Pure and Applied Functional Analysis
Volume 2, Number 2, 2017, 183–220

Yokohama Publishers
ISSN 2189-3764 ONLINE JOURNAL

EXACT PENALTY AND LAGRANGE DUALITY VIA THE DIRECTED SUBDIFFERENTIAL

WOLFGANG ACHTZIGER, ROBERT BAIER, ELZA FARKHI, AND VERA ROSHCHINA

Dedicated to the memory of Alexander Rubinov

Abstract. We present a detailed study of the optimality conditions for constrained nonsmooth optimization problems via the directed subdifferential in the finite-dimensional setting. Three standard approaches from the field of nonlinear programming are considered: the exact l_1-penalty approach, Lagrange duality, and saddle point optimality conditions. The results presented in the paper apply to a large class of problems in which both the objective function and the constraints are directed differentiable (a class that includes definable locally Lipschitz functions and quasidifferentiable functions). All three approaches are illustrated by examples for which the directed subdifferential can be constructed analytically. The visualization parts of the directed subdifferential give additional information on the nature of critical points.

1. Introduction

The goal of this work is to provide a treatment of optimality conditions in terms of the directed subdifferential, giving a clear geometric interpretation whenever possible, and showcasing the suitability of the directed subdifferential for the study of structure of optimization problems. The directed subdifferential was originally introduced in [5] for DC functions, following a suggestion of Alexander Rubinov, and later extended to a significantly more general setting [9] as a means to reconcile the constructive and set-valued approaches to subdifferentials which are not necessarily convex sets. The directed subdifferential incorporates the best of two worlds: it possesses a unique representation, exact calculus rules, and a visualization. All this provides more accurate information about the shape of the subdifferential and the function properties at a given point, compared to the data which can be derived from inexact calculus rules for various common subdifferentials (often being convex sets). In particular, the exact calculus rules allow us to consider the (nonsmooth) l_1-penalty function which enables the direct calculation of minimizers.

This work draws upon the fundamentals of the classic necessary and sufficient optimality conditions for nonsmooth optimization problems that have attracted mathematicians during the past decades. Underpinning the classical approaches [11, 50], the l_1-penalty approach, Lagrange duality (see for both topics e.g., [14, 52,

2010 *Mathematics Subject Classification.* 49J52, 90C26, 26B25.
Key words and phrases. Nonsmooth optimization, Rubinov subdifferential, optimality conditions, exact penalty, Lagrange duality.

23, 16, 46] and [22, Sec. 11] as well as [15, Chap. 7, 10 and 14], [43, Subsec. 1.6.3 and 2.6.3]) and saddle point theorems ([21, 34], [27, Chap. 11]) are well-known techniques used to obtain optimality results. We found overviews on duality and penalty approaches in the books [13, 27, 1] inspirational for our work.

Known results depend on the chosen subdifferential which means that, e.g., the Lagrangian multiplier rule for nonsmooth functions can be found for the Clarke subdifferential, the Dini subdifferential (in a fuzzy form), the proximal, the Michel-Penot, the limiting and the (Mordukhovich) basic subdifferential (see e.g., in [18, 29, 49, 32, 56, 33, 57] and [20, Problem 11.18], [41, Sec. 5.1]). Exact penalization with the l_1-approach often requires good calculus rules for the subdifferential. It is therefore not surprising that several works appeared in this research area [54, 44, 45, 55, 24, 38, 56, 37, 22, 25] which rely on the quasidifferential introduced by Demyanov and Rubinov (see e.g., [22, Sec. III.2, 1.]). Due to its focus on embedding the cone of convex compact subsets into the vector space of pairs of sets [42], the quasidifferential inherits excellent calculus rules from this corresponding vector space.

The *directed subdifferential* which is the focus of this work is based on an idea of embedding, i.e., this subdifferential lives in its own Banach space of directed sets (similarly to the quasidifferential that is effectively an element of Minkowski-Rådström-Hörmander space). A visualization of the directed subdifferential as a usually nonconvex subset of $\mathbb{R}^n$ is possible and has close links to other subdifferentials. The positive part of the visualized directed subdifferential of a DC function f equals to the *Dini subdifferential* $\partial_D f(x)$ in [12, 31], the negative part to the *Dini superdifferential* $\partial_D^{\geq} f(x)$, and the convex hull of its boundary coincides with the *Michel-Penot subdifferential* $\partial_{MP} f(x)$ in [39]. In [6] a possibility for a geometric calculation of the *(Mordukhovich) basic subdifferential* in [40, 41] is found for a difference of convex, positive homogeneous functions in $\mathbb{R}^2$. Parts of the visualization of the directed subdifferential coincides with the basic subdifferential, other parts with the basic superdifferential.

In this work we formulate optimality conditions based on the directed subdifferential for unconstrained nonsmooth optimization problems to constrained problems. Our paper is organised as follows: in Section 2 we briefly remind the definition of the directed subdifferentials for two function classes, as well as optimality conditions and calculus rules that we use in the sequel. The main results of the paper are presented in Sections 3–5, devoted to the penalty approach, Lagrange duality and saddle point optimality conditions, respectively. We subdivide these sections into "Theory" and "Examples" parts for reader's convenience. The last section contains conclusive remarks.

Throughout the paper we use the following notation. For a vector $\lambda = (\lambda_1, \dots, \lambda_n)^T \in \mathbb{R}^n$ we denote by $\lambda \geq 0$ the component-wise inequalities $\lambda_i \geq 0$ for $i = 1, \dots, n$. By $\|\lambda\|_1$ and $\|\lambda\|_2$ we denote the *sum norm* resp. the *Euclidean norm*, i.e., $\|\lambda\|_1 = \sum_{i=1}^n |\lambda_i|$ and $\|\lambda\|_2 = \sqrt{\sum_{i=1}^n \lambda_i^2}$. We also use the *standard scalar product* $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$ on $\mathbb{R}^n$. The *canonical unit vectors* in $\mathbb{R}^n$ are denoted by $e^1, \dots, e^n$. By $\mathbb{R}_+$ we denote the set of non-negative numbers $\{\alpha \in \mathbb{R} \mid \alpha \geq 0\}$, as usual. We denote by $\operatorname{co}(M)$ the *convex hull* of a set $M \subset \mathbb{R}^n$.

2. The directed subdifferential

The *directed subdifferential* defined below is an element of the Banach space $\mathcal{D}(\mathbb{R}^n)$ of *directed sets* which are studied in a series of papers [2, 3, 4]. This Banach space is constructed via the closure of the linear hull of the cone $\mathcal{C}(\mathbb{R}^n)$ of compact, convex, non-empty subsets of $\mathbb{R}^n$ with respect to the norm introduced in [3, Definition 4.1]. An *n-dimensional directed set* $\overrightarrow{A}$ consists of a pair formed by two mappings, an $(n-1)$-dimensional directed set $\overrightarrow{A_{n-1}}$ and a continuous scalar function a_n which are evaluated at a parameter, the direction $l \in S_{n-1}$, i.e.,

$$\overrightarrow{A} := \big(\overrightarrow{A_{n-1}(l)}, a_n(l)\big)_{l \in S_{n-1}} .$$

The *embedding* J_n for a convex set $C \in \mathcal{C}(\mathbb{R}^n)$ introduced in [3, Definitions 3.4 and 4.3] uses the values of the *support function* $\delta^*(l, C) = \max_{c \in C} \langle l, c \rangle$ evaluated in a direction $l \in S_{n-1}$ as second component of the directed set and the *projection* $\Pi_{n-1,l}$ of the *supporting face* $Y(l, C) = \{c \in C \,|\, \langle l, c \rangle = \delta^*(l, C)\}$ into $\mathbb{R}^{n-1}$ which is embedded into $\mathcal{D}(\mathbb{R}^{n-1})$, i.e.,

$$J_n(C) = \Big(J_{n-1}(\Pi_{n-1,l}(Y(l, C))), \delta^*(l, C)\Big)_{l \in S_{n-1}} . \tag{2.1}$$

For brevity we denote $\overrightarrow{v} := J_n(\{v\})$ for a vector $v \in \mathbb{R}^n$.
The projection of $Y(l, A)$ to $\mathbb{R}^{n-1}$ allows a recursive approach for directed sets. The first component of a two-dimensional directed set is a *directed interval* $\overrightarrow{[a, b]}$ introduced by Kaucher [35], where $a > b$ is allowed. This component allows the definition of a difference of directed sets which is inverse to the Minkowski sum and operates, like all other arithmetic operations in this space, separately on both components. Due to the recursive nature of the definitions, directed sets are described via n scalar functions

$$\begin{pmatrix} a_n(l^n) \\ a_{n-1}(l^{n-1}, l^n) \\ \vdots \\ a_2(l^2, \ldots, l^n) \\ a_1(l^1, l^2, \ldots, l^n) \end{pmatrix} \qquad (l^1 \in S_0, l^2 \in S_1, \ldots, l^n \in S_{n-1}) \tag{2.2}$$

evaluated at unit vectors of dimensions 1, 2, ..., n, and the operations are reduced to simple ones on n scalar functions. As an example we mention that the *partial order* $\overrightarrow{A} \le \overrightarrow{B}$ for two n-dimensional directed sets is equivalent to the lexicographic order of the corresponding n describing functions in (2.2) (see [3, Definitions 3.5 and 4.6]). Operations on directed sets extend known operations in $\mathcal{C}(\mathbb{R}^n)$. As an example we state that the *supremum* with respect to the partial order for two embedded sets $C_1, C_2 \in \mathcal{C}(\mathbb{R}^n)$ yields

$$\max\{J_n(C_1), J_n(C_2)\} = J_n(\operatorname{co}(C_1 \cup C_2)) \tag{2.3}$$

by [3, Lemma 3.11 and Proposition 4.20]. Here and below we use the notation $\max\{\overrightarrow{A}, \overrightarrow{B}\} := \sup\{\overrightarrow{A}, \overrightarrow{B}\}$ for the supremum of two two directed sets $\overrightarrow{A}, \overrightarrow{B} \in \mathcal{D}(\mathbb{R}^n)$ defined in [3].

In comparison to other approaches of (minimal) embeddings for $\mathcal{C}(\mathbb{R}^n)$ with equivalence classes of pairs of sets as in [48, 42] or support functions as in [30], directed sets offer a (generally) nonconvex *visualization*

$$V_n(\overrightarrow{A}) := P_n(\overrightarrow{A}) \cup N_n(\overrightarrow{A}) \cup M_n(\overrightarrow{A}) \subset \mathbb{R}^n$$

with a *positive*, a *negative* and a *mixed-type part* ($P_n(\overrightarrow{A})$, $N_n(\overrightarrow{A})$, $M_n(\overrightarrow{A})$) and arrows attached to boundary parts, see [4]. We note that

$$V_n(J_n(C)) = P_n(J_n(C)) = C\ , \quad V_n(\,-J_n(C)) = N_n(\,-J_n(C)) = \ominus C$$

for $C \in \mathcal{C}(\mathbb{R}^n)$, and $\ominus C$ denotes the *pointwise inverse* of a set, i.e., $(-1)\cdot C$.

Directed sets are applied in several fields, and especially the definition and visualization of their difference have many links to existing set differences (e.g., the *geometric difference* $C \overset{*}{-} D$ *of Hadwiger/Pontryagin* in [28, 47] and to the *Demyanov difference* $C \dot{-} D$ in [22, Sec. III.1], [53]). Especially, we have the inclusion chain

$$\big(C \overset{*}{-} D\big) \cup \big(\ominus (D \overset{*}{-} C)\big) \subset V_n(J_n(C) - J_n(D)) \subset C \dot{-} D\ . \tag{2.4}$$

The visualization of the difference of directed sets is always non-empty and neither too small nor too big, but usually nonconvex.

Recall that a function f is called DC (difference of convex functions) iff it can be represented in the form $f = g - h$ with two convex functions g, h.

Definition 2.1 (directed subdifferential)**.** Let $f = g - h$ be a DC function (with $g, h : \mathbb{R}^n \to \mathbb{R}$ convex), and let $x \in \mathbb{R}^n$ be a given point. Then the *directed subdifferential* of f at x is defined as

$$\overrightarrow{\partial} f(x) := J_n(\partial g(x)) - J_n(\partial h(x))\ ,$$

where here and below $\partial g(x)$ denotes the classical (Moreau-Rockafellar) *convex subdifferential* of the function g at x, and J_n is the *embedding* in (2.1).

Remark 2.2. The directed subdifferential is well-defined and does not depend on the chosen DC representation. Also the directed subdifferential of a convex function $g : \mathbb{R}^n \to \mathbb{R}$ coincides with the embedded convex subdifferential, i.e., $\overrightarrow{\partial} g(x) = J_n(\partial g(x))$.

The directed subdifferential can be extended to quasidifferentiable functions [7, 8] and to a more general class of functions defined below. For more details see [9, 10].

Definition 2.3 (directed subdifferentiable function)**.** A function $f : \mathbb{R}^n \to \mathbb{R}$ is called *directed subdifferentiable* at $x \in \mathbb{R}^n$ iff the directional derivative $f'(x;l)$ exists at x for all directions $l \in \mathbb{R}^n$, the mapping $f'(x;\cdot)$ is continuous on $\mathbb{R}^n$ and bounded on S_{n-1} by some constant M and, for $n \geq 2$, its *restriction* $f_l : \mathbb{R}^{n-1} \to \mathbb{R}$ defined for $y \in \mathbb{R}^{n-1}$ by

$$f_l(y) := f'(x; l + \Pi^T_{n-1,l}(y)), \quad \text{where } f'(x;l) := \lim_{h \downarrow 0} \frac{1}{h}\big(f(x+hl) - f(x)\big),$$

is also *directed subdifferentiable* at $y = 0$ with the same constant M for every $l \in S_{n-1}$. Here, $\Pi^T_{n-1,l}$ maps each point of $\mathbb{R}^{n-1}$ back to the hyperplane $\operatorname{span}\{l\}^\perp$

in $\mathbb{R}^n$.
The *directed subdifferential* for such a function is defined as

$$\overrightarrow{\partial} f(x) := \left(\overrightarrow{\partial} f_l(0), f'(l;x) \right)_{l \in S_{n-1}}.$$

Remark 2.4. We mention three special cases of directed subdifferentiable functions. First, a DC function is directed subdifferentiable, since it is a difference of two Lipschitz functions. Secondly, consider *quasidifferentiable functions* for which the directional derivative is a difference of two support functions of sets in $\mathcal{C}(\mathbb{R}^n)$ (see, e.g., [22, Chap. III, (2.2)]). Each quasidifferentiable function is directed subdifferentiable (see, e.g., [9, Remark 5.3]). Finally, locally Lipschitz functions definable on o-minimal structures, and in particular locally Lipschitz semialgebraic functions are directed subdifferentiable.

By the definition each directed subdifferentiable function is directionally differentiable with a directional derivative being continuous with respect to the direction. The following example studying the function from [22, Chap. I, Example 3.2] shows that a directed subdifferentiable or a quasidifferentiable function may not be Lipschitz continuous nor even continuous.

Example 2.5. Consider two functions $p_1, p_2 : \mathbb{R} \to \mathbb{R}$ with $p_1(x) := x^2$, $p_2(x) := -x^2$, and set $\Omega_1 := \operatorname{epi} p_1 \subset \mathbb{R}^2$ and $\Omega_2 := \operatorname{hyp} p_2 \subset \mathbb{R}^2$ as epigraph and hypograph of the parabolas respectively. Furthermore, let $\Omega_3 := \mathbb{R} \times \{0\}$ and set $\Omega := \Omega_1 \cup \Omega_2 \cup \Omega_3$. We set $f(x) := 1$ for $x \in \Omega$, and $f(x) := 0$ for $x \in \mathbb{R}^2 \setminus \Omega$. Fig. 1 shows a plot of the graph of f on the square rectangle $[-2, 2]^2$. As indicated in [22] the function is discontinuous and directionally differentiable at the origin with $f'(0; l) = 0$ for all $l \in \mathbb{R}^n$. Clearly, f is quasidifferentiable at the origin.

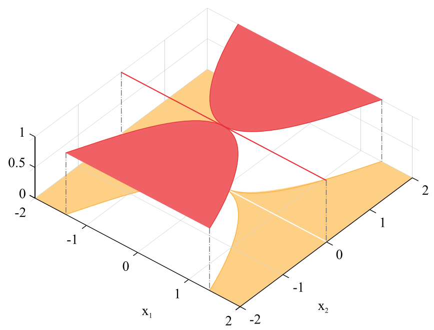


FIGURE 1. A directionally differentiable, discontinuous function (Example 2.5)

Taken into account the reprojection calculated in [7, Example 4.1], we have $\Pi_{1,l} = (l_2, -l_1)$. The restriction is given by

$$f_l(y) = f'(0; l + \Pi_{1,l}^T y) \quad \text{for } y \in \mathbb{R}.$$

Since the directional derivative at the origin is zero for all directions, we have $f_l(y) = 0$ for all $y \in \mathbb{R}$. This shows that f_l is directionally differentiable and fulfills all other requirements so that f is directed subdifferentiable.

The directed subdifferential can be regarded as a generalization of the gradient (see [10, Remark 3.1] for a proof).

Proposition 2.6 (differentiable function)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be (Fréchet-)differentiable at $x \in \mathbb{R}^n$. Then f is directed subdifferentiable, and its directed subdifferential at the point x is given by*

$$\overrightarrow{\partial} f(x) = \overrightarrow{\{\nabla f(x)\}}.$$

We restate the sum rule for the directed subdifferential, and we recall the calculation rule for the directed subdifferential of a maximum of directed subdifferentiable functions proved in [10].

Proposition 2.7 (calculus rules of the directed subdifferential)**.** *Let $f_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ be directed subdifferentiable functions, $i = 1, 2$, and let $\alpha, \beta \in \mathbb{R}$ be given constants. Then the following assertions hold.*

(a) *The function $f := \alpha f_1 + \beta f_2$ is directed subdifferentiable, and its directed subdifferential at arbitrary $x \in \mathbb{R}^n$ is given by*

$$\overrightarrow{\partial} f(x) = \alpha \overrightarrow{\partial} f_1(x) + \beta \overrightarrow{\partial} f_2(x) \, .$$

(b) *The pointwise maximum function $f := \max\{f_1, f_2\}$ with $f(x) := \max\{f_1(x), f_2(x)\}$ is directed subdifferentiable, and its directed subdifferential at arbitrary $x \in \mathbb{R}^n$ is given by*

$$\overrightarrow{\partial} f(x) = \max_{i \in I(x)} \overrightarrow{\partial} f_i(x) \, ,$$

where $I(x) := \{i \in \{1, 2\} \,|\, f_i(x) = f(x)\}$ is the set of active indices.

For proofs see [10, Proposition 2.2 and 2.4].

Theorem 2.8 (optimality conditions)**.** *Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be directed subdifferentiable.*

(a) *(Necessary optimality condition) Let $x^* \in \mathbb{R}^n$ be a local minimizer or a local maximizer, respectively, of f. Then*

$$0 \in P_n(\overrightarrow{\partial} f(x^*)) \quad \textit{resp.} \quad 0 \in N_n(\overrightarrow{\partial} f(x^*)) \, .$$

Moreover, with the development in [10, Proposition 3.1] *this inclusion is equivalent to the inequality*

$$\overrightarrow{0} \leq \overrightarrow{\partial} f(x^*) \quad \textit{resp.} \quad \overrightarrow{0} \leq -\overrightarrow{\partial} f(x^*) \, .$$

(b) *(Sufficient optimality condition) Let $x^* \in \mathbb{R}^n$ satisfy the condition*

$$0 \in \operatorname{int} P_n(\overrightarrow{\partial} f(x^*)) \, .$$

Then x^* is a local minimizer of f (similar for a local maximizer).

For proofs see, e.g., [10, Proposition 3.1] and [8, Proposition 4.4].

Remark 2.9. We introduce the *Rubinov subdifferential* as the visualization of the directed subdifferential [4, 5, 7],

$$\partial_R f(x) := V_n(\overrightarrow{\partial} f(x)) \, .$$

As a direct consequence of (2.4) we have the inclusion chain for subdifferentials,

$$\partial_D f(x) \cup \partial_D^{\geq} f(x) \subset \partial_R f(x) \subset \partial_{MP} f(x) \subset \partial_{Cl} f(x)\ , \tag{2.5}$$

where $\partial_{Cl} f(x)$ denotes the *Clarke generalized gradient*, i.e., *Clarke subdifferential* (see, e.g., [17, 19]).

There is a close connection of directed subdifferentiable functions to *Nesterov's lexicographically smooth functions* in [36] for locally Lipschitz continuous functions, although the lexicographic subdifferential does not coincide with the Rubinov subdifferential in general.

The paper discusses optimization problems of the following type.

$$(\mathbf{P}) \qquad \begin{array}{ll} \min\limits_{x\in\mathbb{R}^n} & f(x) \\ \text{s.t.} & g_i(x) \le 0 \quad (i=1,\dots,m), \\ & h_j(x) = 0 \quad (j=1,\dots,p). \end{array}$$

Let us add a few notations. The set of feasible points of $(\mathbf{P})$, i.e., the *feasible set of* $(\mathbf{P})$, is denoted by

$$\mathcal{F}(\mathbf{P}) := \{x \in \mathbb{R}^n \mid g_i(x) \le 0\ (i=1,\dots,m),\ h_j(x) = 0\ (j=1,\dots,p)\}.$$

Throughout the paper, we assume the following standard assumptions to hold:

A1. The feasible set of $(\mathbf{P})$ is not empty (for simplicity), i.e.,

$$\mathcal{F}(\mathbf{P}) \neq \varnothing. \tag{2.6}$$

A2. The given functions

$$f, g_1, \dots, g_m, h_1, \dots, h_p : \mathbb{R}^n \to \mathbb{R} \text{ are directed subdifferentiable and continuous.} \tag{2.7}$$

As a special case of A2, notice that each DC function is directed subdifferentiable (as outlined above) and continuous.

Furthermore, by f^* we denote

$$f^* := \inf\{\, f(x) \mid x \in \mathcal{F}(\mathbf{P})\,\} \in \mathbb{R} \cup \{-\infty\}$$

which is called the *optimal value of problem* $(\mathbf{P})$. Notice that the optimal value may not be attained or may be infinite, but $f^* < +\infty$ always holds due to assumption (2.6). Finally, when speaking about a "(global) solution x^* of problem $(\mathbf{P})$" we mean a (global) minimizer x^* of $(\mathbf{P})$, i.e., a point $x^* \in \mathcal{F}(\mathbf{P})$ with $f(x^*) = f^*$ $(> -\infty)$.

In the following three sections of the paper we consider three approaches which (first) convert problem $(\mathbf{P})$ into an unconstrained problem, and (secondly) consider optimality conditions based on the directed subdifferential. We assume that the background of some readers may be different from the mainstream of non-smooth optimization. Therefore, we briefly motivate the optimizer's point of view and cite the core references for each approach.

3. The penalty approach

In this section we consider the standard penalty approach to constrained nonlinear programming problems. To keep the exposition self-contained we recall some basics from the theory of penalty functions, and provide links to the literature.

3.1. **Theory.** Recall that a function $p : \mathbb{R}^n \longrightarrow \mathbb{R}$ is called a *penalty function for* $(\mathbf{P})$ iff p is continuous and satisfies the following conditions

$$(3.1) \qquad \begin{array}{rcll} p(x) & = & 0 & \text{for all } x \in \mathcal{F}(\mathbf{P}), \\ p(x) & > & 0 & \text{for all } x \notin \mathcal{F}(\mathbf{P}). \end{array}$$

One of the most useful penalty functions for $(\mathbf{P})$ is the l_1*-penalty function*,

$$(3.2) \qquad p_1 : \mathbb{R}^n \longrightarrow \mathbb{R}, \qquad p_1(x) := \sum_{i=1}^{m} \max\{0, g_i(x)\} + \sum_{j=1}^{p} |h_j(x)| \, .$$

Let p be a penalty function for $(\mathbf{P})$. For a fixed *penalty parameter*

$$\rho > 0$$

consider the following *auxiliary function of* $(\mathbf{P})$,

$$(3.3) \qquad f_\rho : \mathbb{R}^n \longrightarrow \mathbb{R}, \qquad f_\rho(x) := f(x) + \rho\, p(x).$$

If we choose the l_1-penalty function (i.e., "$p \equiv p_1$"), then we use the notation $f_{1,\rho} := f_\rho$, i.e., for all $x \in \mathbb{R}^n$

$$(3.4) \qquad f_{1,\rho}(x) := f(x) + \rho \Bigg(\sum_{i=1}^{m} \max\{0, g_i(x)\} + \sum_{j=1}^{p} |h_j(x)| \Bigg) \, .$$

Coming back to the general situation of arbitrary penalty function p, property (3.1) enforces that $f_\rho(x) = f(x)$ for all $x \in \mathcal{F}(\mathbf{P})$, and $f_\rho(x) > f(x)$ for all $x \notin \mathcal{F}(\mathbf{P})$. We mention that for $\rho \gg 0$ and $x \notin \mathcal{F}(\mathbf{P})$ we have $f_\rho(x) \gg f(x)$ ("penalty"). This leads to the general idea of penalty approaches, namely, replacing the constrained problem $(\mathbf{P})$ by the unconstrained(!) *penalty problem of* $(\mathbf{P})$ (or *auxiliary problem of* $(\mathbf{P})$),

$$(\mathbf{P}_\rho) \qquad \min_{x \in \mathbb{R}^n} f_\rho(x) \, .$$

If the penalty parameter has been chosen large, $\rho \gg 0$, then one hopes that a minimizer of the (unconstrained) penalty problem $(\mathbf{P}_\rho)$ is also a minimizer of the constrained original problem $(\mathbf{P})$ (see, e.g., [13, Chap. 9] or [1, Sec. 8.1] for background and proofs concerning penalty approaches).

For the formulation of a general classical convergence result we introduce the following notation. For fixed penalty parameter $\rho > 0$ denote the *optimal function value* of the penalty problem $(\mathbf{P}_\rho)$ by

$$f_\rho^* := \inf_{x \in \mathbb{R}^n} f_\rho(x) \quad \in \mathbb{R} \cup \{-\infty\} \, .$$

Moreover, when speaking about a "solution x^* of problem $(\mathbf{P}_\rho)$" we mean a global minimizer of the function f_ρ. Since such minimizer x^* in general depends on the parameter ρ, it is denoted by $x_\rho^* := x^*$. Notice, however, that x_ρ^* needs not be unique.

Recall that we assume the standard assumptions A1–A2 to hold throughout the paper.

Theorem 3.1. *For each $\rho > 0$ let the penalty problem $(\mathbf{P}_\rho)$ possess a solution x^*_ρ. Moreover, let the set $\{x^*_\rho \mid \rho > 0\} \subset \mathbb{R}^n$ be bounded. Then*

$$f^* \in \mathbb{R}, \qquad f^* = \lim_{\rho \to +\infty} f^*_\rho,$$

and the limit x^* of each convergent subsequence of x^*_ρ, $\rho \longrightarrow +\infty$, is a solution of the original problem $(\mathbf{P})$, i.e.,

$$x^* \in \mathcal{F}(\mathbf{P}) \qquad \text{and} \qquad f(x^*) = f^*.$$

Moreover, $\lim\limits_{\rho \longrightarrow +\infty} \rho\, p(x^*_\rho) = 0$.

Proof: Assumption A2 in (2.7) implies that all functions f, g_i and h_j are continuous. Moreover, $\mathcal{F}(\mathbf{P}) \neq \varnothing$ by assumption A1 (see (2.6)). These are the conditions such that a classical result on penalty methods can be applied (see, e.g., [13, Theorem 9.2.2]). □

Remark 3.2. Theorem 3.1 describes the situation for global minimizers of $(\mathbf{P})$ and $(\mathbf{P}_\rho)$. Notice, however, that in our (general) situation the function f_ρ may well be non-convex, and therefore the calculation of a global minimizer x^*_ρ of f_ρ may be difficult or, at least, strenuous.

We focus on another very useful detail (whose proof is straightforward due to property (3.1)).

Corollary 3.3. *Let $\overline{\rho} > 0$ be a penalty parameter, and let $x^*_{\overline{\rho}}$ be a solution of problem $(\mathbf{P}_{\overline{\rho}})$. Moreover, let $x^*_{\overline{\rho}}$ be feasible for the original problem $(\mathbf{P})$ (i.e., let $x^*_{\overline{\rho}} \in \mathcal{F}(\mathbf{P})$, or, equivalently, let $p(x^*_{\overline{\rho}}) = 0$). Then the following two assertions hold:*

(a) *The point $x^*_{\overline{\rho}}$ is a solution of $(\mathbf{P})$.*
(b) *For each $\rho \geq \overline{\rho}$ the point $x^*_{\overline{\rho}}$ is a solution of the penalty problem $(\mathbf{P}_\rho)$.*

In view of a practical algorithm, Theorem 3.1 indicates that problem $(\mathbf{P}_\rho)$ must be solved for larger and larger ρ. From a numerical view this may be a delicate matter. It is a pleasant situation if it is not necessary to let ρ tend to infinity. One possibility of such a situation is described above in Corollary 3.3. From a more general view, this property is related to the concept of a *exact penalty function for problem* $(\mathbf{P})$. In this paper we use the following common definition of exactness (see, e.g., [23] for a discussion on different notions of exactness).

Definition 3.4. Let x^* be a solution of problem $(\mathbf{P})$. A penalty function p of problem $(\mathbf{P})$ is called *exact (w.r.t. x^*)* if there exists a *threshold parameter* $\overline{\rho} > 0$ such that for each penalty parameter $\rho \geq \overline{\rho}$ the point x^* is a local minimizer of the auxiliary function f_ρ (i.e, x^* is a local minimizer of the penalty problem $(\mathbf{P}_\rho)$).

The practical significance of this definition lies in the fact that it may be sufficient to solve a single (unconstrained) penalty problem $(\mathbf{P}_\rho)$ to end up with the solution of the constrained problem $(\mathbf{P})$. Needless to say that Definition 3.4 becomes extremely powerful in practice if f_ρ is convex, and then x^* is a global minimizer of f_ρ.

Many investigations have been done on the exactness of penalty functions of type (3.1), see, e.g., [23] and the citations therein. As a general property we first note that p must be nonsmooth even if f is smooth (To see this, consider differentiable f, p, a solution x^* of (**P**), and the practical situation that $\nabla f(x^*) \neq 0$. Assuming that p is exact w.r.t. x^* it follows that $\nabla f_\rho(x^*) = 0$ for all $\rho \geq \overline{\rho}$, and thus $\nabla f(x^*) = 0$, a contradiction). Hence, if exactness plays a role, a nonsmooth penalty function should be considered. In our context we work with functions f, g_i, and h_j which are nonconvex and nonsmooth, anyway. Therefore, and since p_1 is simply structured, the l_1-penalty function is a top candidate for an exact penalty function in our framework.

Remark 3.5. The l_1-penalty function p_1 is well investigated, particularly in view of exactness, in the case when f, g_i, and h_j are locally Lipschitz-continuous. For illustration, we recall the well-known relation of the concept of *calmness* to the exactness of penalty functions (see, e.g., [16]) or to newer criteria based on the Mordukhovich calculus [26]. Calmness is a kind of Lipschitz-condition of the function value around a feasible point of (**P**) with respect to perturbations of the right-hand sides of the constraints of (**P**) (see, e.g., [19, Definition 6.4.1]. There exist also other definitions of calmness, see, e.g., [51]). In practice, calmness is a "mild" condition. Now, let all functions f, g_i, h_j be locally Lipschitz continuous, let x^* be a local minimizer of (**P**), and let (**P**) be calm at x^* in the sense of [19, Definition 6.4.1]. Then the l_1-penalty function p_1 of (**P**) is exact w.r.t. x^* (see [52, Proposition 1] for a proof; see also [19, Proposition 6.4.3] for an analogous result with a slightly different penalty function).

Summarizing, we arrive at the situation that it is of big practical interest to find local/global minimizers of the auxiliary function $f_{1,\rho}$ (see (3.4)). Our intention is to do this by the application of sufficient and necessary optimality conditions via directed subdifferentials (see Theorem 2.8 above). First we present formulas for the directed subdifferential of $f_{1,\rho}$ at an arbitrary point $x \in \mathbb{R}^n$. For this, we introduce the following notation of index sets.

Notation 3.6. Let $x \in \mathbb{R}^n$. We consider the index sets

$$\begin{aligned} I_g^0(x) &:= \{\, i \in \{1, \dots, m\} \mid g_i(x) = 0 \,\}, \\ I_g^+(x) &:= \{\, i \in \{1, \dots, m\} \mid g_i(x) > 0 \,\} \end{aligned}$$

related to the inequality constraints in (**P**), and

$$\begin{aligned} I_h^0(x) &:= \{\, j \in \{1, \dots, p\} \mid h_j(x) = 0 \,\}, \\ I_h^\pm(x) &:= \{\, j \in \{1, \dots, p\} \mid h_j(x) \neq 0 \,\} \end{aligned}$$

related to the equality constraints in (**P**).

These index sets will now be used for an easier expression of the directed subdifferential. The background lies in the application of standard calculus rules for the directed subdifferential. For example, in view of the function $|h_j|$ we make use of the representation $|h_j| = \max\{-h_j, +h_j\}$ in the second formula of the following proposition.

Proposition 3.7. *Let $\rho > 0$ and $x \in \mathbb{R}^n$ be arbitrary. Then the directed subdifferential of the auxiliary function $f_{1,\rho}$ at x is given by the formula*

$$\overrightarrow{\partial} f_{1,\rho}(x) = \overrightarrow{\partial} f(x) + \rho\Big(\sum_{i=1}^{m} \overrightarrow{\partial}(\max\{0, g_i\})(x) + \sum_{j=1}^{p} \overrightarrow{\partial} |h_j|(x) \Big)$$

$$= \overrightarrow{\partial} f(x) + \rho \Big(\sum_{i \in I_g^0(x)} \max \big\{ \overrightarrow{0}, \overrightarrow{\partial} g_i(x) \big\} + \sum_{i \in I_g^+(x)} \overrightarrow{\partial} g_i(x) + \sum_{j \in I_h^0(x)} \max \big\{ -\overrightarrow{\partial} h_j(x), \overrightarrow{\partial} h_j(x) \big\} + \sum_{j \in I_h^{\pm}(x)} \operatorname{sign}(h_j(x)) \overrightarrow{\partial} h_j(x) \Big) .$$

Proof. We apply the calculation rules for directed subdifferentials for the auxiliary function $f_{1,\rho}$ in (3.4) (see Proposition 2.7(a) and (b)). The case for $x \in \mathbb{R}^n$ with $g_i(x) < 0$ can be ignored, since $\overrightarrow{\partial}(\max\{0, g_i\})(x) = \overrightarrow{0}$. The third and fourth term appear by rewriting $|h_j|$ as $|h_j| = \max\{-h_j, h_j\}$. In both cases, if $h_j(x) = 0$ or $h_j(x) \neq 0$, the calculus rule for the maximum of two functions is applied. □

By these explicit formulas we are now able to apply the general optimality conditions from Theorem 2.8 above to our special unconstrained penalty problem $(\mathbf{P}_\rho)$ with $p \equiv p_1$.

Proposition 3.8 (necessary optimality conditions)**.** *Let $\rho > 0$ be a given penalty parameter, and consider the penalty problem $(\mathbf{P}_\rho)$ with the l_1-penalty function $p \equiv p_1$. Moreover, let $x_\rho^* \in \mathbb{R}^n$ be a local minimizer of the auxiliary function $f_{1,\rho}$. Then*

$$0 \in P_n(\overrightarrow{\partial} f_{1,\rho}(x_\rho^*)), \quad \textit{i.e.,} \quad \overrightarrow{0} \leq \overrightarrow{\partial} f_{1,\rho}(x_\rho^*) \,, \tag{3.5}$$

where $\overrightarrow{\partial} f_{1,\rho}(x_\rho^*)$ is given by Proposition 3.7 (for $x := x_\rho^*$).

Proposition 3.9 (sufficient optimality conditions)**.** *Let $\rho > 0$ be a given penalty parameter, and consider the penalty problem $(\mathbf{P}_\rho)$ with the l_1-penalty function $p \equiv p_1$. Moreover, let $\tilde{x} \in \mathbb{R}^n$ be a point satisfying the condition*

$$0 \in \operatorname{int} P_n(\overrightarrow{\partial} f_{1,\rho}(\tilde{x})). \tag{3.6}$$

Then $\tilde{x}$ is a local minimizer of the auxiliary function $f_{1,\rho}$.

3.2. **Examples.** Let us look at some academic examples.

Example 3.10. Consider the particular problem $(\mathbf{P})$ in one variable

$$\begin{aligned} \min_{x \in \mathbb{R}} \quad & x^2 - 8|x| + 7 \\ \text{s.t.} \quad & |x| \leq 1 \,. \end{aligned}$$

It is easy to see that the objective function $f(x) := x^2 - 8|x| + 7$ is strongly monotonically increasing on the interval $[-1, 0]$ and strongly monotonically decreasing on the interval $[0, +1]$. Therefore the local=global minimizers of $(\mathbf{P})$ are the two points $x^{*1} := -1$ and $x^{*2} := +1$ with optimal function value $f^* = f(x^{*\ell}) = 0$ for $\ell = 1, 2$.
Set $g_1(x) := |x| - 1$ for handling the inequality constraint in $(\mathbf{P})$. The auxiliary function $f_{1,\rho}$ is given by

$$f_{1,\rho}(x) = x^2 - 8|x| + 7 + \rho \max\{0, |x| - 1\}$$

$$= \begin{cases} x^2 - 8|x| + 7 & \text{if } |x| \le 1, \\ x^2 - (8-\rho)|x| + (7-\rho) & \text{otherwise.} \end{cases}$$

The standard assumptions A1–A2 are fulfilled, since $0 \in \mathcal{F}(\mathbf{P})$, g_1 is convex and $f(x) = (x^2+7) - 8|x|$ is DC.

(i) calculation of minimizers of the auxiliary function

Before we look at optimality conditions, we analyze $f_{1,\rho}$ by elementary means. We distinguish two cases:

Case 1: It is $0 < \rho < 6$. Then $\bar{x}_\rho := 4 - \frac{\rho}{2} > 1$. Consider $f_{1,\rho}(x)$ for $x > 1$. Then $f'_{1,\rho}(x) = 2x - (8-\rho)$, i.e., $f'_{1,\rho}(x) < 0$ for all $x \in (1, \bar{x}_\rho)$, and $f'_{1,\rho}(x) > 0$ for all $x > \bar{x}_\rho$. For $x \in [0,1]$ we have $f(x) = f_{1,\rho}(x)$, and therefore $f_{1,\rho}$ is strongly monotonically decreasing on $[0,1]$ (as seen above for function f). Summarizing, and using the facts that $f_{1,\rho}$ is continuos and that $f_{1,\rho}(x) = f_{1,\rho}(-x)$ for all $x \in \mathbb{R}$, we deduce that the points

$$x^{*1}_\rho := -\bar{x}_\rho = -(4 - \tfrac{\rho}{2}) \text{ and } x^{*2}_\rho := \bar{x}_\rho = 4 - \tfrac{\rho}{2} \quad \text{for all } \rho \in (0,6) \tag{3.7}$$

are the only local=global minimizers of $f_{1,\rho}$. Their optimal function values are

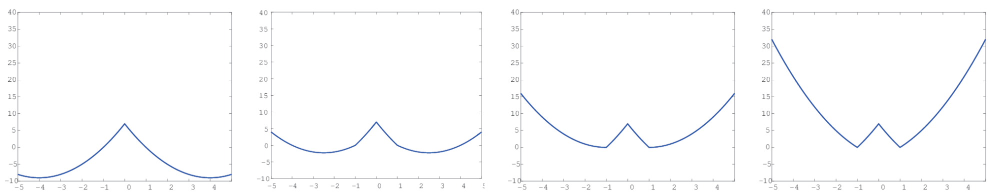
$$f^*_{1,\rho} = f_{1,\rho}(x^{*1}_\rho) = f_{1,\rho}(x^{*2}_\rho) = -\frac{\rho^2}{4} + 3\rho - 9\ .$$

Case 2: It is $\rho \ge 6$. For each $x > 1$ we deduce that $f'_{1,\rho}(x) = 2x - (8-\rho) > 2 - 8 + \rho = -6 + \rho \ge 0$, i.e., $f_{1,\rho}$ is strongly monotonically increasing on the interval $[1, +\infty[$. Analogously to the considerations in case 1 we see that the points

$$x^{*1}_\rho := x^{*1} = -1 \quad \text{and} \quad x^{*2}_\rho := x^{*2} = +1 \qquad \text{for all } \rho \in [6+\infty) \tag{3.8}$$

are the only local=global minimizers of $f_{1,\rho}$. Their optimal function values are $f^*_{1,\rho} = f_{1,\rho}(x^{*1}) = f_{1,\rho}(x^{*2}) = 1 - 8 + 7 = 0$. This concludes case 2.

Fig. 2 shows plots of the objective function f and of the auxiliary function $f_{1,\rho}$ for various values of ρ. Notice that $x^{*\ell}_\rho \notin \mathcal{F}(\mathbf{P})$, $\ell = 1,2$, for all $\rho \in]0,6[$, whereas



(a) obj. fct. f (b) $f_{1,\rho}$ for $\rho = 3$ (c) $f_{1,\rho}$ for $\rho = 6$ (d) $f_{1,\rho}$ for $\rho = 10$

FIGURE 2. The objective function and auxiliary functions for various values of ρ (Example 3.10)

$x^{*\ell}_\rho \in \mathcal{F}(\mathbf{P})$, $\ell = 1,2$, for all $\rho \ge \overline{\rho} := 6$. This also verifies the assertions of Corollary 3.3, and we observe the exactness of the l_1-penalty function p_1. Needless to say that $x^{*\ell}_\rho \longrightarrow x^{*\ell}$, $\ell = 1,2$, for $\rho \nearrow \overline{\rho} = 6$, as predicted by Theorem 3.1 (and, consequently, $f^*_{1,\rho} \longrightarrow f^*$ for $\rho \nearrow \overline{\rho}$).

(ii) calculation of the directed subdifferential of the auxiliary function

For any $\rho > 0$ we look at $f_{1,\rho}$ in its defining representation

$$f_{1,\rho}(x) = x^2 - 8|x| + 7 + \rho \max\{0, |x| - 1\} \qquad \text{for } x \in \mathbb{R}.$$

For any $x \in \mathbb{R}$ Propositions 2.6, 2.7, and 3.7 lead to the formulas

$$\overrightarrow{\partial} f(x) = J_1(\{2x\}) - \begin{cases} 8J_1(\{\text{sign}(x)\}) & \text{if } x \neq 0, \\ 8J_1([-1,1]) & \text{if } x = 0, \end{cases}$$

and

$$\overrightarrow{\partial} f_{1,\rho}(x) = \overrightarrow{\partial} f(x) + \begin{cases} \rho \cdot \overrightarrow{0} & \text{if } x \in (-1,1), \\ \rho \cdot \max\{J_1(\{0\}), J_1(\{\text{sign}(x)\})\} & \text{if } x \in \{-1,1\}, \\ \rho \cdot J_1(\{\text{sign}(x)\}) & \text{if } x \notin [-1,1]. \end{cases}$$

Furthermore, by combination of these two formulas and by using that

$$\max\{J_1(\{0\}), J_1(\{\text{sign}(x)\})\} = J_1(\text{co}\{0, \text{sign}(x)\}) = \begin{cases} [-1,0] & \text{if } x = -1 \\ [0,1] & \text{if } x = +1 \end{cases}$$

we see that

$$\begin{aligned} \overrightarrow{\partial} f_{1,\rho}(x) &= \begin{cases} J_1(\{2x+8-\rho\}) & \text{if } x \in (-\infty,-1), \\ J_1(\{2x+8\}) + \rho \cdot J_1([-1,0]) & \text{if } x = -1, \\ J_1(\{2x+8\}) & \text{if } x \in (-1,0), \\ J_1(\{0\}) - J_1([-8,8]) & \text{if } x = 0, \\ J_1(\{2x-8\}) & \text{if } x \in (0,1), \\ J_1(\{2x-8\}) + \rho \cdot J_1([0,1]) & \text{if } x = 1, \\ J_1(\{2x-8+\rho\}) & \text{if } x \in (1,\infty), \end{cases} \\ &= \begin{cases} J_1(\{2x+8-\rho\}) & \text{if } x \in (-\infty,-1), \\ J_1([6-\rho,6]) & \text{if } x = -1, \\ J_1(\{2x+8\}) & \text{if } x \in (-1,0), \\ -J_1([-8,8]) & \text{if } x = 0, \\ J_1(\{2x-8\}) & \text{if } x \in (0,1), \\ J_1([-6,-6+\rho]) & \text{if } x = 1, \\ J_1(\{2x-8+\rho\}) & \text{if } x \in (1,\infty). \end{cases} \end{aligned} \tag{3.9}$$

(iii) necessary optimality conditions by the directed subdifferential

Now we identify all points x satisfying the necessary optimality condition $0 \in P_1(\overrightarrow{\partial} f_{1,\rho}(x))$ for a local minimizer of $f_{1,\rho}$ (cf. Proposition 3.8). By the formulas in (3.9) we conclude that $0 \in P_1(\overrightarrow{\partial} f_{1,\rho}(x))$ holds if and only if

$$
\begin{array}{llll}
\big(& 0 = 2x + 8 - \rho = 2\big(x - \big(-4 + \frac{\rho}{2}\big)\big) & \text{and } x \in (-\infty, -1) & \big) \text{ or} \\
\big(& 6 - \rho \le 0 & \text{and } x = -1 & \big) \text{ or} \\
\big(& 0 = 2x + 8 & \text{and } x \in (-1, 0) & \big) \text{ or} \\
\big(& 0 = 2x - 8 & \text{and } x \in (0, 1) & \big) \text{ or} \\
\big(& -6 + \rho \ge 0 & \text{and } x = 1 & \big) \text{ or} \\
\big(& 0 = 2x - 8 + \rho = 2\big(x - \big(4 - \frac{\rho}{2}\big)\big) & \text{and } x \in (1, \infty) & \big).
\end{array}
\tag{3.10}
$$

In these considerations the point $\hat{x} := 0$ has been eliminated because its directed subdifferential $\overrightarrow{\partial} f_{1,\rho}(0) = -J_1([-8, 8])$ (see (3.9)) consists only of a negative visualization part, i.e., $P_1(\overrightarrow{\partial} f_{1,\rho}(0)) = \varnothing$ for all $\rho > 0$. Consequently, for any $\rho > 0$ the necessary optimality condition for a local minimizer is not satisfied at $\hat{x} = 0$. We mention, however, that $\hat{x} = 0$ satisfies the necessary optimality conditions for a local *max*imizer of $f_{1,\rho}$ for all $\rho > 0$, because $0 \in N_1(\overrightarrow{\partial} f_{1,\rho}(0)) = [-8, 8]$ for all $\rho > 0$.

Further straightforward elaborations on the conditions in (3.10) and sorting by the magnitude of ρ we arrive at the set X_ρ^{nec} collecting all points satisfying the necessary optimality conditions for local minimizers of $f_{1,\rho}$ given by Proposition 3.8,

$$
\begin{aligned}
X_\rho^{\text{nec}} &:= \{x \in \mathbb{R} \mid 0 \in P_1(\overrightarrow{\partial} f_{1,\rho}(x))\} \\
&= \begin{cases} \big\{ -\big(4 - \frac{\rho}{2}\big),\ +\big(4 - \frac{\rho}{2}\big)\big\} & \text{if } \rho < 6, \\ \{-1, +1\} & \text{if } \rho \ge 6. \end{cases}
\end{aligned}
\tag{3.11}
$$

Hence, for each fixed $\rho > 0$ the necessary optimality conditions identify exactly those points which are the true local=global minimizers of $f_{1,\rho}$ (compare to (3.7) and (3.8))!

(iv) sufficient optimality conditions by the directed subdifferential

For the application of sufficient conditions (cf. Proposition 3.9) we must consider situations where $\operatorname{int}(P_1(\overrightarrow{\partial} f_{1,\rho}(x))) \neq \varnothing$. The formulas in (3.9) show that this is the case if and only if $x \in \{-1, +1\}$ and $\rho > 6$. Moreover, we see that in these situations also the sufficient optimality condition is satisfied, because

$$
\left.
\begin{array}{lcl}
0 \in \operatorname{int}\big(P_1(\overrightarrow{\partial} f_{1,\rho}(-1))\big) &=& (6 - \rho, 6) \\
0 \in \operatorname{int}\big(P_1(\overrightarrow{\partial} f_{1,\rho}(+1))\big) &=& (-6, -6 + \rho)
\end{array}
\right\} \quad \text{for all } \rho > 6.
$$

Hence, for $\rho > 6$ the sufficient optimality conditions hold at the local=global minimizers $x^{*1} = -1$, $x^{*2} = +1$ of $f_{1,\rho}$ (compare to (3.8)).
We add that at the point $\hat{x} = 0$ the sufficient optimality condition for a local *max*imizer of $f_{1,\rho}$ is satisfied for all $\rho > 0$, because $0 \in (-8, +8) = \operatorname{int}(N_1(\overrightarrow{\partial} f_{1,\rho}(0)))$ for all $\rho > 0$.

(v) comparison to the Clarke subdifferential

A careful calculation of the Clarke subdifferential $\partial_{Cl} f_{1,\rho}(x)$ for any $\rho > 0$ and any x leads to formulas which are closely related to the expressions in (3.9). Since the calculus based on the Clarke subdifferential, however, is more unspecific in view of minimization and maximization, one obtains the set of critical points

$$\{x \in \mathbb{R} \mid 0 \in \partial_{Cl} f_{1,\rho}(x)\} = X_\rho^{\text{nec}} \cup \{0\} \qquad \text{for all } \rho > 0$$

with X_ρ^{nec} from (3.11) above. We see that by using the Clarke subdifferential calculus the local maximizer $\hat{x} = 0$ is not excluded from critical points.

Example 3.11. Consider the optimization problem of type (**P**) in two variables with one equality constraint

$$\begin{aligned} \min_{x \in \mathbb{R}^2} \quad & \big\{ \max\{x_1^4 + 2x_2, x_1^4 + x_1^2 + x_2\} - \max\{-x_2^4, -x_2^4 + x_1^2 + x_2\} \big\} \\ \text{s.t.} \quad & x_1 x_2 = 0 \, . \end{aligned}$$

For simplicity in notation we set $F_1, F_2, F_3 \colon \mathbb{R}^2 \longrightarrow \mathbb{R}$,

$$\begin{aligned} F_1(x_1, x_2) &:= \max\{2x_2, x_1^2 + x_2\}, \\ F_2(x_1, x_2) &:= \max\{0, x_1^2 + x_2\}, \\ F_3(x_1, x_2) &:= x_1^4 + x_2^4 \, . \end{aligned}$$

Then the objective function takes the form

$$f = F_1 - F_2 + F_3 \, .$$

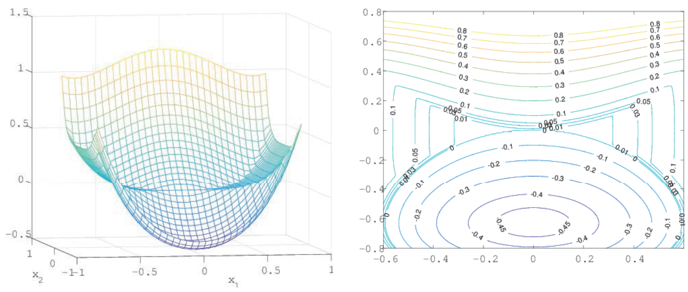
The standard assumptions A1–A2 are satisfied, since $F_1 - F_2$ is DC, F_3 and $h_1(x) = x_1 x_2$ are $\mathcal{C}^2$ and hence, DC. The function $F_1 - F_2$ appears in well-studied academic examples to illustrate different kinds of subdifferentials (see, e.g., [22, Sec. III.4, Ex. 4.2] and [5, Ex. 4.7]). From the latter reference we cite that

$$\overrightarrow{\partial} (F_1 - F_2)(0,0) = J_2(\{(0,1)^T\}) \tag{3.12}$$

while the Clarke subdifferential at the origin is given by

$$\partial_{Cl}(F_1 - F_2)(0,0) = \operatorname{co}(\{\tbinom{0}{0}, \tbinom{0}{1}\}) \, . \tag{3.13}$$

The plots in Fig. 3(a) and Fig. 3(b) show the graph and the level lines, respectively, of the objective function f on the rectangular domain $[-1,1]^2$ respectively



(a) objective function f (b) contour lines of obj. fct. f

FIGURE 3. Plot of the objective function and corresponding contour lines (Example 3.11)

$[-0.6, 0.6]^2$.

(i) calculation of minimizers of the original problem

For feasible points we obtain the objective function values

$$\begin{array}{rcll} f(x_1, 0) & = & x_1^4 & \text{for all } x_1 \in \mathbb{R}, \\ f(0, x_2) & = & x_2^4 + x_2 & \text{for all } x_2 \in \mathbb{R}. \end{array}$$

This shows that the unique global=local minimizer of $(\mathbf{P})$ is the feasible point

$$x^* := (0, -\tfrac{1}{\sqrt[3]{4}})^T = (0, -4^{-1/3})^T \approx (0, -0.6299605)^T$$

with optimal function value

$$f^* = f(x^*) = 4^{-4/3} - 4^{-1/3} \approx -0.4724704 < 0 \,.$$

(ii) calculation of the directed subdifferential of the auxiliary function

Now we have a look at the auxiliary function $f_{1,\rho}$ of $(\mathbf{P})$ for any given $\rho > 0$. With the above defined functions F_1, F_2, F_3, and with the function $F_4 \colon \mathbb{R}^2 \longrightarrow \mathbb{R}$, $F_4(x_1, x_2) := |x_1 x_2|$, we have

$$f_{1,\rho} = F_1 - F_2 + F_3 + \rho F_4.$$

The functions F_1 and F_2 (and F_3) are convex, and F_3 is continuous differentiable. Hence, Remark 2.2 and Propositions 2.6, 2.7 show that

$$\overrightarrow{\partial} F_1(x_1, x_2) = \begin{cases} J_2(\{(0,2)^T\}) & \text{if } x_2 > x_1^2, \\ J_2\left(\operatorname{co}\left(\{\binom{0}{2}, \binom{2x_1}{1}\}\right)\right) & \text{if } x_2 = x_1^2, \\ J_2(\{(2x_1, 1)^T\}) & \text{if } x_2 < x_1^2, \end{cases} \tag{3.14}$$

$$\overrightarrow{\partial} F_2(x_1, x_2) = \begin{cases} J_2(\{(0,0)^T\}) & \text{if } x_2 < -x_1^2, \\ J_2\left(\operatorname{co}\left(\{\binom{0}{0}, \binom{2x_1}{1}\}\right)\right) & \text{if } x_2 = -x_1^2, \\ J_2(\{(2x_1, 1)^T\}) & \text{if } x_2 > -x_1^2, \end{cases} \tag{3.15}$$

$$\overrightarrow{\partial} F_3(x_1, x_2) = J_2(\{4(x_1^3, x_2^3)^T\}) \tag{3.16}$$

for all $(x_1, x_2)^T \in \mathbb{R}^2$. Moreover, writing F_4 in the form $F_4(x_1, x_2) = \max\{-x_1 x_2, +x_1 x_2\}$ for all $x \in \mathbb{R}^2$ Propositions 2.7(b) and 2.6 show that

$$\overrightarrow{\partial} F_4(x_1, x_2) = \begin{cases} J_2(\{(-x_2, -x_1)^T\}) & \text{if } x_1 x_2 < 0 \\ J_2([-|x_2|, +|x_2|] \times \{0\}) & \text{if } x_1 = 0 \text{ and } x_2 \neq 0, \\ J_2(\{0\} \times [-|x_1|, +|x_1|]) & \text{if } x_1 \neq 0 \text{ and } x_2 = 0, \\ J_2(\{(0,0)^T\}) & \text{if } x_1 = x_2 = 0, \\ J_2(\{(x_2, x_1)^T\}) & \text{if } x_1 x_2 > 0 \end{cases} \tag{3.17}$$

for all $(x_1, x_2)^T \in \mathbb{R}^2$ (note that F_4 is Fréchet-differentiable at the origin which is easy to prove). From Proposition 2.7(a) we deduce that

$$\overrightarrow{\partial} f_{1,\rho}(x) = \overrightarrow{\partial} F_1(x) - \overrightarrow{\partial} F_2(x) + \overrightarrow{\partial} F_3(x) + \rho \, \overrightarrow{\partial} F_4(x) \qquad \text{for all } x \in \mathbb{R}^2.$$

By the help of the formulas in (3.14), (3.15), (3.16), and (3.17) above it is easy to calculate $\overrightarrow{\partial} f_{1,\rho}(x)$ at any point $x \in \mathbb{R}^2$.

For brevity, we calculate $\overrightarrow{\partial} f_{1,\rho}(x)$ just at two points, namely, at the optimal point $x^* = (0, -\frac{1}{\sqrt[3]{4}})^T$ from above, and at the origin $\hat{x} := (0,0)^T$.

For $x = x^*$ we obtain

$$\begin{aligned}\overrightarrow{\partial} f_{1,\rho}(x^*) &= J_2(\{(2x_1^*, 1)^T\}) - J_2(\{(0,0)^T\}) + J_2(\{(4(x_1^*)^3, 4(x_2^*)^3)^T\}) \\ &\quad + J_2([x_2^*, -x_2^*] \times \{0\})\end{aligned}$$

[Notice that $x_1^* = 0$.]

$$\begin{aligned} &= J_2(\{(0,1)^T\}) - J_2(\{(0,0)^T\}) + J_2(\{(0, 4(x_2^*)^3)^T\}) \\ &\quad + J_2([x_2^*, -x_2^*] \times \{0\}) \\ &= J_2(\{(0, 1 + 4(x_2^*)^3)^T\}) + J_2([x_2^*, -x_2^*] \times \{0\}) \\ &= J_2([x_2^*, -x_2^*] \times \{1 + 4(x_2^*)^3\})\end{aligned}$$

[Notice that $x_2^* = -\frac{1}{\sqrt[3]{4}}$, i.e., $1 + 4(x_2^*)^3 = 0$.]

$$= J_2([x_2^*, -x_2^*] \times \{0\}).$$

At the origin we obtain

$$\begin{aligned}\overrightarrow{\partial} f_{1,\rho}(0,0) &= J_2\left(\mathrm{co}\left(\tbinom{0}{2}, \tbinom{0}{1}\right)\right) - J_2\left(\mathrm{co}\left(\tbinom{0}{0}, \tbinom{0}{1}\right)\right) + J_2(\{(0,0)^T\}) \\ &\quad + \rho J_2(\{(0,0)^T\}) \\ &= J_2(\{(0,1)^T\}) + J_2(\{(0,0)^T\}) + \rho J_2(\{(0,0)^T\}) \\ &= J_2(\{(0,1)^T\}).\end{aligned}$$

(iii) considerations at the points x^* and at the origin

As calculated above, we have $\overrightarrow{\partial} f_{1,\rho}(x^*) = J_2([x_2^*, -x_2^*] \times \{0\})$ for all $\rho > 0$. This yields $P_2(\overrightarrow{\partial} f_{1,\rho}(x^*)) = [x_2^*, -x_2^*] \times \{0\}$. We conclude that

$$(0,0)^T \in P_2(\overrightarrow{\partial} f_{1,\rho}(x^*)) \qquad \text{for all } \rho > 0,$$

i.e., at the optimal point x^* of (**P**) the necessary optimality condition for a local minimizer of $f_{1,\rho}$ is satisfied for all $\rho > 0$.

Since the interior of $P_2(\overrightarrow{\partial} f_{1,\rho}(x^*))$ is empty, the sufficient condition is not satisfied at x^*.

At the origin we have $\overrightarrow{\partial} f_{1,\rho}(0,0) = J_2(\{(0,1)^T\})$ for all $\rho > 0$ (see above), and therefore neither the necessary optimality condition nor the sufficient optimality condition is satisfied. Since $f'_{1,\rho}((0,0)^T; (0,1)^T) = 1$ and $f'_{1,\rho}((0,0)^T; (0,-1)^T) = -1$, however, we can identify a strict saddle point in the sense of [5, Sec. 5].

(iv) comparison to the Clarke subdifferential

Consequently, neither the necessary optimality condition nor the sufficient optimality condition for a local minimizer of $f_{1,\rho}$ is satisfied at the origin for any $\rho > 0$. This is in contrast to the standard necessary optimality condition based on the Clarke subdifferential calculus. Analogously to the situation for f mentioned above in (3.12) and (3.13) we have that

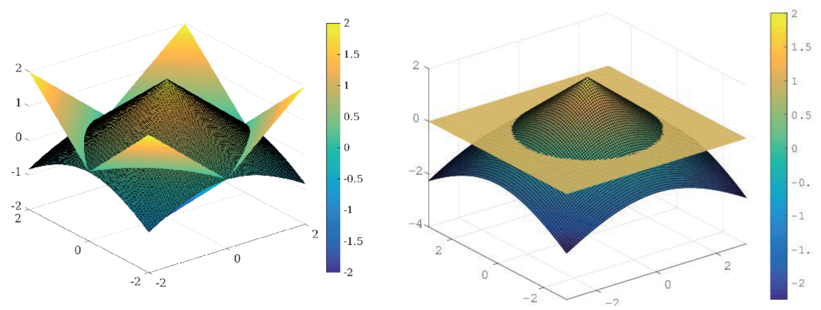
$$(0,0)^T \in \partial_{Cl} f_{1,\rho}(0,0) \qquad \text{for all } \rho > 0.$$

This can be seen as follows. Let $\rho > 0$ be arbitrary, and choose any sequence $(\alpha_k)_{k\in\mathbb{N}} \subset \{\alpha \in \mathbb{R} \mid \alpha > 0\}$ such that $\alpha_k \longrightarrow 0$. Set $\beta_k := \frac{1}{2}\alpha_k^2$ for all k. Then for all $k \in \mathbb{N}$ the function $f_{1,\rho}$ is differentiable at $(\alpha_k, \beta_k)^T$ with

$$\nabla f_{1,\rho}(\alpha_k, \beta_k) = \begin{pmatrix} 2\alpha_k \\ 1 \end{pmatrix} - \begin{pmatrix} 2\alpha_k \\ 1 \end{pmatrix} + \begin{pmatrix} 4\alpha_k^3 \\ 4\beta_k^3 \end{pmatrix} + \rho \begin{pmatrix} \beta_k \\ \alpha_k \end{pmatrix} = \begin{pmatrix} 4\alpha_k^3 + \rho\beta_k \\ 4\beta_k^3 + \rho\alpha_k \end{pmatrix}.$$

Since $\alpha_k \longrightarrow 0$ implies $\beta_k \longrightarrow 0$, we see that $\nabla f_{1,\rho}(\alpha_k, \beta_k) \longrightarrow (0,0)^T$ as $k \longrightarrow +\infty$. This shows that $(0,0)^T \in \partial f_{1,\rho}(0,0)$ (see, e.g., [19, Theorem 2.5.1]).

Example 3.12. Consider the special optimization problem (**P**) with a convex objective function and a single concave inequality constraint.



(a) obj. fct. f and constraint fct. g_1 (b) constraint fct. g_1 and feas. set $\mathcal{F}(\mathbf{P})$

FIGURE 4. Graphs of objective function and constraint (Example 3.12)

$$\begin{aligned} \min_{x \in \mathbb{R}^2} \quad & |x_1| + |x_2| - 2 \\ \text{s.t.} \quad & -\sqrt{x_1^2 + x_2^2} + 2 \le 0 \end{aligned}$$

The feasible set describes “the plane with a hole” (see Fig. 4(b)), i.e.,

$$\mathcal{F}(\mathbf{P}) = \mathbb{R}^2 \backslash \{ x \in \mathbb{R}^2 : \ \|x\|_2 < 2 \} = \{ x \in \mathbb{R}^2 : \ \|x\|_2 \ge 2 \} . \tag{3.18}$$

We set $g_1(x) := -\|x\|_2 + 2$ for all $x \in \mathbb{R}^2$. As in the examples above, we state that the standard assumptions A1–A2 are fulfilled, since $f(x) = \|x\|_1 - 2$ and g_1 are DC and $(2,0)^T \in \mathcal{F}(\mathbf{P})$.

(i) consideration of original problem

We first calculate all global minimizers of (**P**) by analytic means. We use that $\|y\|_1 \ge \|y\|_2$ for all $y \in \mathbb{R}^2$, and therefore

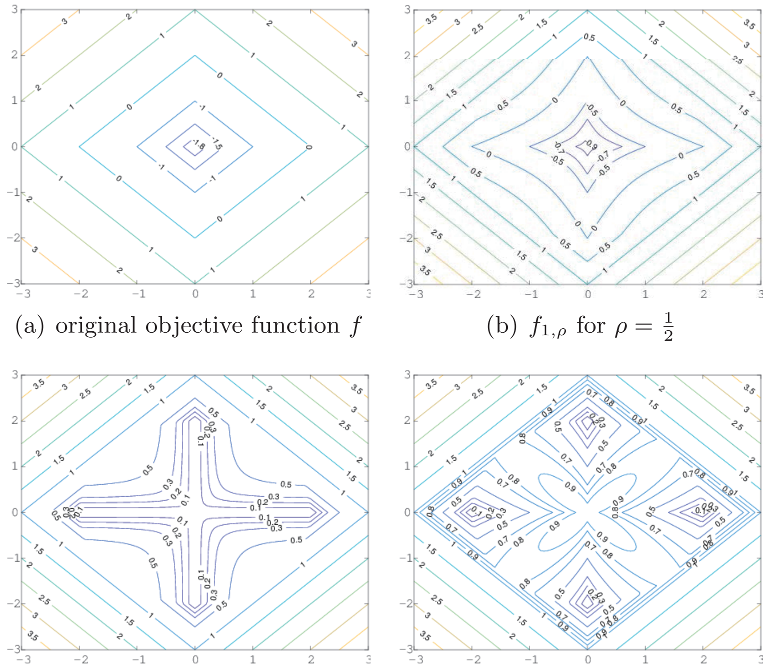
$$\|x\|_1 \ge \|x\|_2 \ge 2 \qquad \text{for all } x \in \mathcal{F}(\mathbf{P})$$

with equality if and only if $\|x\|_1 = \|x\|_2 = 2$. This yields the four (isolated) global minimizers of the problem

$$x^{*1} := (2,0)^T, \quad x^{*2} := (0,2)^T, \quad x^{*3} := (-2,0)^T, \quad x^{*4} := (0,-2)^T \tag{3.19}$$

located on the boundary of $\mathcal{F}(\mathbf{P})$. The corresponding optimal function value is $f^* = 0$.

As in the previous example, we provide some plots. Fig. 4(a) displays the graphs of the convex objective function f (light colors) as well as of the concave constraint function g_1 (dark colors) on the square box $[-2,+2]^2$. The four minimizers $x^{*\ell}$,

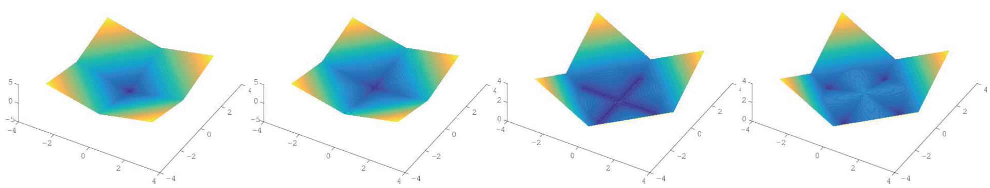


(a) original objective function f (b) $f_{1,\rho}$ for $\rho = \frac{1}{2}$

(c) $f_{1,\rho}$ for $\rho = \overline{\rho} = 1$ (d) $f_{1,\rho}$ for $\rho = \frac{3}{2}$

FIGURE 5. Level lines of objective function f and of $f_{1,\rho}$ for various values of ρ (Example 3.12)

$\ell = 1, 2, 3, 4$, are found on the coordinate axes where $f(x^{*\ell}) = g_1(x^{*\ell}) = 0$. Fig. 4(b) again displays the function graph of g_1 (dark colors) and the feasible set $\mathcal{F}(\mathbf{P})$ in the x_1-x_2-plane (light brown) (cf. (3.18)). Moreover, Fig. 5(a) shows the level lines of the objective function f on the square box $[-3, +3]^2$.



(a) original obj. fct. f (b) $f_{1,\rho}$ for $\rho = \frac{1}{2}$ (c) $f_{1,\rho}$ for $\rho = \overline{\rho} = 1$ (d) $f_{1,\rho}$ for $\rho = \frac{3}{2}$

FIGURE 6. Graphs of objective function f and of auxiliary functions $f_{1,\rho}$ for various values of ρ (Example 3.12)

(ii) consideration of auxiliary problem

Now we analyze the penalty problem $(\mathbf{P}_\rho)$ with penalty function $p \equiv p_1$ and (arbitrary) penalty parameter $\rho > 0$, i.e., minimization of the

$$f_{1,\rho}(x) = \|x\|_1 - 2 + \rho \max\{0, -\|x\|_2 + 2\} \qquad \text{for } x \in \mathbb{R}^2.$$

Fig. 6(a) (again) shows the graph of the objective function f. In comparison to this, the plots in the Figs. 6(b), 6(c), 6(d) show the graph of $f_{1,\rho}$ for the parameter values $\rho = \frac{1}{2}$, $\rho = 1$, and $\rho = \frac{3}{2}$, respectively. Accordingly, the level lines of the auxiliary function $f_{1,\rho}$ are displayed in the Figs. 5(b), 5(c), and 5(d) for the parameter values $\rho = \frac{1}{2}$, $\rho = 1$, and $\rho = \frac{3}{2}$, respectively.

(iii) necessary and sufficient optimality conditions by the directed subdifferential

Now we consider optimality conditions based on the directed subdifferential. Since

$$f_{1,\rho}(x_1, x_2) = f_{1,\rho}(|x_1|, |x_2|) = f_{1,\rho}(|x_2|, |x_1|) \qquad \text{for all } x \in \mathbb{R}^2, \tag{3.20}$$

it suffices to calculate the directed subdifferential only at points x from the set

$$X_0 := \{\, x \in \mathbb{R}^2 \mid x_1 \geq 0,\ x_2 > 0 \,\} \cup \{(0,0)^T\}\,.$$

These calculations are presented in Appendix A. We obtain

$$\overrightarrow{\partial} f_{1,\rho}(x) = \begin{cases} J_2\left(\left\{\begin{pmatrix} 1 - \rho\frac{x_1}{\|x\|_2} \\ 1 - \rho\frac{x_2}{\|x\|_2} \end{pmatrix}\right\}\right) & \text{if } x_1 > 0,\ x_2 > 0,\ \|x\|_2 < 2, \\ J_2\left(\binom{1}{1} + \rho \cdot \operatorname{co}\left\{\binom{0}{0}, -\frac{1}{2}x\right\}\right) & \text{if } x_1 > 0,\ x_2 > 0,\ \|x\|_2 = 2, \\ J_2\left(\left\{\binom{1}{1}\right\}\right) & \text{if } x_1 > 0,\ x_2 > 0,\ \|x\|_2 > 2, \\ J_2([-1,1]^2) - \rho\, J_2(B_1(0,0)) & \text{if } x = (0,0)^T, \\ J_2([-1,1] \times \{1-\rho\}) & \text{if } x_1 = 0,\ 0 < x_2 < 2, \\ J_2([-1,1] \times [1-\rho, 1]) & \text{if } x = (0,2)^T, \\ J_2([-1,1] \times \{1\}) & \text{if } x_1 = 0,\ x_2 > 2 \end{cases} \tag{3.21}$$

where $B_1(0,0) := \{s \in \mathbb{R}^2 : \|s\|_2 \leq 1\}$. From these formulas we calculate the points $x \in X_0$ satisfying the necessary conditions for a local minimizer auxiliary function of $f_{1,\rho}$,

$$X_{0,\rho}^{\text{nec}} := \{\, x \in X_0 \mid (0,0)^T \in P_2(\overrightarrow{\partial} f_{1,\rho}(x)) \,\},$$

and get (see Appendix A, eq. (A.2))

$$X_{0,\rho}^{\text{nec}} = \begin{cases} \{(0,0)^T\} & \text{if } \rho \in (0,1), \\ \{(0,x_2)^T \mid 0 \leq x_2 \leq 2\} & \text{if } \rho = 1, \\ \{(0,2)^T\} & \text{if } \rho \in (1,\sqrt{2}), \\ \{(\delta,\delta)^T \mid 0 < \delta \leq \sqrt{2}\} \cup \{(0,2)^T\} & \text{if } \rho = \sqrt{2}, \\ \{(\sqrt{2},\sqrt{2})^T\} \cup \{(0,2)^T\} & \text{if } \rho > \sqrt{2}. \end{cases} \tag{3.22}$$

For each $\rho > 0$ the function $f_{1,\rho}$ is continuous and coercive (i.e., $f_{1,\rho}(y^k) \longrightarrow +\infty$ for all sequences $(y^k)_k \subset \mathbb{R}^2$ with $\|y^k\|_1 \longrightarrow +\infty$). Hence, the infimal function value of $f_{1,\rho}$ is attained, i.e., $f_{1,\rho}$ possesses global minimizers. Therefore, (3.20) shows that global minimizers of $f_{1,\rho}$ are contained in the set X_0. By Proposition 3.8 these global minimizers are contained in the set $X_{0,\rho}^{\text{nec}}$. Summarizing, for each $\rho > 0$ we find all global minimizers of $f_{1,\rho}$ on X_0 by picking out the points in $X_{0,\rho}^{\text{nec}}$ with the smallest function values. By this we obtain

$$\operatorname*{argmin}_{x\in X_0} f_{1,\rho}(x) = \begin{cases} \{(0,0)^T\} & \text{if } \rho \in (0,1), \\ \{(0,x_2)^T \mid 0 \le x_2 \le 2\} & \text{if } \rho = 1, \\ \{(0,2)^T\} & \text{if } \rho > 1 \end{cases} \tag{3.23}$$

with corresponding optimal function value

$$f^*_{1,\rho} = \begin{cases} 2\rho - 2 & \text{if } \rho \in (0,1), \\ 0 & \text{if } \rho \ge 1. \end{cases} \tag{3.24}$$

Of course, the set of points in X_0 satisfying the sufficient conditions for a local minimizer of $f_{1,\rho}$,

$$X^{\text{suff}}_{0,\rho} := \{\, x \in X_0 \mid (0,0)^T \in \operatorname{int}(P_2(\overrightarrow{\partial} f_{1,\rho}(x)))\,\},$$

must be contained in the set of minimizers $\operatorname*{argmin}_{x\in X_0} f_{1,\rho}(x)$. Let us verify this. By calculations in Appendix A (see eq. (A.3)) we have

$$X^{\text{suff}}_{0,\rho} = \begin{cases} \{(0,0)^T\} & \text{if } \rho \in (0,1), \\ \varnothing & \text{if } \rho = 1, \\ \{(0,2)^T\} & \text{if } \rho > 1, \end{cases} \tag{3.25}$$

and together with (3.23) we see that, indeed, $X^{\text{suff}}_{0,\rho} \subset \operatorname*{argmin}_{x\in X_0} f_{1,\rho}(x)$ holds for all $\rho > 0$ (we even have an equality for all $\rho \neq 1$).

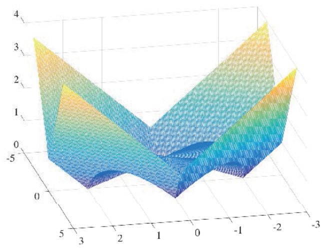
From (3.23) and (3.20) we deduce that the set of all global minimizers of $f_{1,\rho}$ is given by

$$\operatorname*{argmin}_{x\in\mathbb{R}^2} f_{1,\rho}(x) = \begin{cases} \{(0,0)^T\} & \text{if } \rho \in (0,1), \\ \Big(\{0\}\times[-2,+2]\Big) \cup \Big([-2,+2]\times\{0\}\Big) & \text{if } \rho = 1, \\ \{(0,2)^T, (-2,0)^T, (0,-2)^T, (2,0)^T\} & \text{if } \rho \in (1,+\infty). \end{cases} \tag{3.26}$$

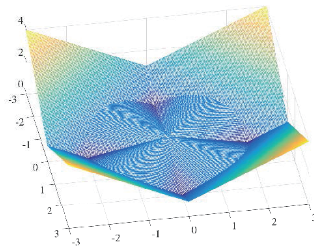
By comparison to (3.19) we observe the exactness of the penalty with $\overline{\rho} := 1$. More precisely, for each $\rho \ge \overline{\rho}$ each global minimizer of (**P**) is also a global minimizer of $f_{1,\rho}$. For $\rho > \overline{\rho} = 1$ also the opposite implication holds.

The above derived results are also nicely observed in the plots of the function graphs and the level lines. Fig. 6(a) (graph of objective function f) and 6(b) (graph of $f_{1,\rho}$ for $\rho = \frac{1}{2}$) illustrate that only the origin $x = 0$ is optimal. This is also illustrated by the level lines (see Figs. 5(a) and 5(b), respectively). For $\rho = 1$ we observe that all the points in $(\{0\}\times[-2,+2])\cup([-2,+2]\times\{0\})$ are optimal (see Fig. 6(c) for the graph of $f_{1,\rho=1}$, and Fig. 5(c) for corresponding level lines). Finally, for $\rho = \frac{3}{2}$ we observe that $f_{1,\rho}$ possesses the four global minimizers $(0,2)^T, (-2,0)^T, (0,-2)^T, (2,0)^T$ (see Fig. 6(d) and 5(d)). Moreover, one detects that the origin is a local maximizer.

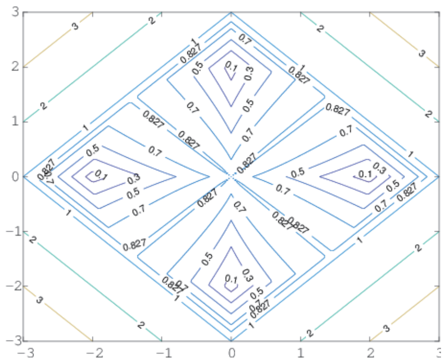
From the expressions in (3.22) and from the calculations in Appendix A we argue that the situation $\rho = \sqrt{2}$ is special. The plots in Fig. 7 show the graph and the level lines of $f_{1,\rho=\sqrt{2}}$. One can see that for this value of ρ the function values $f_{1,\rho=\sqrt{2}}(x)$



(a) auxiliary function $f_{1,\rho=\sqrt{2}}$



(b) auxiliary function $f_{1,\rho=\sqrt{2}}$



(c) contour lines of $f_{1,\rho=\sqrt{2}}$

FIGURE 7. Graphs and contour lines of auxiliary functions $f_{1,\rho}$ for $\rho = \sqrt{2}$ (Example 3.12)

for points x contained in the "cross-shaped" set

$$\{(\delta,\delta)^T \mid \delta \in [-\sqrt{2},+\sqrt{2}]\} \cup \{(-\delta,\delta)^T \mid \delta \in [-\sqrt{2},+\sqrt{2}]\}$$

are constant, and, furthermore, these points are local maximizers of $f_{1,\rho=\sqrt{2}}$. Since $f_{1,\rho=\sqrt{2}}$ is differentiable at these points x, the directed subdifferential $\overrightarrow{\partial} f_{1,\rho=\sqrt{2}}(x)$ consists of only an embedded point.

(iv) visualized directed subdifferential of auxiliary function at minimizer

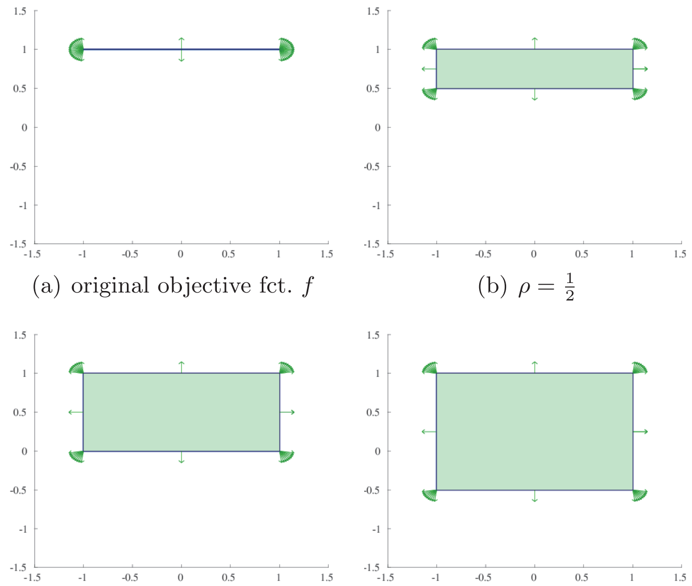
Let us have a look at the visualization of the directed subdifferential at an optimal point of (**P**). Just as an example, let us consider the optimal point

$$x^{*2} = (0,2)^T\,.$$

For any $\rho > 0$ the directed subdifferential of $f_{1,\rho}$ at this point is given by

$$\overrightarrow{\partial} f_{1,\rho}(x^{*2}) = J_2([-1,+1] \times [1-\rho,1])$$

(see (3.21)). Fig. 8 displays visualizations of different directed subdifferentials at the point $x^{*2} = (0,2)^T$. Subfigure 8(a) shows the directed subdifferential of the objective function f of the original problem (**P**). Subfigures 8(b) to 8(d) show the subdifferentials of the auxiliary function $f_{1,\rho}$ for the values $\rho = \frac{1}{2}$, $\rho = \overline{\rho} = 1$, and $\rho = \frac{3}{2}$, respectively.



(a) original objective fct. f (b) $\rho = \frac{1}{2}$

(c) $\rho = \overline{\rho} = 1$ (d) $\rho = \frac{3}{2}$

FIGURE 8. The directed subdifferential $\overrightarrow{\partial} f_{1,\rho}(x^{*2})$ for various values of ρ (Example 3.12)

As already seen above, the necessary optimality condition for a local minimizer of $f_{1,\rho}$ at the point $x^{*2} = (0,2)^T$

$$(0,0)^T \in P_2(\overrightarrow{\partial} f_{1,\rho}(x^{*2}))$$

(i.e., $(0,2)^T \in X_{0,\rho}^{\text{nec}}$) is satisfied if and only if $\rho \geq \overline{\rho} = 1$ (see (3.22)). This is also nicely observed in the plots of Fig. 8.

Analogously, the sufficient optimality condition

$$(0,0)^T \in \operatorname{int}(P_2(\overrightarrow{\partial} f_{1,\rho}(x^{*2})))$$

for a local minimizer of $f_{1,\rho}$ at the point $x^{*2} = (0,2)^T$ is satisfied if and only if $\rho > \overline{\rho} = 1$ (see (3.25)). This is also nicely seen from the plots (see, e.g., Subfig. 8(d)).

(v) sufficient optimality condition by the directed subdifferential for local maximizer at the origin

We mention that for all $\rho > \sqrt{2}$ the origin $\hat{x} = (0,0)^T$ is a local maximizer of $f_{1,\rho}$. This is "detected" by the calculus of the directed subdifferential

$$\overrightarrow{\partial} f_{1,\rho}(0,0) = J_2([-1,1]^2) - \rho J_2(B_1(0,0))$$

(see (3.21)). This difference is closer analyzed in [5, Example 5.7] and in [4, Example 3.20]. In these papers it is illustrated that

$$(0,0)^T \in \operatorname{int} N_2\Big(J_2([-1,1]^2) - \rho J_2(B_1(0,0))\Big) \qquad \Longleftrightarrow \qquad \rho > \sqrt{2}\,.$$

Consequently, for all $\rho > \sqrt{2}$ the sufficient optimality conditions for a local maximizer hold at $\hat{x} = (0,0)^T$. The situation in this example parallels the situation at the (1D)-origin in Example 3.10.

(vi) comparison to the Clarke subdifferential

We close this example with the remark that the use of the Clarke subdifferential calculus generally leads to a larger set of critical points. For example, at the origin $\hat{x} = (0,0)^T$ we have (see below)

$$(0,0)^T \in \partial_{Cl} f_{1,\rho}(\hat{x}) \qquad \text{for all } \rho > 0 \tag{3.27}$$

in contrast to $\hat{x} \in X^{\text{nec}}_{0,\rho}$ iff $\rho \in (0,1]$ (compare to (3.22)).
A simple proof for the inclusion (3.27) is as follows. Let $\rho > 0$, and set $e := (1,1)^T$. Then $f_{1,\rho}$ is differentiable at all points $x = \pm\delta e$ for all $\delta \in (0, \frac{1}{2}\sqrt{2})$ with $\nabla f_{1,\rho}(\pm\delta e) = \pm(1 - \rho/\sqrt{2})e$ for all δ. Hence, trivially, $\lim_{\delta \searrow 0} \nabla f_{1,\rho}(\pm\delta e) = \pm(1-\rho/\sqrt{2})e$, and thus $\pm(1-\rho/\sqrt{2})e \in \partial f_{1,\rho}(0,0)$ (see, e.g., [19, Theorem 2.5.1]). If $\rho = \sqrt{2}$ then we are done. If $\rho \neq \sqrt{2}$ then we use that $\partial f_{1,\rho}(0,0)$ is a convex set and that, trivially, $(0,0)^T \in \text{co}(-(1-\rho/\sqrt{2})e, +(1-\rho/\sqrt{2})e)$.

4. Lagrange duality

In this section we apply Lagrange duality to problem (**P**).

4.1. **Theory.** In the following we briefly state the needed basic notations and results. For an introduction to Lagrange duality we refer, e.g., to [13, Chap. 6] and to [27, Chap. 11].

We start by defining the *Lagrange function of problem* (**P**) (the *Lagrangian of* (**P**) for short),

$$\begin{aligned} &\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^p \longrightarrow \mathbb{R}, \\ &\mathcal{L}(x;\lambda,\mu) := f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{j=1}^p \mu_j h_j(x) \,. \end{aligned} \tag{4.1}$$

Furthermore, we define the *Lagrange dual function of problem* (**P**),

$$\theta : \mathbb{R}^m_+ \times \mathbb{R}^p \longrightarrow \mathbb{R} \cup \{-\infty\}, \qquad \theta(\lambda,\mu) := \inf_{x \in \mathbb{R}^n} \mathcal{L}(x;\lambda,\mu) \,. \tag{4.2}$$

With this notation the *(standard) Lagrange-dual problem of* (**P**) can be stated as

$$(\mathbf{D}) \qquad \begin{array}{cc} \max\limits_{\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^p} & \theta(\lambda,\mu) \\ \text{s.t.} & \lambda \geq 0. \end{array}$$

The optimal function value of this problem (whether attained or not attained; whether finite or not finite) is denoted by

$$\theta^* := \sup_{\lambda \in \mathbb{R}^m_+, \mu \in \mathbb{R}^p} \theta(\lambda,\mu) \qquad \in \mathbb{R} \cup \{-\infty, +\infty\}.$$

We mention that $\theta^* = -\infty$ is possible iff $\theta(\lambda,\mu) = -\infty$ for all $(\lambda,\mu) \in \mathbb{R}^m_+ \times \mathbb{R}^p$. Moreover, we mention that the case $\theta^* = +\infty$ does not occur in our framework, as is shown in the following theorem. In this (standard) theorem we collect the main properties and relations of (**P**) and (**D**).

Theorem 4.1. *Consider problem* (**P**) *and its associated Lagrange-dual problem* (**D**)*. Then the following assertions hold.*

(a) *("weak duality I") For the feasible points in both problems, respectively, we have the inequalities*

$$f(x) \geq \theta(\lambda, \mu) \qquad \textit{for all } x \in \mathcal{F}(\mathbf{P}) \textit{ and all } (\lambda, \mu) \in \mathbb{R}^m_+ \times \mathbb{R}^p.$$

(b) *("weak duality II") For the optimal values f^* and θ^* of* (**P**) *and* (**D**)*, respectively, we have the inequalities*

$$+\infty > \ f^* \ \geq \ \theta^* \ \geq -\infty.$$

(c) *("strong duality", simple situation) Let $(\lambda^*, \mu^*) \in \mathbb{R}^m_+ \times \mathbb{R}^p$ be a maximizer of problem* (**D**)*, and let $x^* \in \mathcal{F}(\mathbf{P})$ be a point such that $\theta(\lambda^*, \mu^*) = f(x^*)$. Then x^* is a solution of problem* (**P**)*.*

Proof: Assertion (a) follows directly from the definitions (or see, e.g., [13, Theorem 6.2.1]). By our general assumption $\mathcal{F}(\mathbf{P}) \neq \varnothing$ (see A1 in (2.6)) we know that $f^* < +\infty$. Hence, the assertions in (b) and in (c) are simple consequences of (a).□

Remark 4.2 (duality gap)**.** From assertion (b) of this theorem we see that $+\infty > f^* \geq \theta^*$. Moreover, $\theta^* > -\infty$ if and only if there is at least one point $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^m_+ \times \mathbb{R}^p$ with $\theta(\bar{\lambda}, \bar{\mu}) > -\infty$. In this situation the (finite!) value $(f^* - \theta^*) \in \mathbb{R}$ is called the *duality gap of the problems* (**P**) *and* (**D**). In the situation of Theorem 4.1(c) the duality gap is zero.
In the pathologic case where the mapping θ is identically to $-\infty$, i.e., $\theta^* = -\infty$, the value $(f^* - \theta^*) = +\infty$ is also denoted as *duality gap* ("infinite duality gap").

As Theorem 4.1(b) shows, we have $\theta(\lambda, \mu) < +\infty$ for all $(\lambda, \mu) \in \mathbb{R}^m_+ \times \mathbb{R}^p$, i.e., $-\theta(\lambda, \mu) > -\infty$ for all $(\lambda, \mu) \in \mathbb{R}^m_+ \times \mathbb{R}^p$. Recall that a function $\Phi : Y \longrightarrow \mathbb{R} \cup \{+\infty\}$ not identically $+\infty$ defined on a non-empty convex set $Y \subset \mathbb{R}^{m+p}$ is called *(proper) convex (on Y)* iff

$$\Phi(\lambda y^1 + (1-\lambda) y^2) \leq \lambda \Phi(y^1) + (1-\lambda)\Phi(y^2) \quad \text{for all } \lambda \in]0,1[\text{ and all } y^1, y^2 \in Y$$

(including infinite function values of Φ), see, e.g., [51, p. 5]. A function $\Psi \colon Y \longrightarrow \mathbb{R} \cup \{-\infty\}$ not identically $-\infty$ defined on a non-empty convex set $Y \subset \mathbb{R}^{m+p}$ is called *concave (on Y)* iff $(-\Psi)$ is convex.

Proposition 4.3. *Let the Lagrange-dual function θ not be identical $-\infty$. Then the function θ is concave on $\mathbb{R}^m_+ \times \mathbb{R}^p$, i.e., the function $(-\theta)$ is convex on $\mathbb{R}^m_+ \times \mathbb{R}^p$.*

Proof: The function $-\theta$ is not identical to $+\infty$, by assumption. Moreover, $-\theta$ does not attain the value $-\infty$ by Theorem 4.1(a). Finally, $-\theta$ is the point-wise supremum function of affine-linear (i.e., convex) functions, and thus is convex itself (see, e.g., [51, Proposition 2.9]). □

This result shows that the dual problem (**D**) can be treated as a convex optimization problem when written in the form

$$(\mathbf{D}') \qquad \begin{array}{cl} \displaystyle\min_{\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^p} & (-\theta)(\lambda, \mu) \\ \text{s.t.} & \lambda \geq 0 \end{array}$$

possessing optimal value $-\theta^*$.

Coming back to the intention of the paper, we would like to apply optimality conditions of unconstrained(!) optimization in terms of directed subdifferentials. If there are no inequality constraints in problem (**P**), i.e., $m = 0$, then problem (**D**) (or, equivalently, problem (**D′**)) is an unconstrained problem,

$$\max_{\mu \in \mathbb{R}^p} \theta(\mu), \qquad \text{resp.} \qquad \min_{\mu \in \mathbb{R}^p} (-\theta)(\mu).$$

By the use of optimality conditions, we may be able to calculate the value $\theta(\mu)$ of the dual function at arbitrary points μ. Using the optimality conditions once more, we may then determine a (globally) optimal point μ^* of the Lagrange-dual problem. In the presence of inequality constraints in (**P**), i.e., $m > 0$, even the simple sign constraints on the variables λ_i in problem (**D**) are an obstacle for the direct application of optimality conditions for unconstrained(!) optimization. Since we have applications in mind, however, where the functions f, g_i, h_j (or only some of them) are nonsmooth anyway, we do not see a hindrance to rewrite problem (**P**) in the following form without(!) inequality constraints. Consider the auxiliary functions

$$\widehat{g}_i : \mathbb{R}^n \longrightarrow \mathbb{R}, \qquad \widehat{g}_i(x) := \max\{0, g_i(x)\}, \qquad i = 1, \dots, m \,. \tag{4.3}$$

Then (**P**) takes the form

$$(\widehat{\mathbf{P}}) \qquad \begin{array}{rl} \min\limits_{x \in \mathbb{R}^n} & f(x) \\ \text{s.t.} & \widehat{g}_i(x) = 0 \quad (i = 1, \dots, m), \\ & h_j(x) = 0 \quad (j = 1, \dots, p). \end{array}$$

The two problems (**P**) and ($\widehat{\mathbf{P}}$) are equivalent in the sense that both problems possess the same feasible set $\mathcal{F}(\mathbf{P})$ and the same solutions x^*.

The Lagrangian of problem ($\widehat{\mathbf{P}}$) is the function (compare to (4.1))

$$\begin{aligned} &\widehat{\mathcal{L}} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \longrightarrow \mathbb{R}, \\ &\widehat{\mathcal{L}}(x; \lambda, \mu) := f(x) + \sum_{i=1}^{m} \lambda_i \widehat{g}_i(x) + \sum_{j=1}^{p} \mu_j h_j(x) \end{aligned} \tag{4.4}$$

where now λ is taken from entire space $\mathbb{R}^m$ (instead from $\mathbb{R}^m_+$ only). In analogous manner, the Lagrange-dual function of problem ($\widehat{\mathbf{P}}$) is defined on the whole space $\mathbb{R}^m \times \mathbb{R}^p$,

$$\widehat{\theta} : \mathbb{R}^m \times \mathbb{R}^p \longrightarrow \mathbb{R} \cup \{-\infty\}, \qquad \widehat{\theta}(\lambda, \mu) := \inf_{x \in \mathbb{R}^n} \widehat{\mathcal{L}}(x; \lambda, \mu) \tag{4.5}$$

(compare to (4.2)). We end up with the corresponding Lagrange-dual problem of ($\widehat{\mathbf{P}}$) which is unconstrained(!),

$$(\widehat{\mathbf{D}}) \qquad \max_{\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^p} \widehat{\theta}(\lambda, \mu) \,.$$

Remark 4.4. Notice that in the situation $m = 1$ and $p = 0$ we obtain a direct relation to the l_1-penalty approach from the previous section. In this situation the definition of $\widehat{g}_1$ (see (4.3)) and the definition of $\widehat{\mathcal{L}}$ (see (4.4)) yield

$$\widehat{\mathcal{L}}(x; \lambda) = f(x) + \lambda \widehat{g}_1(x) = f(x) + \lambda \max\{0, g_1(x)\}$$

for all $x \in \mathbb{R}^n$ and all $\lambda \in \mathbb{R}$ (compare to (3.4) for $m := 1$ and $p := 0$). This means, we have the identity

$$\widehat{\mathcal{L}}(x;\lambda) = f_{1,\rho}(x) \qquad \text{for all } x \in \mathbb{R}^n \text{ and all } \lambda = \rho > 0.$$

4.2. **Examples.**

Example 4.5 (Example 3.10 cont'd)**.** Let us have a look at problem (**P**) from Example 3.10 in its equivalent form ($\widehat{\mathbf{P}}$), i.e., with the inequality constraint expressed as an equality constraint. Its Lagrangian is

$$\widehat{\mathcal{L}}(x;\lambda) = x^2 - 8|x| + 7 + \lambda \max\{0, |x| - 1\} \qquad \text{for } (x,\lambda) \in \mathbb{R} \times \mathbb{R},$$

and the Lagrange-dual map of ($\widehat{\mathbf{P}}$) is given by

$$\widehat{\theta}(\lambda) = \inf_{x \in \mathbb{R}} \Big\{ x^2 - 8|x| + 7 + \lambda \max\{0, |x| - 1\} \Big\} \qquad \text{for } \lambda \in \mathbb{R}.$$

As outlined in Rem 4.4 above, for $\lambda > 0$ the inner min-problem is identical to the l_1-penalty problem solved with $\rho := \lambda$ (see equations (3.7) and (3.8) in Example 3.10). For $\lambda \leq 0$ we obtain the same formulas as for $\rho \in]0, 6[$ (repeat the considerations in Case 1 of Example 3.10), i.e., summarizing,

$$\underset{x \in \mathbb{R}}{\operatorname{argmin}}\, \widehat{\mathcal{L}}(x;\lambda) = \begin{cases} \{-(4 - \frac{\lambda}{2}), +(4 - \frac{\lambda}{2})\} & \text{if } \lambda \in (-\infty, 6), \\ \{-1, +1\} & \text{if } \lambda \in [6, +\infty), \end{cases} \tag{4.6}$$

and

$$\widehat{\theta}(\lambda) = \begin{cases} -\frac{\lambda^2}{4} + 3\lambda - 9 \ (< 0) & \text{if } \lambda \in (-\infty, 6), \\ 0 & \text{if } \lambda \in [6, +\infty). \end{cases}$$

It is easy to see that this function is differentiable (and concave, as expected by Proposition 4.3 above). In the fashion of the considerations in this paper, we apply optimality conditions based on the directed subdifferential. First we apply Proposition 2.6 and obtain that

$$\overrightarrow{\partial}\widehat{\theta}(\lambda) = \begin{cases} \overrightarrow{\{-\frac{1}{2}\lambda + 3\}} & \text{if } \lambda \in (-\infty, 6), \\ \overrightarrow{\{0\}} & \text{if } \lambda \in [6, +\infty). \end{cases}$$

We see that $N_1(\overrightarrow{\partial}\widehat{\theta}(\lambda^*)) = \{-\frac{1}{2}\lambda + 3\}$ for all $\lambda < 6$, and $N_1(\overrightarrow{\partial}\widehat{\theta}(\lambda^*)) = \{0\}$ for all $\lambda \geq 6$. Hence, the necessary optimality condition $0 \in N_1(\overrightarrow{\partial}\widehat{\theta}(\lambda^*))$ for a local maximizer is satisfied if and only if $\lambda^* \geq 6$. The sufficient optimality condition $0 \in \operatorname{int}(N_1(\overrightarrow{\partial}\widehat{\theta}(\lambda)))$ for a local maximizer is not satisfied at any λ. Nevertheless, the function $\widehat{\theta}$ is constant on $[6, +\infty)$, concave on $\mathbb{R}$ (see above) and finite. Hence, the set of optimal points of the Lagrange-dual problem ($\widehat{\mathbf{D}}$) is given by

$$\underset{\lambda \in \mathbb{R}}{\operatorname{argmax}}\, \widehat{\theta}(\lambda) = \{\lambda^* \mid \lambda^* \geq 6\}$$

(coinciding with the set of points satisfying the necessary optimality conditions, cf. the paragraph before). The corresponding optimal function value is

$$\widehat{\theta}^* = 0\ .$$

Since the points $x^{*1} = -1$ and $x^{*2} = +1$ are feasible for (**P**), and since $f(-1) = f(+1) = 0 = \widehat{\theta}^*$, Theorem 4.1(c) proves that $x^{*1} = -1$ and $x^{*2} = +1$ are global minimizers of problem (**P**). Moreover, we see that the duality gap is zero (see Rem 4.2).

Example 4.6 (Example 3.12 cont'd)**.** We consider Lagrange-duality for Example 3.12. For this, we rewrite problem (**P**) in its equivalent form ($\widehat{\mathbf{P}}$), and consider the corresponding Lagrangian $\widehat{\mathcal{L}}$ and its dual problem ($\widehat{\mathbf{D}}$). Since $m = 1$ and $p = 0$, the calculation of the dual function $\widehat{\theta}$ for arguments $\lambda > 0$ is just a copy of some (main) results in Example 3.12.
For $\lambda \leq 0$ we may use optimality conditions as well. Alternatively, we may use the following elementary arguments:
If $\|x\|_2 < 2$ then $x \notin \mathcal{F}(\mathbf{P})$, i.e., (notice that $\lambda \leq 0$, i.e., $-\lambda = |\lambda|$)

$$\widehat{\mathcal{L}}(x;\lambda) = \|x\|_1 - 2 + \lambda(2 - \|x\|_2) = 2\lambda - 2 + \|x\|_1 + |\lambda| \cdot \|x\|_2 \geq 2\lambda - 2.$$

Equality holds iff $\|x\|_1 + |\lambda| \cdot \|x\|_2 = 0$, i.e., iff $x = (0,0)^T$.
If $\|x\|_2 \geq 2$ then $x \in \mathcal{F}(\mathbf{P})$, i.e., $\widehat{g}_1(x) = 0$, and thus

$$\widehat{\mathcal{L}}(x;\lambda) = \|x\|_1 - 2 \geq \|x\|_2 - 2 \geq 0 > -2 \geq 2\lambda - 2.$$

All in all, this shows that

$$\operatorname*{argmin}_{x\in\mathbb{R}^2} \widehat{\mathcal{L}}(x,\lambda) = \{(0,0)^T\}, \qquad \text{and} \qquad \widehat{\theta}(\lambda) = 2\lambda - 2 \qquad \text{for all } \lambda \leq 0.$$

Together with the results from Example 3.12 (see (3.24) and (3.26)) we summarize that

$$\begin{aligned} &\operatorname*{argmin}_{x\in\mathbb{R}^2} \widehat{\mathcal{L}}(x;\lambda) \\ &= \begin{cases} \{(0,0)^T\} & \text{if } \lambda \in (-\infty, 1), \\ \Big(\{0\} \times [-2,+2]\Big) \cup \Big([-2,+2] \times \{0\}\Big) & \text{if } \lambda = 1, \\ \{(0,2)^T, (-2,0)^T, (0,-2)^T, (2,0)^T\} & \text{if } \lambda \in (1,+\infty), \end{cases} \end{aligned} \tag{4.7}$$

and

$$\widehat{\theta}(\lambda) = \begin{cases} 2\lambda - 2 & \text{if } \lambda \in (-\infty, 1), \\ 0 & \text{if } \lambda \in [1, +\infty). \end{cases} \tag{4.8}$$

We observe that the dual function $\widehat{\theta}$ is piecewise linear (and concave, as expected by Proposition 4.3). Although this function is extremely simple, we stick to our methodology and apply optimality conditions. For this we write $\widehat{\theta}$ in the form

$$\widehat{\theta}(\lambda) = \min\{2\lambda - 2, 0\} = -\max\{2 - 2\lambda, 0\} \qquad \text{for all } \lambda \in \mathbb{R}.$$

For $\lambda \neq 1$ we apply Proposition 2.6

$$\overrightarrow{\partial}\widehat{\theta}(1) = -J_1(\partial(-\widehat{\theta})(1)) = -J_1([-2,0]) = -\overrightarrow{[-2,0]}\,,$$

and for $\lambda = 1$ we apply Remark 2.2. Summarizing, the directed subdifferential of $\widehat{\theta}$ is given by

$$\overrightarrow{\partial}\widehat{\theta}(\lambda) = \begin{cases} \overrightarrow{\{2\}} & \text{if } \lambda < 1, \\ -\overrightarrow{[-2,0]} & \text{if } \lambda = 1, \\ \overrightarrow{\{0\}} & \text{if } \lambda > 1. \end{cases}$$

The necessary optimality condition for a local maximizer $0 \in N_1(\overrightarrow{\partial}\widehat{\theta}(\lambda^*))$ is satisfied if and only if $\lambda^* \geq 1$. The sufficient optimality condition for a local maximizer $0 \in \operatorname{int}(N_1(\overrightarrow{\partial}\widehat{\theta}(\lambda^*)))$ is never satisfied.
Clearly, the optimal value of the Lagrange-dual problem $\max_\lambda \widehat{\theta}(\lambda)$ is zero,

$$\widehat{\theta}^* = 0.$$

Since $f^* = 0$ (see Example 3.12), we see that also in this example the duality gap is zero. Once more we obtain that the four points $\{x^{*\ell} \mid \ell = 1,2,3,4\} = \{\pm 2e^i \mid i = 1,2\}$ are solutions of $(\mathbf{P})$ by the application of Theorem 4.1(c).

5. Saddle point optimality conditions

In this section we discuss saddle point optimality criteria, yet another approach for the application of optimality conditions of unconstrained optimization for the detection of solutions of the constrained problem $(\mathbf{P})$.

5.1. **Theory.** As in the previous section we first sketch the standard development, and then change it to our purposes.

Definition 5.1. Consider the Lagrange function $\mathcal{L}$ of problem $(\mathbf{P})$ (see (4.1) above). A triple $(x^*, \lambda^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^p$ is called a *saddle point of* $\mathcal{L}$ if the inequalities

$$\mathcal{L}(x^*; \lambda, \mu) \leq \mathcal{L}(x^*; \lambda^*, \mu^*) \leq \mathcal{L}(x; \lambda^*, \mu^*)$$

hold for all $x \in \mathbb{R}^n$ and all $(\lambda, \mu) \in \mathbb{R}^m_+ \times \mathbb{R}^p$.

We refer to the literature stating standard relations of saddle points of $\mathcal{L}$ and solutions of $(\mathbf{P})$ and $(\mathbf{D})$ (see, e.g., [13, Theorem 6.2.5]). As will become clear below, from analogous reasons as in the previous chapter we want to avoid the constraint "$\lambda \in \mathbb{R}^m_+$" in the saddle point conditions if $m > 0$. Therefore, again consider problem $(\mathbf{P})$ in its equivalent form $(\widehat{\mathbf{P}})$ and its corresponding Lagrangian $\widehat{\mathcal{L}}$ (see previous section). Moreover, consider the Lagrange-dual problem $(\widehat{\mathbf{D}})$ of $(\widehat{\mathbf{P}})$. Since problem $(\widehat{\mathbf{P}})$ possesses only equality constraints, Definition 5.1 directly leads to the inequalities

$$\widehat{\mathcal{L}}(x^*; \lambda, \mu) \leq \widehat{\mathcal{L}}(x^*; \lambda^*, \mu^*) \leq \widehat{\mathcal{L}}(x; \lambda^*, \mu^*) \quad \text{for all } x \in \mathbb{R}^n \text{ and all } (\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p$$

defining a saddle point of $\widehat{\mathcal{L}}$. The next (standard) result fully characterizes saddle points of the Lagrangian $\widehat{\mathcal{L}}$.

Proposition 5.2. *The following three assertions are equivalent.*

- (a) *The triple* $(x^*, \lambda^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ *is a saddle point of* $\widehat{\mathcal{L}}$.
- (b) *The following two conditions (b1) and (b2) are satisfied.*
  - (b1) $\widehat{\mathcal{L}}(x^*; \lambda^*, \mu^*) = \min\{\widehat{\mathcal{L}}(x; \lambda^*, \mu^*) \mid x \in \mathbb{R}^n\}$
  - (b2) *The point* x^* *is feasible for problem* $(\mathbf{P})$, *i.e.,* $x^* \in \mathcal{F}(\mathbf{P})$.
- (c) *The following three conditions (c1) to (c3) are satisfied.*
  - (c1) *The point* x^* *is a solution of the original problem* $(\mathbf{P})$.
  - (c2) *The pair* (λ^*, μ^*) *is a solution of the Lagrange-dual problem* $(\widehat{\mathbf{D}})$.

(c3) *The duality gap of* $(\widehat{\mathbf{P}})$ *and* $(\widehat{\mathbf{D}})$ *is zero, i.e.,* $+\infty > f^* = \widehat{\theta}^* > -\infty$.

The proof of this proposition is a student's exercise (or apply, e.g., [13, Theorem 6.2.5], and notice that $\mathcal{F}(\mathbf{P})$ is the feasible set of $(\widehat{\mathbf{P}})$, and that each solution of $(\widehat{\mathbf{P}})$ is a solution of $(\mathbf{P})$ and vice versa). Notice also that $\widehat{\mathcal{L}}$ is linear in the variables (λ, μ), and therefore the problem $\max_{(\lambda,\mu)} \widehat{\mathcal{L}}(x^*; \lambda, \mu)$ possesses a solution if and only if $x^* \in \mathcal{F}(\mathbf{P})$ (and then the objective function $\widehat{\mathcal{L}}(x^*; \cdot, \cdot)$ is constant on $\mathbb{R}^m \times \mathbb{R}^p$ with value $f(x^*)$). □

Notation 5.3. For fixed (λ, μ) consider the function

$$\widehat{\mathcal{L}}\big|_{[\lambda,\mu]} : \mathbb{R}^n \longrightarrow \mathbb{R}, \qquad \widehat{\mathcal{L}}\big|_{[\lambda,\mu]}(x) := \mathcal{L}(x; \lambda, \mu) \,.$$

Moreover, for fixed (λ, μ) put

$$X^*(\lambda, \mu) := \Big(\operatorname*{argmin}_{x \in \mathbb{R}^n} \widehat{\mathcal{L}}\big|_{[\lambda,\mu]}(x) \Big) \cap \mathcal{F}(\mathbf{P})$$

(where $X^*(\lambda, \mu) = \varnothing$ may well occur).

The combination of this notation with Proposition 5.2(a) and (b) leads to the following result.

Proposition 5.4. *The following two assertions (a) and (b) are equivalent.*

(a) *The triple* $(x^*, \lambda^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ *is a saddle point of* $\widehat{\mathcal{L}}$.
(b) $x^* \in X^*(\lambda^*, \mu^*)$

This proposition tells how saddle points of $\widehat{\mathcal{L}}$ can be calculated (at least, formally) in two steps:

Step 1: For each (fixed) $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p$ calculate the set

$$\operatorname*{argmin}_{x \in \mathbb{R}^n} \widehat{\mathcal{L}}\big|_{[\lambda,\mu]}(x)$$

using optimality conditions of unconstrained optimization.

Step 2: From Step 1 select those (λ, μ) with $X^*(\lambda, \mu) \neq \varnothing$.

Notice that Step 1 is nothing else than for all (λ, μ) finding the corresponding minimizers in x when calculating the value $\widehat{\theta}(\lambda, \mu)$ of the Lagrange-dual map as propagated in the previous section.

5.2. **Examples.**

Example 5.5 (Examples 3.10, 4.5 cont'd)**.** Again we consider Example 3.10 and the further elaborations in Example 4.5. We want to calculate all saddle points of $\widehat{\mathcal{L}}$. As mentioned above ("Step 1"), we first calculate the solution sets

$$\operatorname*{argmin}_{x \in \mathbb{R}^n} \widehat{\mathcal{L}}\big|_{[\lambda]}(x) \qquad \text{for all } \lambda \in \mathbb{R}.$$

As also mentioned above, for each $\lambda \in \mathbb{R}$ this amounts to finding the corresponding minimizers x such that $\widehat{\mathcal{L}}(x; \lambda) = \theta(\lambda)$. This has been done already, see eq. (4.6).

Hence, with $\mathcal{F}(\mathbf{P}) = [-1, +1]$ it is no difficulty to calculate the sets

$$X^*(\lambda) = \begin{cases} \varnothing & \text{if } \lambda \in (-\infty, 6), \\ \{-1, +1\} & \text{if } \lambda \in [6, +\infty). \end{cases}$$

By Proposition 5.4 we see that the set of saddle points of $\widehat{\mathcal{L}}$ is given by

$$\{\,(+1, \lambda) \mid \lambda \geq 6\,\} \,\cup\, \{\,(-1; \lambda) \mid \lambda \geq 6\,\}\ .$$

Finally, by Proposition 5.2(a) and (c) we (once more) get that the points $x^{*1} = -1$ and $x^{*2} = +1$ are solutions of the original problem $(\mathbf{P})$.

Example 5.6 (Exs. 3.12, 4.6 cont'd)**.** Let us again consider problem $(\mathbf{P})$ from Example 3.12 and Example 4.6. We consider the equivalent problem $(\widehat{\mathbf{P}})$, and we try to find saddle points of the corresponding Lagrangian $\widehat{\mathcal{L}}$. According to the previous calculations (see (4.7)), and since $\mathcal{F}(\mathbf{P}) = \{\, x \in \mathbb{R}^2 :\ \|x\|_2 \geq 2\,\}$, it is no difficulty to come up with

$$X^*(\lambda) = \begin{cases} \varnothing & \text{if } \lambda \in (-\infty, 1), \\ \widetilde{X} & \text{if } \lambda \in [1, +\infty) \end{cases} \tag{5.1}$$

where $\widetilde{X} := \{(2,0)^T, (0,2)^T, (-2,0)^T, (0,-2)^T\} = \{x^{*\ell} \mid \ell = 1,2,3,4\}$ with the notation in (3.19). Hence, according to Proposition 5.4 the set of saddle points of $\widehat{\mathcal{L}}$ is given by

$$\{(x^*, \lambda^*) \mid x^* \in \widetilde{X},\ \lambda^* \geq 1\}.$$

Finally, by Proposition 5.2(a) and (c) we (once more) get that the points $x^{*\ell}$, $\ell = 1,2,3,4$, are solutions of the original problem $(\mathbf{P})$.

Conclusions

In all three approaches (l_1-penalty approach, Lagrange duality, saddle point conditions) optimality conditions of unconstrained optimization based on the directed subdifferential can be used for the treatment of the constrained problem. This is possible, at least, for academic problems where the directed subdifferential can be analytically calculated. The power of the approaches lies in the strength of the calculus of directed subdifferentials (i.e., of directed sets). In principle, neither non-convexity nor non-differentiability of the objective function or of the constraint functions cause theoretical difficulties. From this point of view, in practice the discussed l_1-penalty approach is attractive because of its useful exactness properties. The consideration of Lagrange duality and saddle point optimality conditions is applicable, at least, if the number of constraints in the original problem $(\mathbf{P})$ is small. Here the straight calculus of directed subdifferentials also allows the circumvention of sign constraints for multipliers via the treatment of a max-term. This methodology may be favorable in theoretical calculations.

Appendix A. Calculations for Example 3.12

Consider the auxiliary function of problem $(\mathbf{P})$ in Example 3.12,

$$f_{1,\rho}\colon \mathbb{R}^2 \longrightarrow \mathbb{R}, \qquad f_{1,\rho}(x) = \|x\|_1 - 2 + \rho \max\{0, 2 - \|x\|_2\}$$
$$= |x_1| + |x_2| - 2 + \rho \max\{0, 2 - \sqrt{x_1^2 + x_2^2}\}\ .$$

In this appendix we summarize the calculation of the directed subdifferential of $f_{1,\rho}$ for arbitrary $\rho > 0$ and at each point $x \in X_0$ where

$$X_0 := \{ x \in \mathbb{R}^2 \mid x_1 \geq 0,\ x_2 > 0 \} \cup \{(0,0)^T\} .$$

Moreover, we discuss whether necessary and sufficient optimality conditions are satisfied (based on the directed subdifferential).

We start by partitioning of X_0 into the seven subsets

$$\begin{aligned} X_1 &:= \{ x \in \mathbb{R}^2 \mid x_1 > 0, x_2 > 0, \|x\|_2 < 2 \}, \\ X_2 &:= \{ x \in \mathbb{R}^2 \mid x_1 > 0, x_2 > 0, \|x\|_2 = 2 \}, \\ X_3 &:= \{ x \in \mathbb{R}^2 \mid x_1 > 0, x_2 > 0, \|x\|_2 > 2 \}, \\ X_4 &:= \{ (0,0)^T \}, \\ X_5 &:= \{ x \in \mathbb{R}^2 \mid x_1 = 0, 0 < x_2 < 2 \}, \\ X_6 &:= \{ (0,2)^T \}, \\ X_7 &:= \{ x \in \mathbb{R}^2 \mid x_1 = 0, x_2 > 2 \}, \end{aligned}$$

and we consider x from each of these sets, respectively. Moreover, for each $\rho > 0$ and each $i \in \{1, \ldots, 7\}$ we set

$$\begin{aligned} X_{i,\rho}^{\text{nec}} &:= \{ x \in X_i \mid 0 \in P_2(\overrightarrow{\partial} f_{1,\rho}(x)) \} \qquad \text{and} \\ X_{i,\rho}^{\text{suff}} &:= \{ x \in X_i \mid 0 \in \operatorname{int}(P_2(\overrightarrow{\partial} f_{1,\rho}(x))) \} \end{aligned}$$

denoting all points in X_i which satisfy the necessary or, respectively, sufficient optimality conditions for a local minimizer of $f_{1,\rho}$ (cf. the Propositions 3.8 and 3.9).

Case 1: $x \in X_1 = \{ x \in \mathbb{R}^2 \mid x_1 > 0, x_2 > 0, \|x\|_2 < 2 \}$

The function $f_{1,\rho}$ is differentiable at x, because $\max\{0, 2 - \|\tilde{x}\|_2\} = 2 - \|\tilde{x}\|_2$ and $\tilde{x} \neq 0$ for all $\tilde{x} \approx x$. Therefore

$$\begin{aligned} \overrightarrow{\partial} f_{1,\rho}(x) &= J_2\left(\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} \right) + \rho \cdot J_2\left(\left\{ -\frac{1}{2\|x\|_2} \cdot 2x \right\} \right) \\ &= J_2\left(\left\{ \begin{pmatrix} 1 - \rho \frac{x_1}{\|x\|_2} \\ 1 - \rho \frac{x_2}{\|x\|_2} \end{pmatrix} \right\} \right) . \end{aligned}$$

Now we calculate the sets X_1^{nec} and X_1^{suff}. Clearly,

$$P_2(\overrightarrow{\partial} f_{1,\rho}(x)) = \left\{ \begin{pmatrix} 1 - \rho \frac{x_1}{\|x\|_2} \\ 1 - \rho \frac{x_2}{\|x\|_2} \end{pmatrix} \right\} .$$

Hence, $0 \in P_2(\overrightarrow{\partial} f_{1,\rho}(x))$ if and only if $\rho x_1 = \rho x_2 = \|x\|_2$. Straightforward calculations (taking into account that $x \in X_1$) show that this is the case if and only if $0 < x_1 = x_2 < \sqrt{2}$ and $\rho = \|x\|_2 / x_1 = \sqrt{2} x_1 / x_1 = \sqrt{2}$. Hence

$$X_{1,\rho}^{\text{nec}} = \begin{cases} \{(\delta, \delta)^T \mid 0 < \delta < \sqrt{2}\} & \text{if } \rho = \sqrt{2}, \\ \varnothing & \text{otherwise.} \end{cases}$$

Since $P_2(\overrightarrow{\partial} f_{1,\rho}(x))$ is a singleton, its interior is empty. Therefore

$$X_{1,\rho}^{\text{suff}} = \varnothing \qquad \text{for all } \rho > 0.$$

Case 2: $x \in X_2 = \{ x \in \mathbb{R}^2 \mid x_1 > 0,\, x_2 > 0,\, \|x\|_2 = 2 \}$
We use the formula for the maximum of two functions, and obtain

$$\begin{aligned} \overrightarrow{\partial} f_{1,\rho}(x) &= J_2\left(\left\{\begin{pmatrix}1\\1\end{pmatrix}\right\}\right) + \rho\cdot\max\left\{\overrightarrow{0}, J_2\left(\left\{-\frac{1}{2\|x\|_2}\cdot 2x\right\}\right)\right\} \\ &= J_2\left(\left\{\tbinom{1}{1}\right\}\right) + \rho\cdot J_2\left(\operatorname{co}\left\{\tbinom{0}{0}, -\tfrac{1}{2}x\right\}\right) \\ &= J_2\left(\tbinom{1}{1} + \rho\cdot\operatorname{co}\left\{\tbinom{0}{0}, -\tfrac{1}{2}x\right\}\right)\ . \end{aligned}$$

We see that $0 \in P_2(\overrightarrow{\partial} f_{1,\rho}(x))$ if and only if there exists $\lambda \in [0,1]$ such that

$$\tbinom{0}{0} = \tbinom{1}{1} + \rho\cdot\left((1-\lambda)\tbinom{0}{0} + \lambda(-\tfrac{1}{2}x)\right)\ .$$

This identity is equivalent to the component-wise equations

$$\rho\lambda x_1 = 2 \qquad \text{and} \qquad \rho\lambda x_2 = 2\ . \tag{A.1}$$

We see that (necessarily) $x_1 = x_2$ and $\lambda > 0$. Hence, $\|x\|_2 = 2$ implies $x = (\sqrt{2}, \sqrt{2})^T$, and we arrive at $\rho\lambda\sqrt{2} = 2$. The condition $\lambda \le 1$ shows that $\rho \ge \sqrt{2}$. Vice versa (sufficiently), for each $\rho \ge \sqrt{2}$ the equations (A.1) are satisfied with $\lambda := \sqrt{2}/\rho$ and $x := (\sqrt{2}, \sqrt{2})^T$ where, additionally, $\lambda \in (0,1]$. Summarizing,

$$X_{2,\rho}^{\text{nec}} = \begin{cases} \varnothing & \text{if } \rho \in (0, \sqrt{2}) \\ \{(\sqrt{2}, \sqrt{2})^T\} & \text{if } \rho \ge \sqrt{2}. \end{cases}$$

For all $\rho > 0$ and all $x \in X_2$ the interior of the set $\tbinom{1}{1} + \rho\cdot\operatorname{co}\left\{\tbinom{0}{0}, -\tfrac{1}{2}x\right\}$ is empty. Hence,

$$X_{2,\rho}^{\text{suff}} = \varnothing \qquad \text{for all } \rho > 0.$$

Case 3: $x \in X_3 = \{ x \in \mathbb{R}^2 \mid x_1 > 0,\, x_2 > 0,\, \|x\|_2 > 2 \}$
Again the function $f_{1,\rho}$ is differentiable at x where the max term reduces to zero. Therefore

$$\overrightarrow{\partial} f_{1,\rho}(x) = J_2\left(\left\{\tbinom{1}{1}\right\}\right) + \rho\cdot\overrightarrow{0} = J_2\left(\left\{\tbinom{1}{1}\right\}\right)\ .$$

Since $P_2(\overrightarrow{\partial} f_{1,\rho}(x)) = \{(1,1)^T\}$ we have

$$X_{3,\rho}^{\text{nec}} = X_{3,\rho}^{\text{suff}} = \varnothing \qquad \text{for all } \rho > 0.$$

Case 4: $x \in X_4 = \{ (0,0)^T \}$, i.e., $x = (0,0)^T$
The function $\|\,.\,\|_1 \colon y \longmapsto \|y\|_1$ is convex, and thus Remark 2.2 shows that $\overrightarrow{\partial}\|\,.\,\|_1(0,0) = J_2([-1,+1]^2)$. Analogously, $\overrightarrow{\partial}\|\,.\,\|_2(0,0) = J_2(B_1(0,0))$ (where $B_1(0,0) := \{s \in \mathbb{R}^2 : \|s\|_2 \le 1\}$). Since $2 - \|x\|_2 > 0$, we obtain

$$\overrightarrow{\partial} f_{1,\rho}(0,0) = J_2([-1,1]^2) + \rho\big(-J_2(B_1(0,0))\big) = J_2([-1,1]^2) - \rho J_2(B_1(0,0))\,.$$

This difference is closer analyzed in [5, Example 5.7] and in [4, Example 3.20]. In these papers the two equivalences

$$\begin{aligned} (0,0)^T \in P_2\Big(J_2([-1,1]^2) - \rho J_2(B_1(0,0))\Big) &\qquad\iff\qquad \rho \in (0,1]\ , \\ (0,0)^T \in \operatorname{int} P_2\Big(J_2([-1,1]^2) - \rho J_2(B_1(0,0))\Big) &\qquad\iff\qquad \rho \in (0,1) \end{aligned}$$

are illustrated. Consequently,

$$X_{4,\rho}^{\text{nec}} = \left\{ \begin{array}{ll} \{(0,0)^T\} & \text{if } \rho \in (0,1] \\ \varnothing & \text{if } \rho > 1 \end{array} \right\}, \quad \text{and} \quad X_{4,\rho}^{\text{suff}} = \begin{cases} \{(0,0)^T\} & \text{if } \rho \in (0,1) \\ \varnothing & \text{if } \rho \geq 1 \,. \end{cases}$$

Case 5: $x \in X_5 = \{ x \in \mathbb{R}^2 \mid x_1 = 0,\, 0 < x_2 < 2 \}$
Similarly to Case 4 before, $\overrightarrow{\partial} \| \,.\, \|_1(x) = J_2([-1,+1] \times \{1\})$. The max-term in $f_{1,\rho}$ is differentiable at x because $x_1 = 0$ and $x_2 < 2$ (and thus $\max\{0, 2 - \|\tilde{x}\|_2\} = 2 - \|\tilde{x}\|_2$ for all $\tilde{x} \approx x$). Hence,

$$\begin{aligned} \overrightarrow{\partial} f_{1,\rho}(x) &= J_2([-1,1] \times \{1\}) + \rho \cdot J_2\left(\left\{ -\frac{1}{2\|x\|_2} \cdot 2x \right\} \right) \\ &= J_2([-1,1] \times \{1\}) + \rho \cdot J_2\left(\left\{ \left(\begin{smallmatrix} 0 \\ -1 \end{smallmatrix}\right) \right\} \right) \\ &= J_2([-1,1] \times \{1 - \rho\}) \,. \end{aligned}$$

We see that $(0,0)^T \in ([-1,1] \times \{1 - \rho\})$ if and only if $\rho = 1$. We also see that the interior of this set is empty. All in all,

$$X_{5,\rho}^{\text{nec}} = \left\{ \begin{array}{ll} \{0\} \times (0,2) & \text{if } \rho = 1 \\ \varnothing & \text{otherwise} \end{array} \right\}, \quad \text{and} \quad X_{5,\rho}^{\text{suff}} = \varnothing \text{ for all } \rho > 0.$$

Case 6: $x \in X_6 = \{ (0,2)^T \}$, i.e., $x = (0,2)^T$
In this situation, the mapping $\| \,.\, \|_1$ as well as the mapping $(y \longmapsto \max\{0, 2 - \|y\|_2\})$ are nonsmooth at $x = (0,2)^T$. Analogously to above we get

$$\begin{aligned} \overrightarrow{\partial} f_{1,\rho}(x) &= J_2([-1,1] \times \{1\}) + \rho \cdot \max\left\{ \overrightarrow{0}, J_2\left(\left\{ -\frac{1}{2\|x\|_2} \cdot 2x \right\} \right) \right\} \\ &= J_2([-1,1] \times \{1\}) + \rho \cdot \max\left\{ J_2(\{0\}), J_2\left(\left\{ -\frac{1}{x_2} \begin{pmatrix} 0 \\ x_2 \end{pmatrix} \right\} \right) \right\} \\ &= J_2([-1,1] \times \{1\}) + \rho \cdot \max\left\{ J_2(\{0\}), J_2\left(\left\{ \left(\begin{smallmatrix} 0 \\ -1 \end{smallmatrix}\right) \right\} \right) \right\} \\ &= J_2([-1,1] \times \{1\}) + \rho \cdot J_2(\{0\} \times [-1,0]) \\ &= J_2([-1,1] \times [1 - \rho, 1]) \,. \end{aligned}$$

Clearly, $(0,0)^T \in ([-1,1] \times [1-\rho,1])$ if and only if $\rho \geq 1$, and $(0,0)^T \in \operatorname{int}([-1,1] \times [1-\rho,1])$ if and only if $\rho > 1$. Hence,

$$X_{6,\rho}^{\text{nec}} = \left\{ \begin{array}{ll} \varnothing & \text{if } \rho \in (0,1) \\ \{(0,2)^T\} & \text{if } \rho \geq 1 \end{array} \right\}, \quad \text{and} \quad X_{6,\rho}^{\text{suff}} = \begin{cases} \varnothing & \text{if } \rho \in (0,1] \\ \{(0,2)^T\} & \text{if } \rho > 1 \,. \end{cases}$$

Case 7: $x \in X_7 = \{ x \in \mathbb{R}^2 \mid x_1 = 0,\, x_2 > 2 \}$
Again, the max-term reduces to zero in a neighbourhood of x, and hence

$$\overrightarrow{\partial} f_{1,\rho}(x) = J_2([-1,1] \times \{1\}) + \rho \cdot \overrightarrow{0} = J_2([-1,1] \times \{1\}) \,.$$

We see that $(0,0)^T \notin ([-1,1] \times \{1\})$, and therefore

$$X_{7,\rho}^{\text{nec}} = X_{7,\rho}^{\text{suff}} = \varnothing \qquad \text{for all } \rho > 0.$$

This finishes the consideration of all seven cases. Finally, from all cases we collect the points $x \in X_0$ satisfying the optimality conditions. By the above,

$$X_{0,\rho}^{\text{nec}} := \{ x \in X_0 \mid (0,0)^T \in P_2(\overrightarrow{\partial} f_{1,\rho}(x)) \} = \bigcup_{i=1}^{7} X_{i,\rho}^{\text{nec}}$$

$$= \begin{cases} \{(0,0)^T\} & \text{if } \rho \in (0,1), \\ \{(0,x_2)^T \mid 0 \le x_2 \le 2\} & \text{if } \rho = 1, \\ \{(0,2)^T\} & \text{if } \rho \in (1,\sqrt{2}), \\ \{(\delta,\delta)^T \mid 0 < \delta \le \sqrt{2}\} \cup \{(0,2)^T\} & \text{if } \rho = \sqrt{2}, \\ \{(\sqrt{2},\sqrt{2})^T\} \cup \{(0,2)^T\} & \text{if } \rho > \sqrt{2}, \end{cases} \tag{A.2}$$

and

$$X_{0,\rho}^{\text{suff}} := \{ x \in X_0 \mid (0,0)^T \in \text{int}(P_2(\overrightarrow{\partial} f_{1,\rho}(x))) \} = \bigcup_{i=1}^{7} X_{i,\rho}^{\text{suff}}$$

$$= \begin{cases} \{(0,0)^T\} & \text{if } \rho \in (0,1), \\ \varnothing & \text{if } \rho = 1, \\ \{(0,2)^T\} & \text{if } \rho > 1. \end{cases} \tag{A.3}$$

References

[1] A. Bagirov, N. Karmitsa and M. M. Mäkelä, *Introduction to Nonsmooth Optimization. Theory, Practice and Software*, Springer, Cham, 2014.

[2] R. Baier and E. Farkhi, Directed sets and differences of convex compact sets, in: *Systems Modelling and Optimization, Proceedings of the* 18th *IFIP TC7 Conference, Detroit, Michigan, 1997*, M. P. Polis, A. L. Dontchev, P. Kall, I. Lasiecka, and A. W. Olbrot (eds.), Chapman and Hall/CRC, Boca Raton–London–New York–Washington, D.C., 1999, pp. 135–143.

[3] R. Baier and E. Farkhi, Differences of convex compact sets in the space of directed sets. Part I: The space of directed sets, Set-Valued Anal. **9** (2001), 217–245.

[4] R. Baier and E. Farkhi, Differences of convex compact sets in the space of directed sets. Part II: Visualization of directed sets, Set-Valued Anal. **9** (2001), 247–272.

[5] R. Baier and E. Farkhi, The directed subdifferential of DC functions, in: Nonlinear Analysis and Optimization II: Optimization. A Conference in Celebration of Alex Ioffe's 70th and Simeon Reich's 60th Birthdays, 2008, Haifa, Israel, A. Leizarowitz, B. S. Mordukhovich, I. Shafrir, and A. J. Zaslavski (eds.), AMS Contemp. Math. **513**, . AMS and Bar-Ilan University, Providence, RI, USA, 2010, pp. 27–43.

[6] R. Baier, E. Farkhi and V. Roshchina, On computing the Mordukhovich subdifferential using directed sets in two dimensions, in: Variational Analysis and Generalized Differentiation in Optimization and Control, in: Honor of Boris S. Mordukhovich, R. S. Burachik and Jen-Chih Yao (eds.), Springer, New York–Dordrecht–Heidelberg–London, 2010, pp. 59–93.

[7] R. Baier, E. Farkhi and V. Roshchina, The directed and Rubinov subdifferentials of quasidifferentiable functions. Part I: Definition and examples, Nonlinear Anal. **75** (2012), 1074–1088.

[8] R. Baier, E. Farkhi and V. Roshchina, The directed and Rubinov subdifferentials of quasidifferentiable functions. Part II: Calculus, Nonlinear Anal. **75** (2012), 1058–1073.

[9] R. Baier, E. Farkhi and V. Roshchina, Directed subdifferentiable functions and the directed subdifferential without delta-convex structure, J. Optim. Theory Appl. **160** (2014), 391–414.

[10] R. Baier, E. Farkhi and V. Roshchina, From quasidifferentiable to directed subdifferentiable functions: Exact calculus rules, J. Optim. Theory Appl. **171** (2016), 384–401.

[11] M. S. Bazaraa, J. J. Goode and M. Z. Nashed, Optimality criteria in nonlinear programming without differentiability, Oper. Res. **19** (1971), 77–86.

[12] M. S. Bazaraa, J. J. Goode and M. Z. Nashed, On the cones of tangents with applications to mathematical programming, J. Optim. Theory Appl. **13** (1974), 389–426.

[13] M. S. Bazaraa, H. D. Sherali and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, Wiley, University of Michigan, MI, USA, 1993.

[14] A. Ben-Tal and J. Zowe, Necessary and sufficient optimality conditions for a class of nonsmooth minimization problems, Math. Program. **24** (1982), 70–91.

[15] J. F. Bonnans, J. C. Gilbert, C. Lemaréchal and C. Sagastizábal, *Numerical Optimization. Theoretical and Practical Aspects*, Universitext. Springer-Verlag, Berlin, 2003.

[16] J. V. Burke, Calmness and exact penalization, SIAM J. Control Optim. **29** (1991), 493–497.

[17] F. H. Clarke, Generalized gradients and applications, Trans. Amer. Math. Soc. **205** (1975), 247–262.

[18] F. H. Clarke, A new approach to Lagrange multipliers, Math. Oper. Res. **1** (1976), 165–174.

[19] F. H. Clarke, *Optimization and Nonsmooth Analysis*, SIAM, Philadelphia, PA, 1990. First edition published in John Wiley & Sons, Inc., New York, 1983.

[20] F. H. Clarke, Yu. S. Ledyaev, R. J. Stern and P. R. Wolenski, *Nonsmooth Analysis and Control Theory*, Springer-Verlag, New York, 1998.

[21] V. F. Demyanov, Saddle points of quasidifferentiable functions, Vestn. Leningr. Univ., Math. **15** (1983), 39–44.

[22] V. F. Demyanov and A. M. Rubinov, *Constructive Nonsmooth Analysis*, Approximation and Optimization **7**, Peter Lang, Frankfurt am Main–Berlin–Bern–New York–Paris–Wien, 1995. Russian original "*Foundations of Nonsmooth Analysis, and Quasidifferential Calculus*" published in Nauka, Moscow, 1990.

[23] G. di Pillo and L. Grippo, Exact penalty functions in constrained optimization, SIAM J. Control Optim. **27** (1989), 1333–1360.

[24] K. Eppler and B. Luderer, The Lagrange principle and quasidifferential calculus, Wiss. Z. Tech. Univ. Karl-Marx-Stadt **29** (1987), 187–192.

[25] Yan Gao, Optimality conditions with Lagrange multipliers for inequality constrained quasidifferentiable optimization, in: Quasidifferentiability and related topics, V. F. Demyanov and A. M. Rubinov (eds.) Nonconvex Optim. Appl. **43**, . Kluwer Acad. Publ., Dordrecht, 2000, pp. 151–162.

[26] H. Gfrerer and J. V. Outrata, *On Lipschitzian properties of implicit multifunctions*, SIAM J. Optim. **26** (2016), 2160–2189.

[27] O. Güler, *Foundations of Optimization*, Graduate Texts in Mathematics **258**, Springer, New York, 2010.

[28] H. Hadwiger, Minkowskische Addition und Subtraktion beliebiger Punktmengen und die Theoreme von Erhard Schmidt, Math. Z. **53** (1950), 210–218.

[29] J.-B. Hiriart-Urruty, Refinements of necessary optimality conditions in nondifferentiable programming. I, Appl. Math. Optim. **5** (1979), 63–82.

[30] P. L. Hörmander, Sur la fonction d'appui des ensembles convexes dans un espace localement convexe, Ark. Mat. **3** (1954), 181–186.

[31] A. D. Ioffe, Sous-différentielles approchées de fonctions numériques, C. R. Acad. Sci. Paris Sér. I Math. **292** (1981), 675–678.

[32] A. D. Ioffe, Calculus of Dini subdifferentials of functions and contingent coderivatives of set-valued maps, Nonlinear Anal. **8** (1984), 517–539.

[33] A. D. Ioffe, A Lagrange multiplier rule with small convex-valued subdifferentials for nonsmooth problems of mathematical programming involving equality and nonfunctional constraints, Math. Program., Ser. A **58** (1993), 137–145.

[34] V. Jeyakumar, Equivalence of saddle-points and optima, and duality for a class of nonsmooth nonconvex problems, J. Math. Anal. Appl. **130** (1988), 334–343.

[35] E. Kaucher, Interval Analysis in the Extended Interval Space $\mathbb{IR}$, Comput. Suppl. **2** (1980), 33–49.

[36] K. A. Khan, Relating lexicographic smoothness and directed subdifferentiability, Set-Valued Var. Anal. **Online First** (2016), 12 pages. DOI: 10.1007/s11228-016-0375-6.

[37] L. Kuntz and S. Scholtes, Constraint qualifications in quasidifferentiable optimization, Math. Program., Ser. A **60** (1993), 339–347.

[38] B. Luderer and R. Rösiger, On Shapiro's results in quasidifferential calculus, Math. Program., Ser. A **46** (1990), 403–407.

[39] P. Michel and J.-P. Penot, Calcul sous-différentiel pour des fonctions lipschitziennes et non lipschitziennes, C. R. Acad. Sci. Paris Sér. I Math. **298** (1984), 269–272.

[40] B. S. Mordukhovich, Maximum principle in the problem of time optimal response with nonsmooth constraints, Prikl. Mat. Mekh. **40** (1976), 1014–1023.

[41] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation. I. Basic Theory*, Springer-Verlag, Berlin, 2006.

[42] D. Pallaschke and R. Urbański, *Pairs of Compact Convex Sets. Fractional Arithmetic with Convex Sets*, Mathematics and Its Applications **548**, Kluwer Academic Publishers, Dordrecht, 2002.

[43] J.-P. Penot, *Calculus Without Derivatives*, Graduate Texts in Mathematics **266**, Springer-Verlag, New York–Heidelberg–Dordrecht–London, 2013.

[44] L. N. Polyakova, Sufficient conditions for a local extremum of a quasidifferentiable function in the presence of a quasidifferentiable constraint, Vestnik Leningrad. Univ. Mat. Mekh. Astronom. **1985** (1985), 26–30, 121.

[45] L. N. Polyakova, On the minimization of a quasidifferentiable function subject to equality-type quasidifferentiable constraints, in:Quasidifferential Calculus, V. F. Demyanov and L. C. W. Dixon (eds.), Math. Program. Study **29**, Springer, Berlin–Heidelberg, 1986, pp. 44–55. Translated from the Russian.

[46] L. N. Polyakova, *On the method of exact quasidifferentiable penalty functions*, Comput. Math. Math. Phys. **41** (2001), 205–218.

[47] L. S. Pontryagin, *Linear differential games. ii*, Sov. Math., Dokl. **8** (1967), 910–912.

[48] H. Rådström, *An embedding theorem for spaces of convex sets*, Proc. Amer. Math. Soc. **3** (1952), 165–169.

[49] R. T. Rockafellar, *Proximal subgradients, marginal values, and augmented Lagrangians in nonconvex optimization*, Math. Oper. Res. **6** (1981), 424–436.

[50] R. T. Rockafellar, *Lagrange multipliers and optimality*, SIAM Rev. **35** (1993), 183–238.

[51] R. T. Rockafellar, M. Wets and R. J. B. Wets, *Variational Analysis*, Springer, Berlin–Heidelberg, 2009.

[52] E. Rosenberg, *Exact penalty functions and stability in locally Lipschitz programming*, Math. Program. **30** (1984), 340–356.

[53] A. M. Rubinov and I. S. Akhundov, *Difference of compact sets in the sense of Demyanov and its application to non-smooth analysis*, Optimization **23** (1992), 179–188.

[54] A. Shapiro, *On optimality conditions in quasidifferentiable optimization*, SIAM J. Control Optim. **22** (1984), 610–617.

[55] A. Shapiro, *Quasidifferential calculus and first-order optimality conditions in nonsmooth optimization*, in: Quasidifferential Calculus, V. F. Demyanov and L. C. W. Dixon (eds.), Math. Program. Study **29**, Springer, Berlin–Heidelberg, 1986, pp. 56–68.

[56] D. E. Ward, *Differential stability in non-Lipschitzian optimization*, J. Optim. Theory Appl. **73** (1992), 101–120.

[57] Jane J. Ye, *Nondifferentiable multiplier rules for optimization and bilevel optimization problems*, SIAM J. Optim. **15** (2004), 252–274.

Manuscript received December 29 2016
revised April 9 2017

W. Achtziger
Department of Mathematics, Friedrich-Alexander University Erlangen-Nürnberg (FAU), 91058 Erlangen, Germany
E-mail address: achtziger@math.fau.de

R. Baier
Department of Mathematics, University of Bayreuth, 95440 Bayreuth, Germany
E-mail address: robert.baier@uni-bayreuth.de

E. Farkhi
School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel
E-mail address: elza@post.tau.ac.il

V. Roshchina
School of Science, RMIT University, Melbourne, VIC 3001, Australia
E-mail address: vera.roshchina@rmit.edu.au