb{R}^2 \mid d_1 \geq 0, d_2 = 0\} \neq \{d \in \mathbb{R}^2 \mid d_2 = 0\} = \mathcal{T}_{F^{(2)}}^{\text{lin}}(x^*).$$

(2 Points)

Homework Problem 8.4 (Examples and Properties of Polar Cones) 4 Points

(i) Prove Lemma 5.9 of the lecture notes, i. e., for arbitrary sets $M, M_1, M_2 \subseteq \mathbb{R}^n$ the statements

(a) M° is a closed convex cone.

(b) $M_1 \subseteq M_2$ implies $M_2^\circ \subseteq M_1^\circ$.

(ii) Verify the claimed forms of the polar cones in Example 5.10, i. e., the following:

(a) Suppose that A is an affine subspace of $\mathbb{R}^n$ of the form $A = U + \{\bar{x}\}$. Then $A^\circ = \{\bar{x}\}^\circ \cap U^\perp$.

(b) In the absence of inequality constraints, the polar of the linearizing cone $\mathcal{T}_F^{\text{lin}}(x)$ for $x \in F$ has the representation

$$\begin{aligned}\mathcal{T}_F^{\text{lin}}(x)^\circ &= \{s \in \mathbb{R}^n \mid s \text{ is some linear combination of } h_j'(x)^\mathsf{T},\ j = 1, \dots, n_{\text{eq}}\} \\ &= \operatorname{range} h'(x)^\mathsf{T}.\end{aligned}$$

(c) Let $N := (\mathbb{R}_{\geq 0})^n$ denote the non-negative orthant in $\mathbb{R}^n$. Then $N^\circ = (\mathbb{R}_{\leq 0})^n$ is the non-positive orthant.

Solution.

(i) (a) Closedness of M° is simply due the fact that the definition involves a non strict inequality. Convexity is due to the linearity of the defining condition.

(b) This is due to the fact, that having to check a condition for all elements of a larger set is more restrictive.

(ii) (a) Let $A = U + \{\bar{x}\}$. Clearly, for all $s \in \{\bar{x}\}^\circ \cap U^\perp$, and all $\bar{x} + u \in \{\bar{x}\} + U$,

$$s^\mathsf{T}(\bar{x} + u) = s^\mathsf{T}\bar{x} \leq 0,$$

so $\{\bar{x}\}^\circ \cap U^\perp \subseteq (\{\bar{x}\} + U)^\circ$.

Conversely, if $s \in (\{\bar{x}\} + U)^\circ$, then $s^\mathsf{T}\bar{x} = s^\mathsf{T}(\bar{x} + \underbrace{0}_{\in U}) \leq 0$, so $s \in \{\bar{x}\}^\circ$. Also, assuming there were a $u \in U$ with $s^\mathsf{T}u \neq 0$, we know that λu is in the linear subspace U for all $\lambda \in \mathbb{R}$, and therefore

$$s^\mathsf{T}(\bar{x} + \lambda u) = s^\mathsf{T}\bar{x} + \lambda s^\mathsf{T}u \xrightarrow{\lambda \to \operatorname{sgn}(s^\mathsf{T}u)\infty} \infty$$

which is a contradiction to $s \in (\{\bar{x}\} + U^\circ)$, showing that in fact $\{x\}^\circ \cap U^\perp \supseteq (\{\bar{x}\} + U)^\circ$.

(b) This is of course a consequence of Lemma 5.13, which we did not know anything about at the time the remark was made. Luckily, in this special case, we immediately obtain that

$$\mathcal{T}_F^{\text{lin}}(x) = \{d \in \mathbb{R}^n \mid h_i'(x)d = 0,\ i = 1, \dots, n_{\text{eq}}\} = \{d \in \mathbb{R}^n \mid h'(x)d = 0\} = \ker h'(x),$$

which is a linear subspace, hence

$$\ker h'(x)^\circ = \ker h'(x)^\perp = \operatorname{range} h'(x)^\mathsf{T}.$$

(c) Let s be in the the non-positive orthant $(\mathbb{R}_{\leq 0})^n$ and d in the non-negative orthant $(\mathbb{R}_{\geq 0})^n$, then

$$s^\top d = \sum_{i=1}^{n} \underbrace{s_i}_{\leq 0} \underbrace{d_i}_{\geq 0} \leq 0$$

showing that $(\mathbb{R}_{\leq 0})^n \subseteq (\mathbb{R}_{\geq 0})^{n\circ}$.

Assuming there were an $s \in (\mathbb{R}_{\geq 0})^{n\circ}$ and an index i such that $s_i > 0$, then

$$s\top \underbrace{e_i}_{\in \mathbb{R}^n_{\geq 0}} = s_i > 0,$$

which finalizes the proof.

(4 Points)

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