Exercise 12 - Solution

Date issued: 3rd July 2023
Date due: 11th July 2023

Homework Problem 12.1 (Projected conjugate gradient method) 5 Points

Implement the projected M-preconditioned CG method Algorithm 13.2 and visualize the convergence behavior of the method

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{2} d^\intercal A\, d - b^\intercal d \\ \text{subject to} \quad & B\, d = c \end{aligned}$$

for pseudo-randomized problem data (symmetric $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$, $B \in \mathbb{R}^{m \times n}$ and $c \in \mathbb{R}^m$).

Solution.

The solution for this exercise will be delayed.[GM] (5 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^∞ 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

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

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 $\partial f(x) \coloneqq \operatorname{conv} \partial_B f(x)$ is clear from definition.

(3 Points)

Homework Problem 12.3 (Semismooth NCP functions) 6 Points

Show that

$$\Phi_{\min}(a,b) \coloneqq \min\{a,b\} \qquad \text{“min” function,} \tag{13.8a}$$

$$\Phi_{\text{FB}}(a,b) \coloneqq \sqrt{a^2+b^2} - a - b \qquad \text{Fischer-Burmeister function (Fischer, 1992)} \tag{13.8b}$$

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

(i) are NCP functions (Definition 13.4).

(ii) are semismooth everywhere (Definition 13.7).

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 ≥ 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

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

(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 13.6 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:

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

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 (Detecting convergence in primal-dual active set strategies) 6 Points

Consider the primal-dual active set strategy (semismooth Newton, Algorithm 13.10) for the lower bound constrained QP from the lecture notes with the iterates $(d^{(k)}, \mu^{(k)}, \lambda^{(k)})$, initialized with some $(d^{(0)}, \mu^{(0)}, \lambda^{(0)})$.

(i) Show that the residual $F(d^{(k)}, \mu^{(k)}, \lambda^{(k)})$ is nonzero only in its second component for $k \geq 1$.

(ii) Prove that when

$$\mathcal{A}(d^{(k)}, \mu^{(k)}) = \mathcal{A}(d^{(k+1)}, \mu^{(k+1)})$$

for some $k \in \mathbb{N}$ (the primal-dual active index sets coincide for two consecutive iterations) then $(d^{(k+1)}, \mu^{(k+1)}, \lambda^{(k+1)})$ is a solution of the constrained QP.

Solution.

(i) The first and last component of the nonlinear mapping

$$F(d, \mu, \lambda) \coloneqq \begin{pmatrix} A\, d - \mu + B^\intercal \lambda - b \\ \min\{\mu,\ d - \ell\} \\ B\, d - c \end{pmatrix}$$

are linear functions in the arguments, which are smooth. The semismooth Newton algorithm therefore solves these equation with the first iteration (its a linear system solve for a linear systems linearization, ie itself). This is true for each of the following iterations. Accordingly, the residual will only be nonzero in the second of the three components after the first iterateion.

(ii) The iteration giving the $k+1$ iterate is

$$\begin{pmatrix} A\,d^{(k)} - \mu^{(k)} + B^\mathsf{T}\lambda^{(k)} - b \\ \min\{\mu^{(k)},\ d^{(k)} - \ell\} \\ B\,d^{(k)} - c \end{pmatrix} + \begin{bmatrix} A & -\mathrm{Id} & B^\mathsf{T} \\ D_{\mathcal{A}(d^{(k)},\mu^{(k)})} & D_{\mathcal{I}(d^{(k)},\mu^{(k)})} & 0 \\ B & 0 & 0 \end{bmatrix} \begin{pmatrix} d^{(k+1)} - d^{(k)} \\ \mu^{(k+1)} - \mu^{(k)} \\ \lambda^{(k+1)} - \lambda^{(k)} \end{pmatrix} = 0.$$

As we have seen in the previous part of the exercise, $F(d^{(k+1)}, \mu^{(k+1)}, \lambda^{(k+1)}) \in 0 \times \mathbb{R}^n \times 0$, so all that is remaining to show is that $F_2(d^{(k+1)}, \mu^{(k+1)}, \lambda^{(k+1)}) = 0$, which is the complementarity problem of $\mu^{(k+1)}$ and $d^{(k+1)} - \ell$. Since the active and inactive sets for iteration $k, k+1$ coincide, we abbreviate $\mathcal{A} := \mathcal{A}(d^{(k+1)}, \mu^{(k+1)}) = \mathcal{A}(d^{(k)}, \mu^{(k)})$ and the same for the inactive set (where

$$\begin{aligned} \mathcal{A}(d,\mu) &:= \{i \in \{1,\ldots,n_{\text{ineq}}\} \mid \mu \ge d - \ell\} \qquad \textbf{primal-dual active indices} \text{ at } (d,\mu), \\ \mathcal{I}(d,\mu) &:= \{i \in \{1,\ldots,n_{\text{ineq}}\} \mid \mu < d - \ell\} \qquad \textbf{primal-dual inactive indices} \text{ at } (d,\mu) \end{aligned}$$

as in the script), so the second line of the system reduces to

$$\min\{\mu^{(k)},\ d^{(k)} - \ell\} + D_{\mathcal{A}}(d^{(k+1)} - d^{(k)}) + D_{\mathcal{I}}(\mu^{(k+1)} - \mu^{(k)}) = 0.$$

So for active indices $i \in \mathcal{A}$, we obtain that

$$d_i^{(k)} - \ell + d_i^{(k+1)} - d_i^{(k)} = 0$$

which implies that

$$d_i^{(k+1)} = \ell$$

and due to activity we have $\mu_i^{(k+1)} \ge 0$, so the complementarity problem encoded in the second component is satisfied, meaning $\min(\mu^{(k+1)}, d^{(k+1)} - \ell) = 0$.

For inactive indices we obtain

$$\mu_i^{(k)} + (\mu_i^{(k+1)} - \mu_i^{(k)}) = 0$$

which is $\mu_i^{(k+1)} = 0$ and due to inactivity $d^{(k)} > \ell$, so the complementarity problem encoded in the second component is satisfied, meaning $\min(\mu^{(k+1)}, d^{(k+1)} - \ell) = 0$.

(6 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.