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

Exercise 11 - Solution

Date issued: 24th June 2024
Date due: 2nd July 2024

Homework Problem 11.1 (QP Reformulation) 3 Points

Show that solving (9.7) and using the associated Lagrange multiplier (as described in (9.8)) leads to the same next iterate $(x^{(k+1)}, \lambda^{(k+1)})^\intercal$ as solving the "original" QP (9.5).

Solution.

For the "orignal" QP (9.5) we have the following way of deriving the next iterate: We solve

$$\begin{bmatrix} \mathcal{L}_{xx}(x^{(k)}, \lambda^{(k)}) & h'(x^{(k)})^\intercal \\ h'(x^{(k)}) & 0 \end{bmatrix} \begin{pmatrix} d \\ \lambda \end{pmatrix} = - \begin{pmatrix} \nabla_x \mathcal{L}(x^{(k)}, \lambda^{(k)}) \\ h(x^{(k)}) \end{pmatrix}$$

and then set

$$\begin{pmatrix} x^{(k+1)} \\ \lambda^{(k+1)} \end{pmatrix} \coloneqq \begin{pmatrix} x^{(k)} \\ \lambda^{(k)} \end{pmatrix} + \begin{pmatrix} d \\ \lambda \end{pmatrix}.$$

Now, for (9.7), we have associated KKT conditions

$$\begin{bmatrix} \mathcal{L}_{xx}(x^{(k)}, \lambda^{(k)}) & h'(x^{(k)})^\intercal \\ h'(x^{(k)}) & 0 \end{bmatrix} \begin{pmatrix} \tilde{d} \\ \tilde{\lambda}^{(k+1)} \end{pmatrix} = - \begin{pmatrix} \nabla_x f(x^{(k)}) \\ h(x^{(k)}) \end{pmatrix}$$

and then set

$$\begin{pmatrix} \tilde{x}^{(k+1)} \\ \tilde{\lambda}^{(k+1)} \end{pmatrix} \coloneqq \begin{pmatrix} x^{(k)} \\ \tilde{\lambda}^{(k+1)} \end{pmatrix} + \begin{pmatrix} \tilde{d} \\ 0 \end{pmatrix}.$$

(1 Point)

We have to show $\tilde{x}^{(k+1)} = x^{(k+1)}$ and $\tilde{\lambda}^{(k+1)} = \lambda^{(k+1)}$.

Immediately, we see that $d = \tilde{d}$ holds, from the respective second equations, and hence $\tilde{x}^{(k+1)} = x^{(k+1)}$. (1 Point)

Furthermore, we know that $\nabla_x \mathcal{L}(x^{(k)}, \lambda^{(k)}) = \nabla_x f(x^{(k)}) + h'(x^{(k)})^\intercal \lambda^{(k)}$ and thus easily see that

$$\mathcal{L}_{xx}(x^{(k)}, \lambda^{(k)})d + h'(x^{(k)})^\intercal \lambda = -\nabla_x \mathcal{L}(x^{(k)}, \lambda^{(k)}) = -\nabla_x f(x^{(k)}) - h'(x^{(k)})^\intercal \lambda^{(k)}$$

$$\Rightarrow \qquad \mathcal{L}_{xx}(x^{(k)}, \lambda^{(k)})d + h'(x^{(k)})^\intercal (\lambda^{(k)} + \lambda) = -\nabla_x f(x^{(k)}),$$

is equivalent to

$$\mathcal{L}_{xx}(x^{(k)}, \lambda^{(k)})d + h'(x^{(k)})^\intercal \tilde{\lambda}^{(k+1)} = -\nabla_x f(x^{(k)}),$$

which shows $\tilde{\lambda}^{(k+1)} = \lambda^{(k)} + \lambda = \lambda^{(k+1)}$. (1 Point)

Homework Problem 11.2 (Complementarity is equivalent to variational inequality) 3 Points

Prove Lemma 9.4 of the lecture notes, i. e. the equivalence of the KKT complementarity condition

$$\mu \ge 0, \quad g(x) \le 0, \quad \mu^\intercal g(x) = 0 \tag{9.11b}$$

and the variational inequality

$$\mu \in K \quad \text{and} \quad g(x)^\intercal (\nu - \mu) \le 0 \quad \text{for all } \nu \in K \tag{9.12}$$

with the closed convex cone $K := \mathbb{R}^{n_\text{ineq}}_{\ge 0}$ (the non-negative orthant).

Solution.

(3 Points) Suppose that the vectors μ and $g(x)$ satisfy the complementarity system (9.11b). Denote the active and inactive indices at x by $\mathcal{A}(x)$ and $\mathcal{I}(x)$, respectively. We obtain

$$g(x)^\intercal (\nu - \mu) = \sum_{i \in \mathcal{A}(x)} \underbrace{g_i(x)}_{=0} (\nu_i - \mu_i) + \sum_{i \in \mathcal{I}(x)} \underbrace{g_i(x)}_{<0} (\underbrace{\nu_i}_{\ge 0} - \underbrace{\mu_i}_{=0}) \le 0$$

for all $\nu \in K$, i. e., (9.12) holds.

Conversely, suppose that (9.12) is true. For an arbitrary index i, we may have three cases. In case $g_i(x) < 0$, we must have $\mu_i = 0$. Otherwise the choice $\nu = \mu \pm \varepsilon e_i \in K$ (with the i-the standard basis vector e_i) gives a contradiction. In case $g_i(x) = 0$ nothing is to be shown. The case $g_i(x) > 0$ yields a contradiction to (9.12) when we choose $\nu = \mu + e_i$, so it cannot occur. This shows (9.11b).

Homework Problem 11.3 (On the normal cone) 3 Points

Prove Lemma 9.6 of the lecture notes, i. e., the following statements for a set $M \subseteq \mathbb{R}^n$ and $x \in M$:

(i) The normal cone is a closed convex cone.

(ii) $\mathcal{N}_M(x) = (M - \{x\})^\circ$ holds.

Additionally, prove that

(iii) $\mathcal{N}_M(x) \subseteq \mathcal{T}_M(x)^\circ$ but generally $\mathcal{N}_M(x) \neq \mathcal{T}_M(x)^\circ$.

Solution.

(i) This is due to the non-strict defining inequality and the continuity (and linearity) of the scalar product.

(ii) This is a direct consequence of the definition of the polar cone and the Minkowski sum.

(iii) Let $s \in \mathcal{N}_M(x)$ and $d \in \mathcal{T}_M(x)$ with the corresponding sequences $t^{(k)} \searrow 0$ and $x^{(k)} \in M$ such that

$$d^{(k)} := \frac{x^{(k)} - x}{t^{(k)}} \longrightarrow d.$$

Then

$$s^\mathsf{T} d^{(k)} = \frac{\overbrace{s^\mathsf{T}(x^{(k)} - x)}^{\leq 0}}{\underbrace{t^{(k)}}_{\geq 0}} \leq 0$$

and therefore

$$s^\mathsf{T} d \leq 0$$

due to continuity.

The inverse inclusion generally does not hold because the tangent cone (and hence its polar) only considers *local* information, while the normal cone is built on information concerning the entire set.

Consider, e. g., the set

$$M := \{x \in \mathbb{R}^2 \mid x_2 \geq \sqrt{|x_1|}\}$$

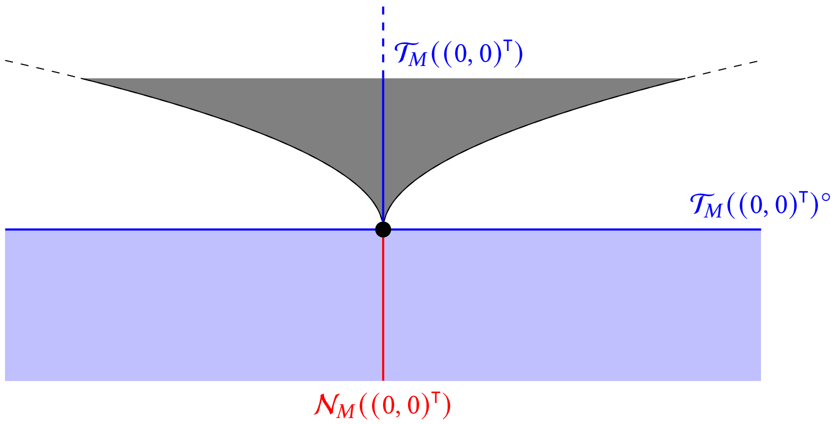

Figure 0.1: Set M (gray), tangent cone with its polar and the normal cone at the origin. The tangent cone and its polar only use local information while the normal cone requires information on the entire set.

at the origin, where

$$\mathcal{T}_M(x) = 0 \times \mathbb{R}_{\geq} \quad \text{and} \quad \mathcal{T}_M(x)^\circ = \mathbb{R} \times \mathbb{R}_{\leq}$$

but

$$\mathcal{N}_M(x) = 0 \times \mathbb{R}_{\leq} \subsetneq \mathcal{T}_M(x)^\circ.$$

(3 Points)

Homework Problem 11.4 (Examples for generalized Newton) 5 Points

For the nonlinear functions $F\colon \mathbb{R} \to \mathbb{R}$ and the set valued functions $N\colon \mathbb{R} \to \mathcal{P}(\mathbb{R})$ below, find all solutions z^* of the generalized equation

$$0 \in F(z) + N(z)$$

and determine, at which solutions the problem is strongly regular.

(i) $F(z) \coloneqq z^2 - 1$ and $N(z) \coloneqq \{0\}$

(ii) $F(z) \coloneqq z^2 - 1$ and $N(z) \coloneqq \mathbb{R}_{\geq}$

(iii) $F(z) \coloneqq (z-1)^2$ and $N(z) \coloneqq \mathcal{N}_{\mathbb{R}_{\geq}}(z)$

Solution.

(*i*) When $N(z) = \{0\}$ for all $z \in \mathbb{R}$, then the inclusion reduces to the standard root-finding problem

$$F(z) = 0$$

which, in the case of the shifted quadratic function has the obvious solution set $\{-1, 1\}$. Note that the linearized inclusion

$$0 \in F(z^{(k)}) + F'(z^{(k)})(z^{(k+1)} - z^{(k)}) + N(z^{(k+1)}) \tag{9.17}$$

that defines the iterates $z^{(k+1)}$ of the generalized Newton method also reduces to

$$0 = F(z^{(k)}) + F'(z^{(k)})(z^{(k+1)} - z^{(k)}) \tag{0.1}$$

and hence to the standard root finding version of Newton's method.

The perturbed system, that we need to analyze for strong regularity is the linear system

$$\Delta = F(z^*) + F'(z^*)(w - z^*)$$

which, due to regularity of $F'(z*) \in \{-2, 2\}$ for all solutions $z^* \in \{-1, 1\}$ has the unique solution

$$w(\Delta) = z^* + \frac{\Delta - F(z^*)}{F'(z^*)}$$

which depend on Δ Lipschitz continuously with modulus $\frac{1}{|F'(z^*)|} = \frac{1}{2}$, so both solutions are strongly regular.

(*ii*) When $N(z) = \mathbb{R}_{\geq}$, then the inclusion

$$0 \in F(z) + N(z)$$

is equivalent to the problem

$$F(z) = z^2 - 1 \leq 0,$$

i. e., finding all arguments that produce a nonpositive function value for the shifted quadratic function, which has the obvious solution set $[-1, 1]$. However, the linearized and perturbed system we want to analyze for strong regularity reads

$$F(z^*) - \Delta + F'(z^*)(w - z^*) = (z^*)^2 - 1 - \Delta + 2z^*(w - z^*) \leq 0,$$

i. e.

$$F(z^*) + F'(z^*)(w - z^*) = (z^*)^2 - 1 + 2z^*(w - z^*) \leq \Delta,$$

which is not uniquely solvable for any of the z^* in the solution set.

For $z^* = 0$, this is clear, because $F'(z^*) = 0$, so for $\Delta \geq F(z^*)$, the perturbed system is solved by and $w \in \mathbb{R}$, while it is not solvable at all for $\Delta < F(z^*)$. For $z^* \in (0, 1]$ we can take, e. g., $\Delta = 0$ and observe that

$$F(z^*) + F'(z^*)(w - z^*) = F(z^*) + 2z^*(w - z^*) \leq F(z^*) \leq 0$$

for any $w \leq z^*$. For $z^* \in [-1, 0)$ we can proceed analogously. Accordingly, we can show that moving to the correct side (depending on whether the linearization of F at z^* is monotonically increasing or decreasing) we decrease the functional value so the inclusion remains satisfied.

Note: The simple idea that standard Newton's method is equivalent to generalized Newton with the trivial cone valued map and finds zeros of functions so generalized Newton with a half space as the relevant cone should be able to find solution of sign problems is not necessarily correct. In this example, none of the solutions are isolated so strong regularity will not be satisfied. This is not an application that generalized Newton methods are "designed for".

(iii) Note that the standard quadratic function is no longer shifted down, but to the right with apex at $(1, 0)$. The inclusion problem

$$0 \in F(z) + N(z)$$

means that $z \in \mathbb{R}_{\geq}$ (because otherwise the normal cone is empty), that $F(z) \geq 0$ if $z = 0$ and $F(z) = 0$ if $z > 0$, so the inclusion is equivalent to the nonlinear complementarity problem

$$F(z) \geq 0 \perp z \geq 0.$$

The single zero of $F(z)$ is at $z = 1 \geq 0$ while the single zero of the identity is at $z = 0$ with $F(0) = 1$ so the solution set is $\{0, 1\}$.

The point $z^* = 1$ is not strongly regular, because $F'(1) = 0$, so the perturbed, linear inclusion

$$\Delta \in F(z^*) + F'(z^*)\,(w - z^*) + N(w)$$

reduces to

$$\Delta \in \mathcal{N}_{\mathbb{R}_{\geq}}(w)$$

which, for $\Delta = 0$, is solved by any $w \in \mathbb{R}_{\geq}$.

At the point $z^* = 0$, we have $F(z^*) = 1$ and $F'(z^*) = -2$, so the linear perturbed system becomes

$$\Delta \in 1 - 2w + \mathcal{N}_{\mathbb{R}_\geq}(w)$$

For $\Delta \neq 1$, this is uniquely solvable by $w(\Delta) \coloneqq \frac{1-\Delta}{2} \neq 0$ (the cone collapses to the trivial $\{0\}$ in this case), which Lipschitz continuously depends on Δ with modulous $\frac{1}{2}$. I. e., we can take $\delta < 1$ to show strong regularity.

Note: The two possible solutions are isolated so strong regularity is possible. Only where the tangent degenerates, we loose information, similarly to standard Newton.

(5 Points)

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