R. Herzog, G. Müller, E. Herberg, M. Hashemi
Heidelberg University



# Exercise 13 - Solution

Date issued: 10th July 2023
Date due: 18th July 2023

**Homework Problem 13.1** (Differentiability of the $\ell_1$-merit function) 2 Points

Verify that the directional derivative

$$\pi_1'(x; d) \coloneqq \lim_{t \searrow 0} \frac{\pi_1(x + t\, d) - \pi_1(x)}{t}$$

of the $\ell_1$-penalty part

$$\pi_1(x) \colon \mathbb{R}^n \to \mathbb{R}, \quad \pi_1(x) \coloneqq \sum_{i=1}^{n_{\text{ineq}}} \max\{0, g_i(x)\} + \sum_{j=1}^{n_{\text{eq}}} |h_i(x)|$$

of the $\ell_1$-merit function exists everywhere and is given by

$$\begin{aligned} \pi_1'(x; d) = \sum_{\substack{i=1 \\ g_i(x)<0}}^{n_{\text{ineq}}} 0 \quad &+ \sum_{\substack{i=1 \\ g_i(x)=0}}^{n_{\text{ineq}}} \max\{0, g_i'(x)\, d\} \quad + \sum_{\substack{i=1 \\ g_i(x)>0}}^{n_{\text{ineq}}} g_i'(x)\, d \\ &+ \sum_{\substack{j=1 \\ h_j(x)<0}}^{n_{\text{eq}}} -h_j'(x)\, d \quad + \sum_{\substack{j=1 \\ h_j(x)=0}}^{n_{\text{eq}}} |h_j'(x)\, d| \quad + \sum_{\substack{j=1 \\ h_j(x)>0}}^{n_{\text{eq}}} h_j'(x)\, d \end{aligned} \tag{14.2}$$

for $d \in \mathbb{R}^n$.

**Solution.**

We can rewrite $|\cdot| = \max(0, \cdot) + \max(0, -\cdot)$ and only need to show that for differentiable $f \colon \mathbb{R} \to \mathbb{R}$, we have that

$$(\max(0,\cdot)\circ f)'(x;d) = \begin{cases} 0 & f(x) < 0 \\ \max(0, f'(x)d) & f(x) = 0 \\ f'(x)d & f(x) > 0 \end{cases}.$$

The first and third case are obvious as the function is either locally equal to 0 or to $f$ (due to continutiy of $f$).

For the second case, note that max is directionally differentiable and Lipschitz continuous, which implies that a chain rule for directional derivatives holds. For any $\tilde{d} \in \mathbb{R}$, we have that

$$\max{}'(0,0;\tilde{d}) = \lim_{t\searrow 0} \frac{\max(0, t\tilde{d})}{t} = \lim_{t\searrow 0} t\frac{\max(0,\tilde{d})}{t} = \max(0,\tilde{d}),$$

so the chain rule for the directional derivative for $x$ with $f(x) = 0$ yields

$$(\max(0,\cdot)\circ f)'(x;d) = \max{}'(0, f(x); f'(x;d)) = \max{}'(0,0; \underbrace{f'(x;d)}_{\tilde{d}}) = \max(0; f'(x;d)).$$

(2 Points)

**Homework Problem 13.2** (Penalty reformulation of infeasible SQP-subproblems) 1 Points

Show that the penalty reformulation

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{2} d^\intercal A\, d - b^\intercal d + \gamma\left[\mathbf{1}^\intercal v + \mathbf{1}^\intercal w + \mathbf{1}^\intercal t\right], \quad \text{where } (d, v, w, t) \in \mathbb{R}^n \times \mathbb{R}^{n_{\text{eq}}} \times \mathbb{R}^{n_{\text{eq}}} \times \mathbb{R}^{n_{\text{ineq}}} \\ \text{subject to} \quad & B_{\text{eq}}\, d - c_{\text{eq}} = v - w \\ \text{and} \quad & B_{\text{ineq}}\, d - c_{\text{ineq}} \le t \\ \text{as well as} \quad & v \ge 0,\ w \ge 0,\ t \ge 0 \end{aligned} \tag{14.10}$$

of the SQP-subproblem-type problem

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{2} d^\intercal A\, d - b^\intercal d, \quad \text{where } d \in \mathbb{R}^n \\ \text{subject to} \quad & B_{\text{eq}}\, d - c_{\text{eq}} = 0 \\ \text{and} \quad & B_{\text{ineq}}\, d - c_{\text{ineq}} \le 0 \end{aligned} \tag{14.9}$$

is always feasible.

**Solution.**

This is straightforward to see, as choosing any $d \in \mathbb{R}^n$, there is an infinite number of choices for $v, w, t$ such that the constraints are satisfied. One canonical one would be (interpreting the max componentwise)

$$\begin{aligned} v(d) &:= \max(B_{\text{eq}} d - c_{\text{eq}}, 0), \\ w(d) &:= \max(-(B_{\text{eq}} d - c_{\text{eq}}), 0), \\ t &:= \max(B_{\text{ineq}}\, d - c_{\text{ineq}}, 0). \end{aligned}$$

Note that these relations necessarily have to hold at any minimizer of the penalized problem (14.10), as we can see because otherwise reducing the objective value is obviously possible with the choice above. (1 Point)

**Homework Problem 13.3** (Smoothness properties of exact penalty functions) 6 Points

Consider the constrained optimization problem

$$\text{minimize} \quad f(x) \quad \text{where} \quad x \in \mathcal{F} \tag{P}$$

for a functional $f\colon \mathbb{R}^n \to \mathbb{R}$ and a nonempty feasible set $\mathcal{F} \subseteq \mathbb{R}^n$. Further, define the penalized (unconstrained) problems

$$\text{minimize} \quad f(x) + \gamma\pi(x) \quad \text{where} \quad x \in \mathbb{R}^n \tag{$\text{P}_\gamma$}$$

for a penalty function $\pi\colon \mathbb{R}^n \to \mathbb{R}$ and a (penalty) parameter $\gamma > 0$.

**Note:** A penalty function is defined as satisfying $\pi(x) = 0$ for $x \in \mathcal{F}$ and $\pi(x) > 0$ for $x \in \mathbb{R}^n \setminus \mathcal{F}$.

Show the following:

(*i*) If $x^* \in \mathcal{F}$ is a local/global solution for ($\text{P}_\gamma$) for a $\gamma^* > 0$, then it is a local/global solution for (P) and for ($\text{P}_\gamma$) for any $\gamma \geq \gamma^*$.

(*ii*) If there exist a $\gamma^* > 0$ and an $x^* \in \mathbb{R}^n$, such that $x^*$ is a global solution of ($\text{P}_\gamma$) for all $\gamma \geq \gamma^*$, then $x^*$ is a global solution to (P).

(*iii*) Let $f$ be differentiable. If $x^* \in \mathcal{F}$ is a local solution to (P) and to ($\text{P}_\gamma$) for a $\gamma^* > 0$, then $\pi$ is not differentiable at $x^*$ or $f'(x^*) = 0$.

What does Statement (*iii*) mean for exact penalization methods in general?

**Solution.**

We define $\Phi(x, \gamma) \coloneqq f(x) + \gamma \pi(x)$.

($i$) Let $x^*$ be a local solution for ($\text{P}_\gamma$) for a $\gamma^* > 0$, i. e., there is $\varepsilon \in (0, \infty]$ such that

$$\Phi(x^*, \gamma^*) \le \Phi(x, \gamma^*) \quad \forall x \in B_\varepsilon(x^*).$$

By definition of a penalty function, we have that

$$\Phi(x, \gamma) = f(x) \quad \forall x \in \mathcal{F},\ \gamma > 0$$

and therefore

$$\Phi(x, \gamma) = \Phi(x, \gamma^*) \quad \forall x \in \mathcal{F},\ \gamma > 0.$$

Additionally, also by definition,

$$\Phi(x, \gamma^*) \le \Phi(x, \gamma) \quad \forall x \in \mathbb{R}^n,\ \gamma > \gamma^*.$$

This combines to

$$\Phi(x^*, \gamma) = \Phi(x^*, \gamma^*) \le \Phi(x, \gamma^*) \le \Phi(x, \gamma) \quad \forall x \in B_\varepsilon(x^*),\ \gamma > \gamma^*, \tag{0.1}$$

which shows the minimizer property for all penalized problems with sufficiently large penalty parameter.

For the original problem (P), simply note that (0.1) implies that

$$f(x^*) \le f(x) \quad \forall x \in B_\varepsilon(x^*) \cap \mathcal{F}.$$

The global case is covered by $\varepsilon = \infty$ above.

($ii$) Let $x^*$ and $\gamma^*$ be given as in the problem statement. We show that $x^*$ must be feasible, because Statement ($i$) then shows that $x^*$ is a global solution to the original problem (P).

Suppose $x^* \notin \mathcal{F}$, then, because $\mathcal{F}$ is nonempty, we can fix any other $x \in \mathcal{F}$, so by definition of the penalty function, we have that

$$\pi(x^*) > 0 \quad \text{and} \quad \pi(x) = 0$$

so that

$$\Phi(x, \gamma) = f(x) \le f(x^*) + \gamma \pi(x^*) = \Phi(x^*, \gamma) \quad \forall \gamma > \max\left(\gamma^*, \frac{f(x) - f(x^*)}{\pi(x^*)}\right),$$

contradicting the assumption.

$(iii)$ By Statement $(i)$, we know that $x^*$ is a minimizer of $\Phi(\cdot, \gamma)$ for all $\gamma > \gamma^*$.

Assume that $\pi$ is differentiable at $x^*$, then the first order optimality conditions for the unconstrained, penalized problem yield

$$0 = \frac{\partial}{\partial x}\Phi(x^*, \gamma) = f'(x^*) + \gamma\pi'(x^*) \quad \forall\gamma \geq \gamma^* > 0,$$

which can only be true for $f'(x^*) = \pi'(x^*) = 0$.

If $f'(x^*) = 0$, then $x^*$ is a stationary point of the unconstrained problem, which is somewhat of a degenerate case, where the constraints can be disregarded. This means that exact penalization can only be done using nonsmooth penalty functions - i. e., in practice, we either have to deal with nonsmoothness in penalty functions or with penalty parameters that need to be driven to infinity. **Note:** The augmented Lagrangian approach introduces an exact but modified penalization that guesses and updates correct multipliers along the iterations.

(6 Points)

**Homework Problem 13.4** (KKT conditions for convex multiobjective optimization) 6 Points

The filter-globalization strategy for the SQP method is based on ideas from multiobjective optimization, where the local/global minimizer concepts known from singleobjective optimization are replaced with so called Pareto-optimal points, i. e. points whose function value tuples are not dominated. This exercise gives a small insight into the field of multiobjective optimization.

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a $C^1$-function whose components are convex.

$(i)$ Show that the set $f(\mathbb{R}^n) + \mathbb{R}^m_{\geq 0}$ is convex.

$(ii)$ Use the proper separation theorem for convex sets (e. g. Herzog, 2022, Satz 15.30) to show that all Pareto optimal points of the multiobjective optimization problem are global minimizers of corresponding single-objective optimization problems.

**Note:** Replacing multiobjective problems by corresponding single objective problems is known as scalarization.

$(iii)$ Derive first order necessary Pareto-optimality conditions.

**Solution.**

$(i)$ Let $a,\ b \in f(\mathbb{R}^n) + \mathbb{R}^m_{\geq 0}$, i. e.

$$a = f(x) + c$$
$$b = f(y) + d$$

for $x,\ y \in \mathbb{R}^n$ and $c, d \in \mathbb{R}^m_{\geq 0}$, and let $\alpha \in [0, 1]$.

Since all component functions of $f$ are convex, we know that

$$f(\alpha x + (1-\alpha)y) \leq_{\mathbb{R}^m} \alpha f(x) + (1-\alpha)f(y)$$

where $\leq_{\mathbb{R}^m}$ means that all components of the vectors in $\mathbb{R}^m$ satisfy the inequality, i. e., an ordering with respect to the nonnegative cone.

Accordingly

$$\begin{aligned}\alpha\, a + (1-\alpha)\, b &= \alpha\,(f(x) + c) + (1-\alpha)\,(f(y) + d)\\ &= \alpha\, f(x) + (1-\alpha)\, f(y) + \alpha\, c + (1-\alpha)\, d\\ &= \underbrace{f(\alpha\, x + (1-\alpha)\, y)}_{\in f(\mathbb{R}^n)} + \underbrace{[\alpha\, f(x) + (1-\alpha)\, f(y) - f(\alpha\, x + (1-\alpha)\, y)] + \alpha\, c + (1-\alpha)\, d}_{\in \mathbb{R}^m_{\geq 0}},\end{aligned}$$

showing convexity of the set in question. (2 Points)

$(ii)$ Let $x \in \mathbb{R}^n$ be a Pareto optimal point, i. e., its function value tupel $f(x) \in \mathbb{R}^m$ is not being dominated by any other function value tuple. Equivalently,

$$\big(f(x) + \mathbb{R}^m_{\leq 0}\big) \cap f(\mathbb{R}^n) = \{f(x)\},$$

which shows that $f(x)$ is on the boundary of $f(\mathbb{R}^n) + \mathbb{R}^m_{\geq 0}$ and not in its relative interior (which coincides with its interior in this case). The proper separation theorem applied to the point $f(x)$ and the set $f(\mathbb{R}^n) + \mathbb{R}^m_{\geq 0}$ therefore yields a properly separating hyperplane with a normal vector $\alpha \in \mathbb{R}^m$, i. e., there is an $\alpha \in \mathbb{R}^m$ such that

$$\alpha^\mathsf{T} f(x) \leq \alpha^\mathsf{T}(f(y) + a) \quad \forall y \in \mathbb{R}^n, a \in \mathbb{R}^m_{\geq 0}$$

with $\alpha \neq 0$. This implies that $\alpha \geq_{\mathbb{R}^m} 0$ and that

$$\alpha^\mathsf{T} f(x) \leq \alpha f(y) \quad \forall y \in \mathbb{R}^n$$

meaning that $x$ is a global minimizer for the weighted-sum problem

$$\text{Minimize} \quad \sum_{i=1}^m \alpha_i(x) f_i(x) \quad \text{where } x \in \mathbb{R}^n \tag{0.2}$$

Note that we can of course simply multiply $\alpha$ by the inverse of $\sum_{i=1}^m \alpha_i$ to obtain $\alpha_i \geq 0$, $\sum_{i=1}^m \alpha_i = 1$ in (0.2). (2 Points)

(*iii*) Since any Pareto optimal point is the global minimizer of a weighted-sum problem of the type (0.2), a first order Pareto optimality condition is simply the first order optimality condition of the weighted sum problem, i. e., we obtain the result that if a point $x \in \mathbb{R}^n$ is Pareto optimal for $f \colon \mathbb{R}^n \to \mathbb{R}^m$, then there exists $\alpha \in \mathbb{R}^m$ such that

$$\begin{aligned} \alpha_i \geq 0, \sum_{i=1}^m \alpha_i = 1,\ i = 1, \ldots, m \\ \sum_{i=1}^m \alpha_i f_i'(x) = 0. \end{aligned}$$

(2 Points)

**Note:** Note that $\alpha$ is not necessarily unique, as you can easily see when considering a constant function $f \colon \mathbb{R}^n \to \mathbb{R}^m$.

**Note:** The derivations above can be easily extended to cover convex constraints as well to derive KKT conditions. Generally, Pareto optimality can be reformulated using a multi-input maximum value function and you obtain very similar KKT type conditions in the general case using subdifferential calculus.

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

## References

Herzog, R. (2022). *Grundlagen der Optimierung*. Lecture notes. URL: https://tinyurl.com/scoop-gdo.