SIAM J. OPTIM.
Vol. 29, No. 2, pp. 1473–1501

CHARACTERIZING AND TESTING SUBDIFFERENTIAL REGULARITY IN PIECEWISE SMOOTH OPTIMIZATION*

ANDREA WALTHER† AND ANDREAS GRIEWANK‡

Abstract. Functions defined by evaluation programs involving smooth elementals and absolute values as well as the max and min operators are piecewise smooth. They can be approximated by piecewise linear functions in abs-linear form. Using this second-order approximation, in [*Optim. Methods Softw.,* 31 (2016), pp. 904–930] we derived local first- and second-order minimality conditions for the underlying piecewise smooth functions. They are necessary and sufficient, respectively. These generalizations of the classical KKT and sufficient second-order theory assumed that the given abs-linear approximation satisfies the linear independence kink qualification (LIKQ). In this paper we relax LIKQ to the Mangasarian–Fromovitz kink qualification (MFKQ) and develop constructive conditions for several local convexity concepts. These are the existence of a supporting hyperplane at a given point for the function itself (SUP), the same for its local piecewise linearization (FOS), and the convexity of its local piecewise linearization on a neighborhood (FOC). As a consequence we show that first-order convexity in the sense of (FOC) is always required by subdifferential regularity (REG) as defined in Rockafellar and Wets [*Variational Analysis*, Springer, Cham, 1998], and is even equivalent to it under MFKQ. Whereas it was observed by Griewank and Walther that testing for local minimality (MIN) is polynomial under LIKQ, in this paper we show that, even under this strong linear independence kink qualification, testing for FOC and thus REG is co-NP-complete. We conjecture that this is also true for testing MFKQ itself.

Key words. subdifferential regularity, first-order-convexity, Clarke generalized gradient, Mordukhovich subdifferential, linear independence kink qualification, Mangasarin–Fromovitz kink qualification, abs-normal form

AMS subject classifications. 49J52, 90C56

DOI. 10.1137/17M115520X

1. Introduction and motivation.

We view this paper as part of an ongoing effort to make the concepts and results of the extensive literature on nonsmooth analysis accessible and implementable for computational practitioners. As in algorithmic or automatic differentiation [7], the key assumption facilitating this process is that the functions of interest are given by evaluation programs whose individual instructions can be easily analyzed and approximated. In the classical smooth case all of them are assumed to be differentiable near the evaluation points of interest. By allowing piecewise linear elemental functions such as abs, min, and max as part of the mix, we arrive at a large class of piecewise smooth functions, called abs-normal functions, that can still be analyzed by slight extensions of algorithmic differentiation tools. However, the resulting extended program does not produce *gradients*, *Jacobians*, or *Hessians*, but represents a procedure for evaluating the so-called abs-linearization $\Delta\varphi(x;\Delta x)$ as a local piecewise linear model of φ developed at x and evaluated at Δx. The construction of this approximation was first given in [5]. It simply amounts to replacing all smooth elemental instructions by their tangent, and all piecewise linear elementals by themselves. As shown in [5], we have uniform second-order contact, i.e.,

$$\varphi(x+\Delta x)-\varphi(x)=\Delta\varphi(x;\Delta x)+\mathcal{O}(\|\Delta x\|^2). \tag{1}$$

*Received by the editors November 2, 2017; accepted for publication (in revised form) March 7, 2019; published electronically May 28, 2019.
http://www.siam.org/journals/siopt/29-2/M115520.html

†Institut für Mathematik, Universität Paderborn, 33098 Paderborn, Germany (Andrea.Walther@uni-paderborn.de).
‡School of Mathematical Sciences and Information Technology, Yachay Tech, Hacienda San José, Urcuquí, 100119, Ecuador (griewank@yachaytech.edu.ec).

In contrast to directional differentiation, the order term in this generalized Taylor expansion is uniform, i.e., does not depend on the direction $\Delta x/\|\Delta x\|$. Since the discrepancy $\varphi(x+\Delta x) - \varphi(x) - \Delta\varphi(x;\Delta x)$ is of second order it possesses a derivative that vanishes at $\Delta x = 0$. However, for fixed x the discrepancy function is generally not differentiable with respect to Δx in a neighborhood of the origin. In other words we do not have strong Bouligand differentiability as discussed in [25].

Nevertheless, it is not surprising that quite a few local properties of $\varphi(x)$ near some x are inherited by $\Delta\varphi(x;\Delta x)$ near the origin $\Delta x = 0$. It is not very difficult to check that this is true in particular for the following properties:

- local minimality (MIN);
- local convexity (CON);
- existence of a supporting affine function (SUP).

Note that the relation between the two sides of (1), and hence the implications with respect to the above properties, is not symmetric. The right-hand side $\Delta\varphi(x;\Delta x)$ is by construction piecewise linear, whereas the left-hand side belongs to our class of nonsmooth functions. For example, this means that $\Delta\varphi(x;\Delta x)$ has a sharp minimum, i.e., locally $\Delta\varphi(x_*;\Delta x) \geq c\|x - x_*\|$, if and only if it has a strict minimum in that $\Delta\varphi(x_*;\Delta x) > 0$ for small $\Delta x \neq 0$. Of course this equivalence does not hold for the nonlinear version. In the companion paper [10], we explore these optimality conditions and the resulting rates of convergence for successive abs-linearization methods. Since $\Delta\varphi(x;\Delta x)$ is a first-order approximation of $\varphi(x)$, we will refer to its properties as first-order minimality (FOM), first-order convexity (FOC), and first-order supportability or stability (FOS) of the underlying function φ. These are necessary conditions for φ itself to have the corresponding properties MIN, CON, SUP. Conversely, some properties of $\varphi(x)$ can be deduced from those of its linearization $\Delta\varphi(x;\Delta x)$ at the origin, provided certain additional assumptions are satisfied.

In the abs-linear form, $\Delta\varphi(x;\Delta x)$ is represented by several real matrices and vectors, which can be analyzed by various linear algebra procedures. For an objective function $\varphi(x)$, it was shown in [8] how this information can be used to characterize local optimality, i.e., property MIN in terms of generalized KKT and positive curvature conditions. This analysis was based on a generalization of the linear independence constraint qualification (LICQ) called the linear independence kink qualification (LIKQ), which is generic [6], i.e., always satisfiable by arbitrary small perturbations of a given abs-normal form. Some of these perturbations may alter the inherent structural properties of a given problem, which is why we wish to relax LIKQ, just like LICQ in the smooth, constrained case. A popular relaxation is the Mangasarian–Formovitz constraint qualification (MFCQ), which generalizes to the Mangasarian–Fromovitz kink qualification (MFKQ) as defined later, in Definition 2.12.

To maintain equalities instead of inclusions in generalized differentiation one needs subdifferential regularity as defined, for example, in [1, Def. 3.5]. This well-known nonsmoothness property is a necessary condition for partial smoothness [17], which yields in turn a special case of the $\mathcal{VU}$ decomposition [20]. Our main thrust is to characterize subdifferential regularity (REG) for our class of nonsmooth functions. Therefore, one main contribution of this paper is the proof that REG at a point x always requires FOC. For this reason, we also refer to REG as a *convexity property*, although strictly speaking it does not require proper convexity. This can be seen from the simple example $\varphi(x) = |\sin(x)|$, which is not CON but FOC, even at $x = 0$, where $\Delta\varphi(0;\Delta x) = |\Delta x|$ as detailed later, in Example 2.6. An immediate consequence of FOC is first-order support, i.e., the existence of a supporting hyperplane g of $\Delta\varphi(x;\Delta x)$ at $\Delta x = 0$ such that the shifted function $(\Delta\varphi(x;\Delta x) - g^\top \Delta x)$ has 0 as a local minimizer. From the Taylor expansion above we see immediately that this is equivalent to g being a regular subgradient of φ at x as defined in [24]. Regular subgradients are also called Fréchet subdifferentials, which were extensively studied in [16]. However, in that contribution there are no specific results for piecewise smooth functions. We find here using [8] that, under LIKQ, Fréchet subdifferentiability can be verified in polynomial time.

The existence of a supporting hyperplane at a given point can also be interpreted as multiphase stability in the following sense.

LEMMA 1.1. *A continuous function $\varphi : \mathbb{R}^n \mapsto \mathbb{R}$ possesses a regular subgradient g at $x \in \mathbb{R}^n$ if and only if for $\mu_j \in \mathbb{R}$, $\mu_j \geq 0$, and $x_j \in \mathbb{R}^n$, $j \in \{1, \ldots, n\}$, the problem*

$$\min \sum_{j=0}^{n} \mu_j \varphi(x_j) \quad s.t. \quad \sum_{j=0}^{n} \mu_j x_j = x \;\; and \;\; \sum_{j=0}^{n} \mu_j = 1 \tag{2}$$

has the local minimizers $x_j = x$ for all j.

Proof. It is easy to see that the above minimization problem is, for any $\rho > 0$, weakly dual (as introduced in [3, Prop. 3.1]) to the maximization problem

$$\mathrm{conv}_\rho(\varphi)(x) \equiv \max g^\top x \quad \text{s.t.} \quad g^\top \tilde{x} \leq \varphi(\tilde{x}) \quad \text{for} \quad \tilde{x} \in B_\rho(x),$$

where $B_\rho(x)$ denotes the ball with radius ρ centered at x. It is shown in [3, Prop. 4.1] that there is no duality gap, which proves the assertion. ◻

The physical interpretation of this result is that x is a *feed vector* whose components represent the (molar) concentrations of various species in a mixture, e.g., hydrocarbons in a crude oil. Then, with $\varphi(x)$ denoting the Gibbs free energy, the mixture can split up into various submixtures called *phases* $\mu_j x_j$ in order to minimize the resulting mixed energy as a target function of (2). Locally this can only yield a reduction compared to the feed energy $\varphi(x)$ if there is no supporting hyperplane at x; see, e.g., [19, 22]. On the other hand the existence of such a hyperplane immediately implies the stability of the feed x as its own single phase.

The logical relations between the various properties defined above are given in the implication chain shown in Figure 1.

$$\text{CON} \Longrightarrow \text{REG} \;\underset{\substack{\Longleftarrow \\ (\text{MFKQ})}}{\Longrightarrow}\; \text{FOC} \Longrightarrow \text{FOS} \;\underset{\substack{\Longrightarrow \\ (\text{MFKQ}\wedge\text{POSC})}}{\Longleftarrow}\; \text{SUP} \Longleftarrow \text{MIN}$$

FIG. 1. *Relations between convexity properties of φ.*

The one-way implications at the beginning and at the end follow directly from the definitions. The relation of interest is the near equivalence between REG and FOC and the close connection between SUP of φ and that of its piecewise linearization $\Delta\varphi(x; \Delta x)$, namely FOS. Here, we need the MFKQ for the more difficult, converse implication. That is, we develop as the first main contribution of this paper a constructive condition for a local convexity concept, i.e., the convexity of the local piecewise linearization on a neighborhood. Based on this, we prove that FOC is always required by REG and is even equivalent to it under MFKQ. Whereas it was observed in [8] that testing for MIN is polynomial under LIKQ, the second main contribution of this paper is the proof that, even under this strong kink qualification, testing for FOC and thus

REG is co-NP-complete. We conjecture that this is also true for testing MFKQ itself. In contrast, testing for FOS under LIKQ is found here to be polynomial. The implication arrow from FOS to SUP, and thus MIN, represents a second-order sufficiency condition under MFKQ. It can be derived as a small modification of the condition given in [10, Thms. 4.1 and 4.3].

The paper is organized as follows. In section 2, we first introduce the representation of piecewise smooth functions in abs-normal form. Furthermore, we give five different example functions that will be used to illustrate the concepts and results throughout the paper. Then, we introduce the two kink qualifications LIKQ and (the weaker) MFKQ. In section 3 we first review some classical concepts of nonsmooth analysis and then prove the key result of this paper, namely the near equivalence between REG and FOC. Support conditions and thus Fréchet subdifferentiability are analyzed in section 4. Section 5 discusses the computational complexity of testing for convexity. The paper concludes with a summary and outlook in section 6.

2. Kink qualifications for nonsmooth problems. For the definition of kink qualifications, we consider the class of objective functions that are defined as finite compositions of smooth elemental functions and the absolute value function $\text{abs}(x) = |x|$ on an open domain. This also includes $\max(x, y)$, $\min(x, y)$, and the positive part function $\max(0, x)$, which can be reformulated in terms of an absolute value. The inclusion of the Euclidean norm as an elementary function would lead to objectives that are still Lipschitz continuous and lexicographically differentiable [23] but no longer piecewise smooth as defined in [25, Chap. 4].

The abs-normal form. To derive the abs-normal form for the class of piecewise smooth functions considered here, we define and number all arguments of absolute value evaluations successively as *switching variables* z_i for $i = 1, \ldots, s$. Without loss of generality we can number the switching variables in a topological order such that z_j can only influence z_i if $j < i$. Hence, we obtain the components of $z = z(x)$ one by one as piecewise smooth Lipschitz continuous functions of x. Then, we formulate the calculation of all switching variables as equality constraints. Furthermore, we introduce the vector of the absolute values of the switching variables as an extra argument of the then smooth target function f and the equality constraints F. Thus, we obtain a piecewise smooth representation of $y = \varphi(x)$ in the so-called abs-normal form

$$z = F(x, |z|), \tag{3}$$

$$y = f(x, |z|), \tag{4}$$

where for $\mathcal{D} \subset \mathbb{R}^n$ open $F : \overline{\mathcal{D}} \times \overline{\mathbb{R}^s_+} \mapsto \mathbb{R}^s$ and $f : \overline{\mathcal{D}} \times \overline{\mathbb{R}^s_+} \mapsto \mathbb{R}$ with $\overline{\mathcal{D}} \times \overline{\mathbb{R}^s_+} \subset \mathbb{R}^{n+s}$. Note that the vector z is implicitly defined by (3). Sometimes, we write

$$\varphi(x) \equiv f(x, |z(x)|)$$

to denote the objective directly in terms of the argument vector x only.

Example 2.1 (half pipe). The so-called half-pipe function given by

$$\varphi : \mathbb{R}^2 \mapsto \mathbb{R}, \qquad \varphi(x_1, x_2) = \max(x_2^2 - \max(x_1, 0), 0), \tag{5}$$

can be stated as

$$\varphi(x_1, x_2) = \tfrac{1}{2}\left(x_2^2 - \tfrac{1}{2}(x_1 + |x_1|) + \left|x_2^2 - \tfrac{1}{2}(x_1 + |x_1|)\right|\right) \tag{6}$$

using the so-called switching variables

$$z_1 = x_1 \equiv F_1(x, |z|) \qquad \text{and} \qquad z_2 = x_2^2 - \tfrac{1}{2}(x_1 + |z_1|) \equiv F_2(x, |z|). \tag{7}$$

Hence, it has the abs-normal form

$$z = F(x, |z|), \qquad y = \tfrac{1}{2}\left(x_2^2 - \tfrac{1}{2}(x_1 + |z_1|) + |z_2|\right) \equiv f(x, |z|).$$

In this paper, we are mostly interested in the case where the nonlinear elementals are all at least continuously differentiable, yielding the following function class.

DEFINITION 2.2. *For any $d \in \mathbb{N}$ and $\mathcal{D} \subset \mathbb{R}^n$, the set of functions $\varphi : \overline{\mathcal{D}} \mapsto \mathbb{R}$ defined by an abs-normal form* (3), (4) *with $f, F \in C^d(\overline{\mathcal{D}} \times \overline{\mathbb{R}^s_+})$ is denoted by $C^d_{\text{abs}}(\overline{\mathcal{D}})$. Since F and f are smooth in the respective arguments, the derivatives*

$$L \equiv \frac{\partial}{\partial |z|} F(x, |z|) \in \mathbb{R}^{s \times s}, \qquad Z \equiv \frac{\partial}{\partial x} F(x, |z|) \in \mathbb{R}^{s \times n}, \tag{8}$$

$$a \equiv \frac{\partial}{\partial x} f(x, |z|) \in \mathbb{R}^n, \quad \textit{and} \quad b \equiv \frac{\partial}{\partial |z|} f(x, |z|) \in \mathbb{R}^s \tag{9}$$

are well defined on $\overline{\mathcal{D}} \times \overline{\mathbb{R}^s_+}$ when interpreting $|z| \in \overline{\mathbb{R}^s_+}$ as a (nonnegative) variable vector.

Due to our assumption on the numbering of switching variables, the derivative matrix L is strictly lower triangular. Note that a function $\varphi \in C^d_{\text{abs}}(\overline{\mathcal{D}})$ may have various abs-normal forms, as shown below, for the example proposed by Hiriart-Urruty and Lemaréchal in [13, Chap. VIII, sect. 2.2]. The properties occurring in Figure 1 are independent of the particular representation, except for the kink qualifications LIKQ and MFKQ introduced below.

The combinatorial aspect of the evaluation can be expressed in terms of the signature vector $\sigma(x) \equiv \text{sgn}(z(x))$ and the corresponding diagonal matrix $\Sigma(x) = \text{diag}(\sigma(x))$. Throughout the paper, we will write $z = z(x)$, $\sigma = \sigma(x)$, and $\Sigma = \Sigma(x)$ for brevity if the dependence on the argument x is clear. However, we will also frequently consider the situation where σ varies over all possibilities $\{-1, 0, 1\}^s$. As observed already in [11] for the nonlinear case, the limiting gradients of φ in the vicinity of x as defined later in Definition 3.2 are given by

$$g_\sigma^\top \equiv a^\top + b^\top \Sigma (I - L\Sigma)^{-1} Z = a^\top + b^\top (\Sigma - L)^{-1} Z, \tag{10}$$

where the last equality only holds if $\sigma \in \{-1, 1\}^s$ so that Σ is nonsingular and thus its own inverse. The signature vectors define the domains

$$S_\sigma = \{x \in \mathbb{R}^n \mid \text{sgn}(z(x)) = \sigma\} \tag{11}$$

as a decomposition of the argument space, where one has

$$\varphi(x) = \varphi_\sigma(x) \qquad \text{for all } x \in S_\sigma \tag{12}$$

and φ_σ is one of finitely many differentiable selection functions in the sense of Scholtes [25]. At a given x, the nonsmoothness of the function φ is caused by the *active* switching variables $z_i(x) = 0$ for $1 \le i \le s$. We collect them in the active switch set

$$\alpha = \alpha(x) \equiv \{1 \le i \le s \mid \sigma_i(x) = \text{sgn}(z_i(x)) = 0\} \quad \text{of size} \quad |\alpha(x)| = s - |\sigma(x)|,$$

with $|\sigma| \equiv \|\sigma\|_1$ and $|\alpha|$ defined correspondingly. That is, $|\cdot|$ denotes a set's cardinality. We will distinguish two different scenarios for the activity pattern α.

Definition 2.3 (localization). *Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a C^d_{abs} function. If all switching variables vanish for a given point x, i.e.,*

$$z = z(x) = 0 \quad \text{and} \quad \alpha(x) = \{1, \ldots, s\},$$

we say that the switching and the function φ are localized at x. Otherwise, the switching and the function itself are nonlocalized.

Note that for each fixed $\sigma \in \{-1, 0, 1\}^s$ and corresponding $\Sigma = \operatorname{diag}(\sigma)$ the system

$$z = F(x, \Sigma z)$$

is $C^d(\overline{\mathcal{D}} \times \overline{\mathbb{R}^s_+})$ and has, by the implicit function theorem, a locally unique solution $z^\sigma = z^\sigma(x)$ with the well-defined Jacobian

$$\nabla z^\sigma \equiv \frac{\partial}{\partial x} z^\sigma = (I - L\Sigma)^{-1} Z \in \mathbb{R}^{s \times n}, \tag{13}$$

where Z and L are evaluated at $(x, z^\sigma(x))$. These derivative matrices can also be used to generate a local piecewise linear model, as follows.

Definition 2.4. *For $\varphi \in C^d_{\text{abs}}(\overline{\mathcal{D}})$, the abs-linear form $\Delta\varphi : \mathbb{R}^n \mapsto \mathbb{R}$, $\Delta y = \Delta\varphi(x; \Delta x)$, developed at x and evaluated for Δx is defined by*

$$\begin{bmatrix} z \\ \Delta y \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} Z & L \\ a & b \end{bmatrix} \begin{bmatrix} \Delta x \\ \Sigma z \end{bmatrix}$$

with $c_1 \in \mathbb{R}^s$, $c_2 \in \mathbb{R}$, $\sigma = \sigma(\Delta x) \equiv \operatorname{sign}(z(\Delta x)) \in \{-1, 0, 1\}^s$ now depending on Δx and $\Sigma \equiv \operatorname{diag}(\sigma)$ such that $\Delta\varphi(x; 0) = \varphi(x)$.

Example problems. To illustrate the kink qualifications and the regularity results derived in this paper, we consider the following five examples.

Example 2.5 (HUL). Hiriart-Urruty and Lemaréchal highlighted in [13] the piecewise linear, convex function $\varphi : \mathbb{R}^2 \to \mathbb{R}$,

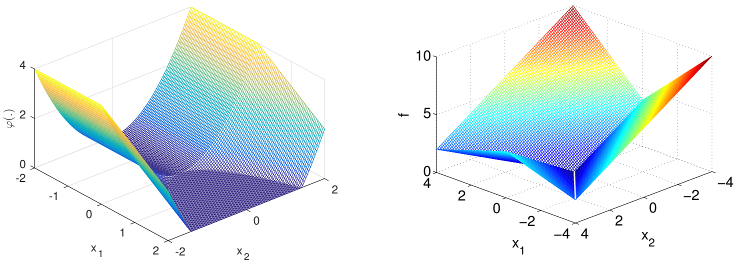
$$\varphi(x_1, x_2) = \max\{-100, 3x_1 - 2x_2, 2x_1 - 5x_2, 3x_1 + 2x_2, 2x_1 + 5x_2\}. \tag{14}$$

To derive an abs-normal form for this function one could either use the formulation

$$\begin{aligned} &\varphi(x) = \max\{\max\{\max\{\max\{y_0(x), y_1(x)\}, y_2(x)\}, y_3(x)\}, y_4(x)\} \\ \text{with} \quad &y_0(x) = -100,\ y_1(x) = 3x_1 - 2x_2,\ y_2(x) = 2x_1 - 5x_2, \\ &y_3(x) = 3x_1 + 2x_2,\ y_4(x) = 2x_1 + 5x_2, \end{aligned} \tag{15}$$

which yields, with a representation based on the absolute value, the four switching variables

$$\begin{aligned} z_1 &= -100 - 3x_1 + 2x_2, \\ z_2 &= -50 - \tfrac{1}{2}x_1 + 4x_2 + \tfrac{1}{2}|z_1|, \\ z_3 &= -25 - 1.25x_1 - 6x_2 + \tfrac{1}{4}|z_1| + \tfrac{1}{2}|z_2|, \\ z_4 &= -12.5 + \tfrac{35}{8}x_1 + \tfrac{9}{2}x_2 + \tfrac{1}{8}|z_1| + \tfrac{1}{4}|z_2| + \tfrac{1}{2}|z_3|, \end{aligned}$$



(a) $\varphi(x)$ *as defined in* (6).

(b) $\varphi(x)$ *as defined in* (18) *for* $\varepsilon = 0.5$.

FIG. 2. *Half-pipe example and gradient cube example for* $n = 2$.

or use the mathematically equivalent description

$$\varphi(x) = \max\{\max\{-100, 2x_1 + 5|x_2|\}, 3x_1 + 2|x_2|\}, \tag{16}$$

which requires only the three switching variables

$$z_1 = x_2, \quad z_2 = -100 - 2x_1 - 5|z_1|, \quad z_3 = -50 - 2x_1 + \tfrac{1}{2}|z_1| + \tfrac{1}{2}|z_2|.$$

As we will see later, the two representations have quite different properties.

Example 2.6 (abs-sin). As a simple nonconvex example in one dimension, we will employ

$$\varphi : \mathbb{R} \mapsto \mathbb{R}, \qquad \varphi(x) = |\sin(x)|. \tag{17}$$

Example 2.7 (half pipe). The half-pipe function $\varphi : \mathbb{R}^2 \mapsto \mathbb{R}$ defined in (5) is also nonconvex, as illustrated in Figure 2(a).[1]

Example 2.8 (gradient cube). Here, we consider the gradient cube example as introduced in [8] for $n = 2$. It is illustrated in Figure 2(b) and defined by

$$\varphi : \mathbb{R}^2 \mapsto \mathbb{R}, \qquad \varphi(x_1, x_2) = |x_2 - |x_1|| + \varepsilon|x_1|. \tag{18}$$

For piecewise linear functions, local convexity is of course neither sufficient nor necessary for local optimality. The second observation is borne out by an inverted lemon squeezer, which has a unique global minimum at the center but is of course not convex, as illustrated by the next example function.

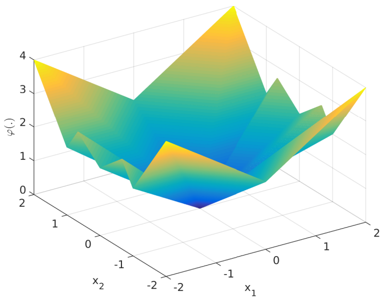
Example 2.9 (lemon squeezer). For $q \in \mathbb{N} \cup \{0\}$ and $\varepsilon \in \mathbb{R}$ given, we define the function

$$\varphi : \mathbb{R}^2 \mapsto \mathbb{R}, \qquad \varphi(x) = \sum_{i=0}^{q} (y_{2i}(x) + \varepsilon y_{2i+1}(x)) \quad \text{with} \tag{19}$$

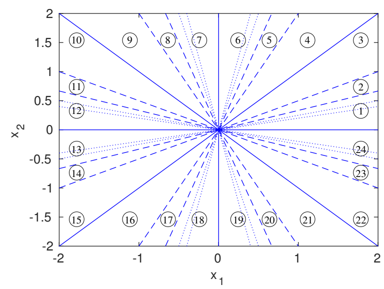
$$y_0(x) = |x_1| + |x_2|, \quad y_1(x) = |x_1 + x_2| + |x_1 - x_2|,$$
$$y_i(x) = |x_1 + i\,x_2| + |x_1 - i\,x_2| + |x_1 + x_2/i| + |x_1 - x_2/i|, \; i > 2,$$

which is illustrated in Figure 3(a) for $q = 1$. Hence, φ has $s = 8q+4$ switching variables yielding 2^{8q+4} definite signature vectors. The solid lines in Figure 3(b) represent the

[1]Color figures are available in the online version of this paper.



(a) $\varphi(\cdot)$ *for* $q = 1$.



(b) *Kinks in the argument space for* $q = 0, 1, 2$.

FIG. 3. *Function of* (19) *with* $\varepsilon = -0.5$.

kinks for $q = 0$, the dashed lines the additional kinks for $q = 1$, and the dotted lines the additional kinks for $q = 2$. The numbers $i \in \{1, \ldots, 24\}$ in the circles identify the corresponding selection function φ_i for later use. As can be seen, for $q = 1$ only 24 definite signature vectors out of the 4096 possibilities belong to nonempty subdomains of the argument space. For $q = 2$ the effect is even more pronounced with only 40 definite signature vectors out of 1 048 576 possibilities having nonempty subdomains in the argument space. Obviously, all $8q + 4 > 2 = n$ switching variables are active at the origin $0 \in \mathbb{R}^2$.

Kink qualifications. We will now examine under what conditions the sets S_σ as defined in (11) satisfy the classical constraint qualifications LICQ or MFCQ in some neighborhood of a given point $\mathring{x}$ with signature $\mathring{\sigma} \equiv \sigma(\mathring{x})$. By continuity of $z(x)$ it follows immediately that all nonvanishing components $\mathring{\sigma}_j \neq 0$ force the components σ_j of σ at points in the neighborhood to have the same sign. In other words, for some ball B_ρ about $\mathring{x}$ with radius $\rho > 0$ sufficiently small the intersection $B_\rho \cap S_\sigma$ can only be nonempty if

$$\sigma \succeq \mathring{\sigma} \quad \text{in the sense that} \quad \sigma_j \mathring{\sigma}_j \geq \mathring{\sigma}_j^2 \ \text{ for } \ j = 1, \ldots, s. \tag{20}$$

That is, $\sigma \succeq \mathring{\sigma}$ denotes $\sigma_j = \mathring{\sigma}_j$ if $\sigma_j \neq 0$ and no restriction on σ_j if $\sigma_j = 0$. This partial ordering of the signature vectors was used in [11]. As in the piecewise linear case we find that the closure $\bar{S}_\sigma$ of any S_σ is contained in the *extended closure*

$$\hat{S}_\sigma \equiv \{x \in \mathbb{R}^n : \sigma \succeq \sigma(x)\} \supset \bar{S}_\sigma.$$

Since $\preceq$ is a partial ordering we have the monotonicity property

$$\tilde{\sigma} \preceq \sigma \quad \Longrightarrow \quad \hat{S}_{\tilde{\sigma}} \subset \hat{S}_\sigma.$$

According to this monotonicity property, one has $\hat{S}_{\tilde{\sigma}} \subset \hat{S}_\sigma$ for a σ being definite, i.e., $0 \neq \sigma_i$ for all $i = 1, \ldots, s$, and $\tilde{\sigma} \preceq \sigma$. For this reason, from now on we can consider only maximal $\hat{S}_\sigma$, which are characterized by σ being a definite signature vector, for the examination of convexity. We will abbreviate this definiteness by $0 \notin \sigma$ and note that then $\Sigma = \Sigma^{-1}$ is an involutory matrix.

In particular, we have near $\mathring{x}$ the local decomposition property

$$\bar{B}_\rho = \bigcup_{0 \notin \sigma \succeq \mathring{\sigma}} \left(\hat{S}_\sigma \cap \bar{B}_\rho \right).$$

Using the smooth vector function z^σ as defined above we can for definite σ describe the $\hat{S}_\sigma$ in the usual representation of constrained sets as

$$\hat{S}_\sigma \equiv \{x \in \mathbb{R}^n : \sigma_i\, z_i^\sigma(x) \;\geq 0 \text{ for } i = 1, \dots, s\}.$$

As observed in [8, sect. 3.2] the point $\mathring{x}$ is a local minimizer of $\varphi(x)$ if and only if it is a local minimizer of each one of the *branch problems*

$$\min f_\sigma(x) \equiv f(x, \Sigma z^\sigma(x)) \quad \text{s.t.} \quad x \in \hat{S}_\sigma \quad \text{with} \quad 0 \notin \sigma \succeq \mathring{\sigma}. \tag{21}$$

It is natural to look at constraint qualifications for these problems, which will be useful for the regularity analysis in section 3. For any such definite σ the constraints that are active at $\mathring{x}$ have the same indices $i \in \mathring{\alpha} \equiv \alpha(\mathring{x})$, but the corresponding constraints $\sigma_i\, z_i^\sigma(x) \geq 0$ are not the same since σ is different. According to (13), the Jacobian of all constraints is given by

$$\Sigma \nabla z^\sigma = \Sigma(I - L\Sigma)^{-1} Z = (\Sigma - L)^{-1} Z \in \mathbb{R}^{s \times n},$$

where we have used the invertibility of $\Sigma = \Sigma^{-1}$ due to the definiteness of σ. One can show that the Jacobian of only the active constraints has a very similar structure.

LEMMA 2.10 (Jacobian of active constraints). *Consider the branch problem* (21) *for a given fixed definite signature vector* $\sigma \in \{-1, 1\}^s$ *with the corresponding signature matrix* Σ. *For* $\mathring{x} \in \mathbb{R}^n$ *with the signature vector* $\mathring{\sigma} \equiv \mathring{\sigma}(x)$ *and the resulting signature matrix* $\mathring{\Sigma}$, *the Jacobian of the constraints active at* $\mathring{x}$ *is given by*

$$J_\sigma \equiv (\sigma_i \nabla z_i^\sigma)_{i \in \mathring{\alpha}} = \check{\Sigma}(I - \check{L}\check{\Sigma})^{-1} \check{Z} = (\check{\Sigma} - \check{L})^{-1} \check{Z} \in \mathbb{R}^{|\mathring{\alpha}| \times n} \tag{22}$$

with matrices $\check{Z} \in \mathbb{R}^{|\mathring{\alpha}| \times n}$, $\check{\Sigma} \in \operatorname{diag}\{-1, 1\}^{|\mathring{\alpha}|}$ *diagonal, and* $\check{L} \in \mathbb{R}^{|\mathring{\alpha}| \times |\mathring{\alpha}|}$ *strictly lower triangular.*

Proof. Since $\sigma \succeq \mathring{\sigma}$, there exists a diagonal matrix Γ with $\operatorname{diag}(\Gamma) \in \{-1, 0, 1\}^s$ such that

$$\Sigma = \mathring{\Sigma} + \Gamma \qquad \text{and} \qquad \mathring{\Sigma}\,\Gamma = 0.$$

This yields

$$\Gamma \nabla z^\sigma = \Gamma(I - L\mathring{\Sigma} - L\Gamma)^{-1} Z = \Gamma[I - (I - L\mathring{\Sigma})^{-1} L\Gamma]^{-1} (I - L\mathring{\Sigma})^{-1} Z.$$

We define

$$\mathring{L} \equiv (I - L\mathring{\Sigma})^{-1} L, \qquad \mathring{Z} \equiv (I - L\mathring{\Sigma})^{-1} Z,$$

and $\mathring{P} \equiv |\Gamma|$. Here and throughout, $|\cdot|$ denotes that the absolute value of a matrix is taken componentwise to get a new matrix. Then, the matrix $\mathring{P}$ can be used to cancel out the inactive constraints and we obtain, with $\Gamma = \mathring{P}\Gamma = \Gamma\mathring{P}$ and $\mathring{P} = \mathring{P}\mathring{P}$, that

$$\Gamma \nabla z^\sigma = \Gamma\mathring{P}[I - \mathring{L}\Gamma]^{-1} \mathring{Z} = \Gamma\mathring{P}[I - \mathring{P}\mathring{L}\mathring{P}\Gamma]^{-1} \mathring{P}\mathring{Z} = \mathring{P}\Gamma[I - \mathring{P}\mathring{L}\mathring{P}\mathring{P}\Gamma]^{-1} \mathring{P}\mathring{Z},$$

where the penultimate identity follows from the Neumann series. Extracting the submatrices

$$\begin{aligned} &\check{Z} = (\mathring{P}\mathring{Z})_{i \in \mathring{\alpha}, 1 \leq j \leq n} \in \mathbb{R}^{|\mathring{\alpha}| \times n}, \quad \check{\Sigma} = (\mathring{P}\Gamma)_{i \in \mathring{\alpha}, j \in \mathring{\alpha}} \in \{-1, 1\}^{|\mathring{\alpha}| \times |\mathring{\alpha}|}, \quad \text{and} \\ &\check{L} = (\mathring{P}\mathring{L}\mathring{P})_{i \in \mathring{\alpha}, j \in \mathring{\alpha}} \in \mathbb{R}^{|\mathring{\alpha}| \times |\mathring{\alpha}|}, \end{aligned} \tag{23}$$

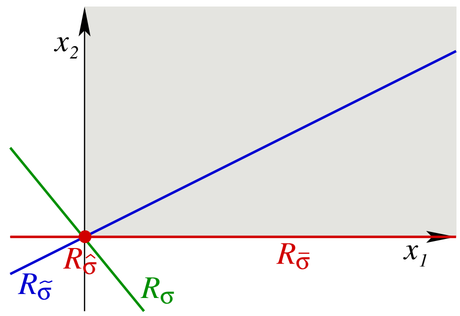


FIG. 4. *Four different scenarios for MFCQ with $n = 1$ and $|\mathring{\alpha}| = 2$.*

one arrives at the reduced identity

$$J_\sigma \equiv (\sigma_i \nabla z_i^\sigma)_{i\in\mathring{\alpha}} = \check{\Sigma}(I - \check{L}\check{\Sigma})^{-1}\check{Z} = (\check{\Sigma} - \check{L})^{-1}\check{Z} \in \mathbb{R}^{|\mathring{\alpha}|\times n}$$

for the Jacobian of the active constraints, as stated in the assertion. ∎

This relation was also derived in [8] for nonlocalized points by eliminating the z_i with $i \notin \mathring{\alpha}$ using the implicit function theorem. Since $|\det(\check{\Sigma} - \check{L})| = 1$, we obtain the following result.

COROLLARY 2.11 (uniformity of rank and null space). *The active Jacobian J_σ has, for all $\sigma \succeq \mathring{\sigma}$, the same rank $r \leq \min(|\mathring{\alpha}|, n)$ and the same null space as $\check{Z}$, which is spanned by some orthogonal matrix $\check{U} \in \mathbb{R}^{n\times(n-r)}$ such that $\check{Z}\check{U} = 0 \in \mathbb{R}^{|\mathring{\alpha}|\times(n-r)}$. All Jacobians J_σ have full rank $r = |\mathring{\alpha}| \leq n$ if and only if the $|\mathring{\alpha}| \times n$ matrix $\check{Z}$ has full rank $\mathring{\alpha} \leq n$. Hence, at $\mathring{x}$ either all branch problems satisfy LICQ or none of them do. Otherwise, if the columns of $\check{Z}$ are linearly independent such that $r = n < |\mathring{\alpha}|$, then the null space of J_σ contains only the null vector $0 \in \mathbb{R}^n$ for all $\sigma \succeq \mathring{\sigma}$.*

Due to this uniformity the constraint property LICQ is easy to check in polynomial time. In contrast, the Mangasarian–Fromovitz constraint qualification [18] for some $\sigma \succeq \mathring{\sigma}$ requires that

$$J_\sigma v = (\check{\Sigma} - \check{L})^{-1}\check{Z}v > 0 \tag{24}$$

has some solution $v \in \mathbb{R}^n$. There is also the possibility that $J_\sigma v \geq 0$ has only the trivial solution $v = 0$, in which case the branch problem is trivial, since S_σ is only a singleton, and can be excluded from further consideration. The latter possibility is not of much interest in the smooth case, but here it is quite likely to arise for certain signatures σ. Geometrically, this means that if the linear subspace

$$R_\sigma \equiv \{(\check{\Sigma} - \check{L})^{-1}\check{Z}v : v \in \mathbb{R}^n\}$$

intersects the positive orthant of $\mathbb{R}^{|\mathring{\alpha}|}$ in its interior, then we have MFCQ, and if it intersects only at the origin we have the trivial case. This is sketched in Figure 4, where MFCQ holds for the signature vectors σ and $\tilde{\sigma}$ but is violated for the signature vector $\bar{\sigma}$. Furthermore, $\hat{\sigma}$ represents the trivial case.

Dual formulation. By the usual duality relations for constraints of linear programs, MFCQ is violated for a particular $\sigma \succeq \mathring{\sigma}$ if and only if

$$\mu^\top J_\sigma = \mu^\top(\check{\Sigma} - \check{L})^{-1}\check{Z} = 0 \in \mathbb{R}^n \quad \text{for some} \quad 0 \neq \mu \geq 0 \in \mathbb{R}^{|\mathring{\alpha}|}, \tag{25}$$

since these are the constraints dual to the original MFCQ conditions

$$J_\sigma v = (\check{\Sigma} - \check{L})^{-1}\check{Z}v > 0 \in \mathbb{R}^{|\mathring{\alpha}|} \quad \text{for some} \quad 0 \neq v \in \mathbb{R}^n.$$

For any objective, the convex combination (25) implies that the Fritz John condition

$$\mu_0 g_\sigma(\mathring{x}) = \sum_{i \in \mathring{\alpha}} \mu_i \mathring{\sigma}_i \nabla z^\sigma(\mathring{x}) \quad \text{with} \quad \mu_0 = 0$$

holds, which is necessary but in no way sufficient for the optimality of φ_σ on the degenerate subdomain $\hat{S}_\sigma$.

MFCQ on an example. The following simple example shows that MFCQ may hold for one definite σ but not for another. Consider, for the switching variables of the half-pipe example given in (7), the localized case when $\mathring{x} = 0 \in \mathbb{R}^2$ with $n = 2$, $s = 2$, and

$$\check{Z} = Z = \begin{bmatrix} 1 & 0 \\ -\frac{1}{2} & 0 \end{bmatrix}, \quad \check{L} = L = \begin{bmatrix} 0 & 0 \\ -\frac{1}{2} & 0 \end{bmatrix}$$

yielding the active Jacobians

$$J_{(1,1)} = \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}, \quad J_{(1,-1)} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \quad J_{(-1,1)} = J_{(-1,-1)} = \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix}.$$

The corresponding range $R_{(1,1)} = \{(v_1, -v_1)^\top \mid (v_1, v_2)^\top \in \mathbb{R}^2\}$ intersects the positive orthant only at the origin but for all $(v_1, v_2)^\top \in \mathbb{R}^2$ with $v_1 = 0$ and $v_2 \in \mathbb{R}$. Hence, MFCQ is violated for the subdomain

$$\hat{S}_{(1,1)} = \{x \in R^2 \mid x_1 \geq 0 \text{ and } x_2^2 \geq x_1\}.$$

The vectors for the dual criterion are given by the left null vector $0 \neq \mu = (\tilde{\mu}, \tilde{\mu})^\top \in \mathbb{R}^2$, $\tilde{\mu} > 0$, confirming the violation of MFCQ. In contrast, the range $R_{(1,-1)} = \{(v_1, v_1)^\top \mid (v_1, v_2)^\top \in \mathbb{R}^2\}$ intersects the interior of the positive orthant for all $(v_1, v_2)^\top \in \mathbb{R}^2$ with $v_1 > 0$. Hence, MFCQ is satisfied on the subdomain

$$\hat{S}_{(1,-1)} = \{x \in \mathbb{R}^2 \mid x_1 \geq 0 \text{ and } x_2^2 \leq x_1\}.$$

For the remaining definite signatures, one obtains $\hat{S}_{(-1,1)} = \{x \in \mathbb{R}^2 \mid x_1 \leq 0\}$ and the degenerate polyhedron $\hat{S}_{(-1,-1)} = \{x \in \mathbb{R}^2 \mid x_1 \leq 0, x_2 = 0\}$ with

$$R_{(-1,1)} = R_{(-1,-1)} = \{(-v_1, 0) \mid v_1 \in \mathbb{R}\}.$$

It follows that they intersect the positive orthant only on its boundary for $v_1 \leq 0$. This can also be seen from the nonzero left null vectors $0 \neq \mu = (0, \mu_2)^\top \in \mathbb{R}^2$, $\mu_2 > 0$, confirming the violation of MFCQ. As stated in Corollary 2.11, one has

$$\operatorname{rank}(J_{(1,1)}) = \operatorname{rank}(J_{(1,-1)}) = \operatorname{rank}(J_{(-1,1)}) = \operatorname{rank}(J_{(-1,-1)}) = 1,$$

and all Jacobians have the same null space with the basis $u = (0, 1)^\top$.

If $\check{Z}$ has full row rank, and thus represents a surjective mapping from $\mathbb{R}^n$ onto $\mathbb{R}^{|\mathring{\alpha}|}$, then the criterion given by (24) is always satisfied. Other than that we do not know of any simple condition on $\check{Z}$, and possibly $\check{L}$, that would guarantee that all J_σ, and thus the corresponding $\hat{S}_\sigma$, satisfy MFCQ. For this reason, we define as follows an extension of the definition of LIKQ in [8].

Definition 2.12 (LIKQ and MFKQ). *For $\varphi \in C^1_{\text{abs}}(\overline{\mathcal{D}})$ according to Definition* 2.2 *and a given point $\mathring{x}$, consider the reduced quantities $\check{Z}$ and $\check{L}$ as defined in Lemma* 2.10 *and the Jacobian of the active constraints J_σ as defined in* (22)*. Then we say that LIKQ is satisfied at $\mathring{x}$ if $\check{Z} \in \mathbb{R}^{|\mathring{\alpha}|\times n}$ has full rank $|\mathring{\alpha}|$. More generally, we say that MFKQ holds if for all $\sigma \succeq \mathring{\sigma}$ as defined in* (20) *the vector inequality $J_\sigma v > 0$ is solvable for some $v \in \mathbb{R}^n$ or the problem is trivial in that $J_\sigma v \geq 0$ has only the solution $v = 0 \in \mathbb{R}^n$.*

Similar to the situation for smooth optimization, it follows easily that LIKQ implies MFKQ. In [8] we showed that the two nonsmooth versions of the chained Rosenbrock function suggested by Nesterov (according to [12]) satisfy LIKQ everywhere, and that LIKQ holds for their *natural* abs-normal representation, i.e., without any modification or preprocessing. That allows the complete characterization of the unique minimizer, excluding in particular the exponential number of stationary points that may entrap BFGS and other (generalized) gradient-based solvers. An optimization algorithm that makes this distinction constructively is currently under development.

While LIKQ just requires a rank determination for $\check{Z}$, we have so far not found a way to avoid the combinatorial effort of testing the weaker condition MFKQ for each branch problem defined by $\hat{\sigma} \succeq \mathring{\sigma}$. Indeed, we conjecture that MFKQ cannot be tested in polynomial time.

Lemma 2.13 (kink qualifications for example problems). *For the kink qualifications LIKQ and MFKQ, one has the following at $\mathring{x} = 0 \in \mathbb{R}$ and $\mathring{x} = 0 \in \mathbb{R}^2$, respectively.*

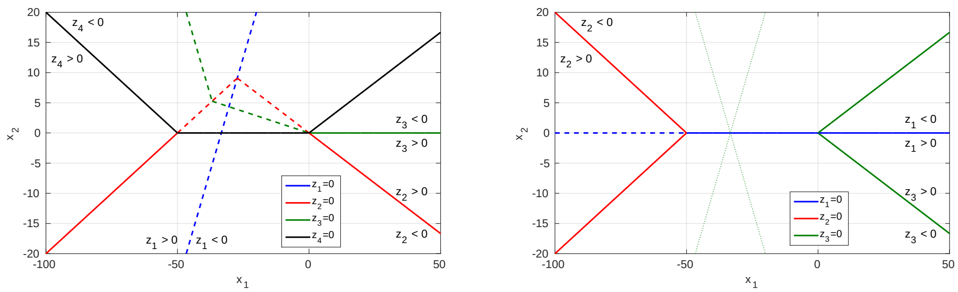
	HUL		Abs-sin	Half pipe	Gradient cube	Lemon squeezer
	(15)	(16)				
LIKQ	⦸	✓	✓	⦸	✓	⦸
MFKQ	✓	✓	✓	⦸	✓	✓

Proof. For the HUL example represented by (15), one has for $\mathring{x} = 0 \in \mathbb{R}^2$ that $s = 4$, $n = 2$, and $\mathring{\alpha} = \{2, 3, 4\}$. Therefore, LIKQ is obviously violated, since three out of four switching variables are active; see Figure 5(a). In the neighborhood of $\mathring{x}$, there are six polyhedra, i.e., less than 2^s, and they are all open. It also follows immediately from Figure 5(a) that MFKQ holds, since $\varphi(x)$ is already linear and therefore the linearizations of all sets S_σ have a nonempty interior. The dashed lines represent hidden kinks, where intermediate quantities undergo nonsmooth transitions, but the final function $\varphi(x)$ is completely smooth. Such hidden kinks might slow down the progress of an algorithm, but should not prevent it from functioning properly.

Using (16) as an alternative, one has at $\mathring{x} = 0 \in \mathbb{R}^2$ that $s = 3$, $\mathring{\alpha} = \{1, 3\}$,

$$Z = \begin{pmatrix} 0 & 1 \\ 2 & 0 \\ 2 & 0 \end{pmatrix}, \qquad \text{and} \qquad \check{Z} = \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix},$$

such that LIKQ, and hence also MFKQ, holds. One can check that all points where kinks intersect are indeed LIKQ points, since always only two of the three switching variables are active and their gradients are linearly independent; see Figure 5(b). Here, the dotted lines represent possibly active switches that occur in the theoretical formulation. However, they can never actually be active due to the contradicting requirements on x_1 and x_2.



(a) *Nondifferentiable points for* (15). (b) *Nondifferentiable points for* (16).

FIG. 5. *Kink structure for different formulations of* (14).

For the abs-sin example, one has $s = 1$, $z_1 = \sin(x)$,

$$\nabla z_\alpha^\sigma(0) = \nabla z^\sigma(0) = \cos(0) = 1, \quad L = 0, \quad a = 0, \quad b = 1,$$

such that LIKQ holds, which implies MFKQ also holds.

For the half-pipe example, at $\mathring{x} = 0$ the switching variables given in (7) yield

$$\check{Z} = Z = \begin{bmatrix} 1 & 0 \\ -\frac{1}{2} & 0 \end{bmatrix}, \ \check{L} = L = \begin{bmatrix} 0 & 0 \\ -\frac{1}{2} & 0 \end{bmatrix}, \ a = (-0.25 \ 0)^\top, \text{ and } b = (-0.25 \ 0.5)^\top.$$

This implies that we do not have LIKQ since Z does not have full row rank. Also, as already shown above, MFKQ is violated, since MFCQ cannot hold for $\sigma = (1, 1)$.

For the gradient cube example, (18) yields the switching variables $z_1 = x_1$ and $z_2 = x_2 - |z_1|$, so that

$$\check{Z} = Z = I, \ \check{L} = L = \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix}, \ a = (0 \ 0)^\top, \text{ and } b = (\varepsilon \ 1)^\top$$

for all $x \in \mathbb{R}^2$. Hence, LIKQ does hold at $\mathring{x}$, implying MFKQ also holds.

Finally, for the lemon squeezer example one has the $s = 8q+4$ switching variables

$$\begin{aligned} &z_1 = x_1, \ z_2 = x_2, \ z_3 = x_1 + x_2, \ z_4 = x_1 - x_2, \\ &z_{4i+1} = x_1 + i\,x_2, \ z_{4i+2} = x_1 - i\,x_2, \ z_{4i+3} = x_1 + x_2/i, \ z_{4i+4} = x_1 - x_2/i \end{aligned}$$

for $i = 1, \ldots, 2q$. Hence,

$$\begin{aligned} &\check{Z} = Z = \begin{pmatrix} 1 & 0 & 1 & 1 & \cdots & 1 & 1 & 1 & 1 & \cdots \\ 0 & 1 & 1 & -1 & \cdots & i & -i & 1/i & -1/i & \cdots \end{pmatrix}^\top \in \mathbb{R}^{(8q+4)\times 2}, \\ &a = 0, \quad \text{and} \quad b = (1, 1, \varepsilon, \varepsilon, 1, 1, 1, 1, \varepsilon, \varepsilon, \varepsilon, \varepsilon, \ldots), \end{aligned}$$

and obviously LIKQ cannot hold at $\mathring{x}$. Using $L = 0 \in \mathbb{R}^{s\times s}$, one has $J_\sigma = \Sigma^{-1}\check{Z} = \Sigma\check{Z}$. Since $\mathring{\sigma} = 0 \in \mathbb{R}^s$, MFKQ requires that either $J_\sigma v > 0$ is solvable for some $v \in \mathbb{R}^2$ or $J_\sigma v \geq 0$ has only the trivial solution for all $\Sigma \in \{-1, 1\}^s$. The strict inequality $J_\sigma v > 0$ yields the inequalities

$$\begin{aligned} &\sigma_1 v_1 > 0, \ \sigma_2 v_2 > 0, \ \sigma_3(v_1 + v_2) > 0, \ \sigma_4(v_1 - v_2) > 0, \\ &\sigma_{4i+1}(v_1 + iv_2) > 0, \ \sigma_{4i+2}(v_1 - iv_2) > 0, \\ &\sigma_{4i+3}(v_1 + v_2/i) > 0, \ \sigma_{4i+4}(v_1 - v_2/i) > 0 \qquad \text{for } i = 1, \ldots, 2q. \end{aligned}$$

Therefore, either one can find values of v_1 and v_2 for a given Σ such that all inequalities are fulfilled or there are contradicting strict inequalities yielding $v = 0 \in \mathbb{R}^s$ as the only solution of $J_\sigma v \geq 0$. It follows that MFKQ holds at $\mathring{x} = 0 \in \mathbb{R}^2$ for all $q \in \mathbb{N}$. $\square$

As can be seen from the HUL example, the representation of a function may have a considerable influence on the kink qualifications. To avoid LIKQ being violated, one should try to introduce as few kinks as possible. Furthermore, note that the lemon squeezer illustrates nicely the situation when MFKQ may hold but LIKQ does not.

3. Convexity conditions.

Smooth optimality conditions for local minima usually combine a stationarity condition with a convexity condition. Even in the unconstrained smooth but singular case, functions need not be convex near minimizers, e.g.,

$$\varphi(x_1, x_2) \equiv x_2^2 - 2x_2x_1^2 + \epsilon x_1^4 = (x_2 - x_1^2)^2 + (\epsilon - 1)x_1^4 \quad \text{for} \quad \epsilon > 1.$$

Taking the root $\sqrt{\varphi}$ to eliminate the singularity of the Hessian, one obtains a nonsmooth problem that is still nonconvex. In some applications, such as multiphase equilibria of mixed fluids, a lack of convexity may lead to the instability of single phase equilibria, as discussed in the introduction. Therefore, later in this section we will examine in more detail the conditions for convexity in the vicinity of a given point, irrespective of whether or not the point is even stationary. Such a verification of convexity is of interest not only for optimality but also, for example, for computer graphics. For a different class of piecewise-defined functions, such convexity tests were defined, for example, in [2, 4]. As for optimality (see [8, 10]), we can obtain necessary first-order conditions for convexity. We begin with a review of various established concepts for generalized derivatives and their relation for C^d_{abs} functions.

Some generalized derivatives of C^d_{abs} functions. One possibility is to define (regular) subdifferentials according to [21, 24].

DEFINITION 3.1 (Mordukhovich subgradients). *For a function $\varphi : \mathbb{R}^n \mapsto \mathbb{R}$ and a point $x \in \mathbb{R}^n$ the subderivative $d\varphi(x)(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}$ is defined as*

$$d\varphi(x)(w) = \liminf_{h \searrow 0, \bar{w} \to w} \frac{\varphi(x + h\bar{w}) - \varphi(x)}{h},$$

and the set of regular subgradients is given by

$$\widehat{\partial}^M \varphi(x) = \{g \in \mathbb{R}^n \mid \langle g, w \rangle \leq d\varphi(x)(w) \text{ for all } w \in \mathbb{R}^n\}.$$

This allows us to define the set of (general) subgradients as the outer semicontinuous envelope

(26)
$$\partial^M \varphi(x) = \{g \in \mathbb{R}^n \mid \exists \{x_k\}_{k \in \mathbb{N}} : x_k \to x,\ \varphi(x_k) \to \varphi(x),\ g_k \in \widehat{\partial}^M \varphi(x_k),\ g_k \to g\}.$$

The function $\varphi(\cdot)$ is called regular at x with $\partial^M \varphi(x) \neq \emptyset$ if $\varphi(\cdot)$ is locally lower semicontinuous at x and $\widehat{\partial}^M \varphi(x) = \partial^M \varphi(x)$.

Since we consider C^d_{abs} functions $\varphi(\cdot)$ throughout the whole paper, all $\varphi(\cdot)$ are lower semicontinuous and $\partial^M \varphi(x) \neq \emptyset$ holds everywhere. Hence, we only have to verify $\widehat{\partial}^M \varphi(x) = \partial^M \varphi(x)$ to show regularity of $\varphi(\cdot)$ in a given point x.

Another widely used derivative concept is based on limits of classical gradients. There, one exploits Rademacher's theorem, which guarantees that Lipschitz continuous functions such as the C^d_{abs} functions considered here are almost everywhere differentiable. Let $D_\varphi \subset \overline{\mathcal{D}}$ denote the set where the locally Lipschitz continuous function φ is differentiable in the classical sense, i.e., for each $x \in D_\varphi$ the classical gradient $\nabla\varphi(x)$ exists. Then one has the following.

Definition 3.2 (limiting gradients and Clarke subdifferential). *For a locally Lipschitz continuous function $\varphi : \overline{\mathcal{D}} \mapsto \mathbb{R}$ and a point $x \in \overline{\mathcal{D}}$, the set of limiting gradients is given by*

$$\partial^L\varphi(x) = \{g \in \mathbb{R}^n \mid \exists\{x_k\}_{k\in\mathbb{N}} : x_k \in D_\varphi,\ x_k \to x,\ \nabla\varphi(x_k) \to g\}.$$

This set is frequently also called Bouligand subdifferential. It forms the basis for the Clarke subdifferential defined by

$$\partial^C\varphi(x) = \operatorname{conv}\{\partial^L\varphi(x)\}.$$

Finally, for the C^d_{abs} functions considered here, one can define the following rather new derivative concept using the abs-linearization as introduced in (1).

Definition 3.3 (conical gradients). *For a C^d_{abs} function $\varphi : \overline{\mathcal{D}} \mapsto \mathbb{R}$ and a point $x \in \mathbb{R}^n$, the set of conical gradients is given by*

$$\partial^K\varphi(x) = \{g \in \mathbb{R}^n \mid g \in \partial^L_{\Delta x}\Delta\varphi(x;\Delta x)|_0\}.$$

These conical gradients and their generalization, conical Jacobians, are considered, for example, in [14, 15]. For the elements $g \in \partial^K\varphi(x)$, there must exist a signature vector $\sigma \in \{-1,0,1\}^s$ with $g = g_\sigma$ as defined in (10) such that the tangent cone of the coincidence set $\{x \in \mathbb{R}^n \mid \varphi(x) = \varphi_\sigma(x)\}$ at $\mathring{x}$ has a nonempty interior; see [5].

Example 3.4 (generalized derivatives for the abs-sin example). The function introduced in (17) is not differentiable in the classical sense at $\mathring{x} = 0 \in \mathbb{R}$. For the other derivative concepts, one obtains

$$\begin{aligned} d\varphi(0)(w) = |w| \quad &\Rightarrow \quad \widehat{\partial}^M\varphi(0) = [-1,1] = \partial^M\varphi(0),\\ \partial^L\varphi(0) = \{-1,1\} \quad &\Rightarrow \quad \partial^C\varphi(0) = [-1,1]. \end{aligned}$$

The abs-linearization of $\varphi(\cdot)$ at $\mathring{x} = 0$ is given by $\Delta\varphi(0;\Delta x) = |\Delta x|$, which is convex despite the fact that $\varphi(\cdot)$ itself is not convex at $\mathring{x} = 0$. It follows that

$$\partial^L\Delta\varphi(0;0) = \{-1,1\} = \partial^K\varphi(0).$$

As one can see, for this example, one obtains the inclusions

$$\partial^K\varphi(0) \subsetneq \widehat{\partial}^M\varphi(0) = \partial^M\varphi(0) \qquad \text{and} \qquad \partial^K\varphi(0) = \partial^L\varphi(0) \subsetneq \partial^C\varphi(0).$$

Example 3.5 (generalized derivatives for the half-pipe example). For the function given by (6), one can check that $\varphi(\cdot)$ is indeed differentiable in the classical sense at $\mathring{x} = 0 \in \mathbb{R}^2$ with $\nabla\varphi(0) = (0,0)$. Furthermore, one finds that

$$\begin{aligned} &\widehat{\partial}^M\varphi(0) = \{(0,0)\} \subsetneq \partial^M\varphi(0) = \{(0,0),(-1,0)\} = \partial^L\varphi(0)\\ \Rightarrow \quad &\partial^C\varphi(0) = \{(v,0) \mid v \in [-1,0]\}. \end{aligned}$$

Hence, in this case $\widehat{\partial}^M\varphi(0)$ is a proper subset of $\partial^M\varphi(0)$, such that $\varphi(\cdot)$ is not regular at $\mathring{x} = 0$. Since $\Delta\varphi(0;\Delta x) \equiv 0$, one has

$$\partial^K\varphi(0) = \partial^L\Delta\varphi(0;0) = \{(0,0)\}.$$

This yields the inclusions

$$\{\nabla\varphi(0)\} = \widehat{\partial}^M\varphi(0) \subsetneq \partial^M\varphi(0) = \partial^L\varphi(0) \qquad \text{and} \qquad \partial^K\varphi(0) \subsetneq \partial^C\varphi(0).$$

Note that the generalized derivatives may contain more elements than the classical gradient since $\{\nabla\varphi(0)\}$ is a proper subset of $\partial^M\varphi(0)$, $\partial^L\varphi(0)$, and $\partial^C\varphi(0)$.

Example 3.6 (generalized derivatives for the gradient cube example). The function of the gradient cube example is again not differentiable at $\mathring{x} = (0,0)^\top \in \mathbb{R}^2$. Possible candidates for a regular subgradient are given by the gradients of the selection functions $\varphi_i(\cdot)$, $1 \le i \le 4$, i.e.,

$$g_1 = (-1+\varepsilon, 1), \quad g_2 = (1-\varepsilon, 1), \quad g_3 = (-1-\varepsilon, -1), \quad \text{and} \quad g_4 = (1+\varepsilon, -1).$$

The property of g being a regular subgradient of $\varphi(\cdot)$ at the argument $\mathring{x}$ is equal to

$$\liminf_{x\to\mathring{x}, x\neq\mathring{x}} \frac{\varphi(x) - \varphi(\mathring{x}) - \langle g, x-\mathring{x}\rangle}{\|x-\mathring{x}\|} \ge 0. \tag{27}$$

One can now check that this inequality holds for g_1 and g_2 if $\varepsilon \ge 1$. For g_3 and g_4, the condition holds if $\varepsilon \ge -1$. This yields

$$\widehat{\partial}^M\varphi(0) = \operatorname{conv}\{g_1, g_2, g_3, g_4\} = \partial^M\varphi(0) \quad \text{if } \varepsilon \ge 1,$$

since $\varphi(\cdot)$ is a convex function for $\varepsilon \ge 1$, and then the second equality is given by [24, Prop. 8.12]. For $\varepsilon \in [-1, 1)$ one has to examine the situation more closely, since $\varphi(\cdot)$ is no longer convex. As can be seen from Figure 2, in this case the gradients of $\varphi_3(\cdot)$ and $\varphi_4(\cdot)$ define supporting hyperplanes for $\varphi(\cdot)$. A third supporting hyperplane is determined by the function values for $x_1 = x_2 > 0$ and $-x_1 = x_2 > 0$, the normal vector of which is given by $g_5 \equiv (0, \varepsilon)$. For this vector, one can again check that (27) holds if $\varepsilon \in [-1, 1]$. It follows that

$$\widehat{\partial}^M\varphi(0) = \operatorname{conv}\{g_3, g_4, g_5\},$$

whereas, with $\widehat{\partial}^M\varphi(0) \subset \partial^M\varphi(0)$ according to [24, Thm. 8.6], (26) yields

$$\partial^M\varphi(0) = \widehat{\partial}^M\varphi(0) \cup \operatorname{conv}\{g_1, g_4\} \cup \operatorname{conv}\{g_2, g_3\}.$$

Finally, for $\varepsilon < -1$, one has

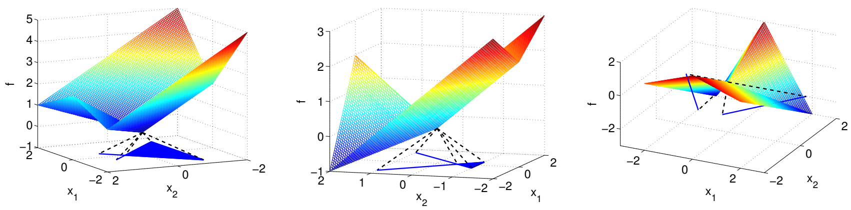
$$\widehat{\partial}^M\varphi(0) = \emptyset \qquad \text{and} \qquad \partial^M\varphi(0) = \operatorname{conv}\{g_1, g_4\} \cup \operatorname{conv}\{g_2, g_3\}.$$

Figure 6 illustrates the Mordukhovich subdifferentials for different values of ε as dark blue areas and lines. Since $\varphi(\cdot)$ is already piecewise linear, one has $\varphi(x) = \Delta\varphi(0; x)$ and

$$\partial^K\varphi(0) = \partial^L\varphi(0) = \{g_1, g_2, g_3, g_4\} \qquad \text{and} \qquad \partial^C\varphi(0) = \operatorname{conv}\{g_1, g_2, g_3, g_4\}$$

for all values of ε.

Example 3.7 (generalized derivatives for the lemon squeezer example). The corresponding $\varphi(\cdot)$ is not differentiable at $\mathring{x} = (0,0)^\top \in \mathbb{R}^2$. For notational simplicity, we



(a) $\partial^M\varphi(0)$ *for* $\varepsilon = 0.5$. (b) $\partial^M\varphi(0)$ *for* $\varepsilon = -0.5$. (c) $\partial^M\varphi(0)$ *for* $\varepsilon = -1.5$.

FIG. 6. *Mordukhovich subdifferentials for the gradient cube problem.*

only consider the case when $q = 1$ here. Possible candidates for a regular subgradient are given by the gradients of the linear selection functions $\varphi_i(\cdot)$, $1 \leq i \leq 24$, i.e.,

$$
\begin{array}{lll}
g_1 = (5+6\varepsilon, 1), & g_2 = (5+4\varepsilon, 1+6\varepsilon), & g_3 = (3+4\varepsilon, 5+6\varepsilon), \\
g_4 = (3+2\varepsilon, 5+8\varepsilon), & g_5 = (1+2\varepsilon, 6+8\varepsilon), & g_6 = (1, 6+26\varepsilon/3), \\
g_7 = (-1, 6+26\varepsilon/3), & g_8 = (-1-2\varepsilon, 6+8\varepsilon), & g_9 = (-3-2\varepsilon, 5+8\varepsilon), \\
g_{10} = (-3-4\varepsilon, 5+6\varepsilon), & g_{11} = (-5-4\varepsilon, 1+6\varepsilon), & g_{12} = (-5-6\varepsilon, 1), \\
g_{13} = (-5-6\varepsilon, -1), & g_{14} = (-5-4\varepsilon, -1-6\varepsilon), & g_{15} = (-3-4\varepsilon, -5-6\varepsilon), \\
g_{16} = (-3-2\varepsilon, -5-8\varepsilon), & g_{17} = (-1-2\varepsilon, -6-8\varepsilon), & g_{18} = (-1, -5-26\varepsilon/3), \\
g_{19} = (1, -6-26\varepsilon/3), & g_{20} = (1+2\varepsilon, -6-8\varepsilon), & g_{21} = (3+2\varepsilon, -5-8\varepsilon), \\
g_{22} = (3+4\varepsilon, -5-6\varepsilon), & g_{23} = (5+4\varepsilon, -1-6\varepsilon), & g_{24} = (5+6\varepsilon, -1).
\end{array}
$$

For $\varepsilon \geq 0$, $\varphi(\cdot)$ is convex, yielding

$$\widehat{\partial}^M\varphi(0) = \operatorname{conv}\{g_i,\ i = 1, \ldots, 24\} = \partial^M\varphi(0).$$

For $\varepsilon < 0$, it can be shown that the kinks between g_i and g_{i+1} yield a convex part if i is even and a concave part if i is odd; see Figure 3(b). Hence, for i even, i.e., in the convex case, all elements of the convex hull of g_i and g_{i+1} are regular subgradients, whereas when i is odd, i.e., in the concave case, the set of regular subgradients at this kink is empty. For $0 > \varepsilon > -9/13$, this yields

$$\partial^M\varphi(0) = \bigcup_{i\in\{2,4,\ldots,22\}} \operatorname{conv}\{g_i, g_{i+1}\} \cup \{g_{24}, g_1\}$$

and an even more complicated set $\widehat{\partial}^M\varphi(0)$, which is therefore not stated here. If $\varepsilon < -9/13$, the kinks defined by the selection functions $\varphi_6(\cdot)$ and $\varphi_7(\cdot)$ as well as $\varphi_{18}(\cdot)$ and $\varphi_{19}(\cdot)$ attain negative values. Hence, no supporting hyperplane exists at $\mathring{x} = 0$, yielding

$$\widehat{\partial}^M\varphi(0) = \emptyset.$$

Since $\varphi(\cdot)$ is itself already piecewise linear, it follows that

$$\partial^K\varphi(0) = \partial^L\varphi(0) = \{g_i,\ i = 1, \ldots, 24\} \quad \text{and} \quad \partial^C\varphi(0) = \operatorname{conv}\{g_i,\ i = 1, \ldots, 24\}$$

for all values of ε.

As illustrated by these small examples, the relations of the different concepts for generalized derivatives are by no means trivial. For this reason, we will now examine these relations more closely.

Relations between generalized derivatives. Exploiting the $C^d_{\rm abs}$ structure, one obtains the following results.

PROPOSITION 3.8 (limiting, Mordukhovich, and Clarke subdifferentials). *For the $C^d_{\rm abs}$ functions $\varphi : \overline{\mathcal{D}} \to \mathbb{R}$, $\overline{\mathcal{D}} \subset \mathbb{R}^n$, and $x \in \mathbb{R}^n$, the inclusions*

$$\emptyset \neq \partial^L \varphi(x) \subset \partial^M \varphi(x) \subset \partial^C \varphi(x)$$

hold. Furthermore, the function $\varphi(\cdot)$ is regular in $x \in \mathbb{R}^n$ if and only if

$$\partial^M \varphi(x) = \partial^C \varphi(x).$$

Proof. It follows from Definition 3.2 that for an element $g \in \partial^L \varphi(x)$ there exists a sequence $\{x_k\}_{k\in\mathbb{N}}$ such that $\varphi(\cdot)$ is differentiable at x_k and $\nabla\varphi(x_k) \to g$. According to [24, Ex. 8.8a], one then has $\{\nabla\varphi(x_k)\} = \widehat{\partial}^M \varphi(x_k)$, yielding $g \in \partial^M \varphi(x)$.

Now assume that $g \in \partial^M \varphi(x)$. If $\varphi(\cdot)$ is smooth in a neighborhood of x, then according to [24, Ex. 8.8b] one has

$$\{\nabla\varphi(x_k)\} = \partial^L \varphi(x) = \partial^M \varphi(x) = \partial^C \varphi(x).$$

If $\varphi(\cdot)$ is not smooth in a neighborhood of x, then at least one switching variable is active at x. To illustrate the situation, assume that only one switching variable vanishes, i.e., at x one has, for two selection functions as defined in (12), that $\varphi(x) = \varphi_{\sigma_1}(x) = \varphi_{\sigma_2}(x)$ with the two gradients $g_{\sigma_1}(x)$ and $g_{\sigma_2}(x)$. It follows from the definition of $\partial^M\varphi(x)$ that $g_{\sigma_1}(x), g_{\sigma_2}(x) \in \partial^M\varphi(x)$. Furthermore, if (27) also holds for the convex combinations of $g_{\sigma_1}(x)$ and $g_{\sigma_2}(x)$, then they are also contained in $\partial^M\varphi(x)$. That is, one of the following two cases occurs:

$$\partial^M\varphi(x) = \{g_{\sigma_1}(x), g_{\sigma_2}(x)\} \quad \text{or} \quad \partial^M\varphi(x) = \operatorname{conv}\{g_{\sigma_1}(x), g_{\sigma_2}(x)\};$$

see Examples 3.5 and 3.4, respectively, for an illustration. Therefore, it follows that

$$\partial^M\varphi(x) \subset \operatorname{conv}\{g_{\sigma_1}(x), g_{\sigma_2}(x)\} = \operatorname{conv}\left\{\partial^L\varphi(x)\right\} = \partial^C\varphi(x).$$

If more switching variables vanish at x, it follows similarly that the corresponding gradients $g_{\sigma_i}(x)$, $i = 1, \ldots, l$, of the l selection functions with $\varphi(x) = \varphi_{\sigma_i}(x)$ are contained in $\partial^M\varphi(x)$. Depending on (27), convex combinations of these gradients or of a proper subset of these gradients are also elements of $\partial^M\varphi(x)$; see Example 3.6 with $\varepsilon < 1$ for an illustration. This yields

$$\partial^M\varphi(x) \subset \operatorname{conv}\{g_{\sigma_i}(x) \mid \varphi(x) = \varphi_{\sigma_i}(x),\ i = 1, \ldots, l\} = \operatorname{conv}\left\{\partial^L\varphi(x)\right\} = \partial^C\varphi(x).$$

For proving the second assertion, first assume that $\varphi(\cdot)$ is regular in x. Then one has that $\widehat{\partial}^M\varphi(x) = \partial^M\varphi(x)$ is convex; see [24, Thm. 8.6]. Using the same argument as above, it follows for the selection functions with $\varphi(x) = \varphi_{\sigma_i}(x)$, $i = 1, \ldots, l$, that

$$\partial^M\varphi(x) = \operatorname{conv}\{g_{\sigma_i}(x) \mid \varphi(x) = \varphi_{\sigma_i}(x),\ i = 1, \ldots, l\} = \operatorname{conv}\left\{\partial^L\varphi(x)\right\} = \partial^C\varphi(x).$$

Now assume that $\partial^M\varphi(x) = \partial^C\varphi(x)$ holds. Then, $\partial^M\varphi(x)$ is a convex set. This can only be the case if

$$\partial^M\varphi(x) = \operatorname{conv}\{g_{\sigma_i}(x) \mid \varphi(x) = \varphi_{\sigma_i}(x),\ i = 1, \ldots, l\}.$$

Then, it follows from the $C^d_{\rm abs}$ structure of the functions considered here and the definition of the regular Mordukhovich subgradient that $\partial^M\varphi(x) = \widehat{\partial}^M\varphi(x)$ must hold. Hence, $\varphi(\cdot)$ is regular in x. ∎

PROPOSITION 3.9 (conical and limiting gradients). *Let $\varphi : \overline{\mathcal{D}} \to \mathbb{R}$, $\overline{\mathcal{D}} \subset \mathbb{R}^n$, be a C^d_{abs} function. Then, one has*

$$\partial^K \varphi(x) \subset \partial^L \varphi(x)$$

for all $x \in \mathbb{R}^n$. Furthermore, if MFKQ holds at $\mathring{x} \in \overline{\mathcal{D}}$, then

$$\partial^K \varphi(\mathring{x}) = \partial^L \varphi(\mathring{x}).$$

Proof. The first inclusion was shown in [5, Prop. 9]. Therefore, we only have to prove the second equality. Due to the piecewise smoothness of φ, every limiting gradient is the gradient of a selection function φ_σ, which coincides with φ on S_σ. Because of MFKQ, the tangent cone of S_σ coincides with that of its linearization and has a nonempty interior. Thus, the assertion follows from [5, Cor. 1], which states that the gradient of φ_σ must then be conical. ∎

Hence, since MFKQ holds at $\mathring{x} = 0$ for Examples 2.6 and 2.8, the equality of the conical and limiting gradients at $\mathring{x} = 0$ derived in Examples 3.4 and 3.6, respectively, is to be expected. On the other hand, MFKQ does not hold at $\mathring{x} = 0$ for Example 2.7. As illustrated in Example 3.5, in this case $\partial^K \varphi(x)$ is a proper subset of $\partial^L \varphi(x)$, which is possible since MFKQ is not satisfied.

Relations between different convexity properties. Based on the abs-linearization as defined in Definition 2.4, we introduce the concept of first-order convexity in the following way.

DEFINITION 3.10 (first-order convexity (FOC)). *The C^d_{abs} function $\varphi : \mathbb{R}^n \to \mathbb{R}$ is said to be* convex of first order *at a point $\mathring{x}$ if its abs-linearization $\Delta\varphi(\mathring{x}; \cdot) : \mathbb{R}^n \to \mathbb{R}$ is convex on some ball about the argument $\Delta x = 0$.*

Hence, a function $\varphi(\cdot)$ is called first-order convex in some ball about $\mathring{x}$ if its abs-linearization as a first-order model in $\mathring{x}$ is convex. For this new concept, one has the following.

THEOREM 3.11 (regularity and FOC). *The C^d_{abs} function $\varphi : \mathbb{R}^n \to \mathbb{R}$ is first-order convex in some ball about $\mathring{x}$ if $\varphi(\cdot)$ is regular at $\mathring{x}$. Furthermore, if MFKQ holds at $\mathring{x} \in \mathbb{R}^n$, then $\varphi(\cdot)$ is first-order convex in some ball about $\mathring{x}$ if and only if $\varphi(\cdot)$ is regular at $\mathring{x}$.*

Proof. Assume that $\varphi(\cdot)$ is regular at $\mathring{x}$. Then, it follows from Proposition 3.8 that

$$\widehat{\partial}^M \varphi(\mathring{x}) = \partial^M \varphi(\mathring{x}) = \partial^C \varphi(\mathring{x}) = \operatorname{conv}\left\{\partial^L \varphi(\mathring{x})\right\}.$$

Furthermore, one obtains from Proposition 3.9 that

$$\partial^K \varphi(\mathring{x}) \subset \partial^L \varphi(\mathring{x}).$$

This implies for $g \in \partial^L \varphi(\mathring{x})$ as limiting gradient of the abs-linearization $\Delta\varphi(x; \cdot)$ that at the argument $\Delta x = 0$ one has $g \in \partial^K \varphi(\mathring{x}) \subset \hat{\partial}^M \varphi(\mathring{x})$. Hence, g defines a supporting hyperplane of $\varphi(\mathring{x})$ at $\mathring{x}$. Using this property and the approximation order given by (1) it follows that g also defines a supporting hyperplane for $\Delta\varphi(\mathring{x}; \cdot)$ at $\Delta x = 0$. Since this holds for all elements $g \in \partial^L \Delta\varphi(\mathring{x}; 0)$, $\Delta\varphi(\mathring{x}; \cdot)$ must be convex in some ball about $\Delta x = 0$. Therefore, $\varphi(\cdot)$ is first-order convex in some ball about $\mathring{x}$.

Now assume in addition that MFKQ holds for $\varphi(\cdot)$ at $\mathring{x}$. It is left to show that FOC then implies regularity of $\varphi(\cdot)$ at $\mathring{x}$. Using Proposition 3.9, the approximation order of the linearization, and FOC, one has that all elements of $\partial^L\varphi(\mathring{x})$ are supporting hyperplanes of $\varphi(\cdot)$ at $\mathring{x}$ yielding

$$\partial^L\varphi(\mathring{x}) \subset \widehat{\partial}^M\varphi(\mathring{x}).$$

Since the definition of $\widehat{\partial}^M\varphi(\mathring{x})$ then also includes all convex combinations of two arbitrary elements of $\partial^L\varphi(\mathring{x})$, one obtains

$$\begin{aligned} &\partial^C\varphi(\mathring{x}) = \text{conv}\left\{\partial^L\varphi(\mathring{x})\right\} \subset \widehat{\partial}^M\varphi(\mathring{x}) \subset \partial^M\varphi(\mathring{x}) \subset \partial^C\varphi(\mathring{x}) \\ \Rightarrow \quad &\widehat{\partial}^M\varphi(\mathring{x}) = \partial^M\varphi(\mathring{x}) \end{aligned}$$

and therefore the regularity of $\varphi(\cdot)$ at $\mathring{x}$. ∎

As can be seen, FOC implies the inclusion $\partial^L\varphi(x) \subset \partial^M\varphi(x)$. On the other hand, one might assume that this inclusion also implies FOC and therefore regularity. However, Example 2.8 for $\varepsilon \in [-1, 1)$ shows that this is not the case.

The findings so far can be summarized as follows. For a C^d_{abs} function $\varphi(\cdot)$,

- in the general case we have

$$\emptyset \neq \partial^K\varphi(x) \subset \partial^L\varphi(x) \subset \partial^M\varphi(x) \subset \partial^C\varphi(x)$$

  as well as

$$\varphi(\cdot) \text{ regular in } \mathring{x} \quad \Leftrightarrow \quad \partial^C\varphi(\mathring{x}) = \partial^M\varphi(\mathring{x}) \quad \Rightarrow \quad \varphi(\cdot) \text{ FOC in } \mathring{x},$$

- if MFKQ holds in $\mathring{x}$, one has additionally

$$\partial^K\varphi(\mathring{x}) = \partial^L\varphi(\mathring{x}) \qquad \text{and} \qquad \varphi(\cdot) \text{ regular at } \mathring{x} \quad \Leftrightarrow \quad \varphi(\cdot) \text{ FOC at } \mathring{x}.$$

As an immediate corollary the implication chain CON $\Longrightarrow$ REG $\Longrightarrow$ FOC (that is, convexity of C^d_{abs} functions is inherited by their abs-linearizations) holds, as we claimed in the introduction.

COROLLARY 3.12. *The C^d_{abs} function $\varphi : \mathbb{R}^n \to \mathbb{R}$ can only be convex on some ball about a point $\mathring{x}$ if its abs-linearization $\Delta\varphi(\mathring{x};\cdot) : \mathbb{R}^n \to \mathbb{R}$ is convex for Δx in some ball about the origin, i.e., if $\varphi(\cdot)$ is first-order convex at $\mathring{x}$.*

Proof. If $\Delta\varphi(\mathring{x};\Delta x)$ is not convex on a ball about the origin on which it is homogeneous, then there exist increments $u \neq v$ from within this ball such that

$$\Delta\varphi(\mathring{x};(u+v)/2) = \epsilon + \tfrac{1}{2}[\Delta\varphi(\mathring{x};u) + \Delta\varphi(\mathring{x};v)] \quad \text{with} \quad \varepsilon > 0.$$

Due to the homogeneity, we find, for $\tau \in (0, 1)$, that

$$\Delta\varphi(\mathring{x};\tau\,(u+v)/2) = \tau\epsilon + \tfrac{1}{2}[\Delta\varphi(\mathring{x};\tau\,u) + \Delta\varphi(\mathring{x};\tau\,v)].$$

Since $\Delta\varphi(\mathring{x};\Delta x)$ is a second-order approximation of $\varphi(\mathring{x}+\Delta x) - \varphi(\mathring{x})$, we have

$$\begin{aligned} \varphi(\mathring{x}+\tau(u+v)/2) + \mathcal{O}(\tau^2) &= \varphi(\mathring{x}) + \Delta\varphi(\mathring{x};\tau\,(u+v)/2) \\ &= \tau\epsilon + \tfrac{1}{2}[\varphi(\mathring{x}+\tau u) + \varphi(\mathring{x}+\tau v)] + \mathcal{O}(\tau^2). \end{aligned}$$

Hence, it is clear that for τ small enough the positive $\tau\epsilon$ term will dominate the two $\mathcal{O}(\tau^2)$ terms, which shows that φ itself cannot be convex, yielding a contradiction. ∎

Note that the convexity of $\Delta\varphi(\mathring{x};\cdot)$ is necessary but not sufficient for the convexity of $\varphi(\cdot)$, as illustrated by Example 2.6.

Now the question is, of course, how we can determine whether the abs-linear approximation is convex. This can be answered as follows.

THEOREM 3.13. *Suppose that, for the C^d_{abs} function $\varphi : \mathbb{R}^n \to \mathbb{R}$, the corresponding abs-normal form is localized at $\mathring{x}$, i.e., $\mathring{\alpha} = \alpha(\mathring{x}) = \{1,\ldots,s\}$. Furthermore, assume that LIKQ holds at $\mathring{x}$ so that $s \le n$. Then the abs-linearization of φ at $\mathring{x}$ is locally convex if and only if, componentwise,*

$$b^\top (I - DL)^{-1} \ge 0 \quad \textit{whenever} \quad |D| \le 1, \tag{28}$$

where D ranges over all diagonal matrices and the inequality is meant componentwise. In the nonlocalized case, L and b in (28) *must be replaced by $\check{L}$ as defined in* (23) *and $\check{b} = (b_i)_{i\in\mathring{\alpha}}$.*

Proof. For notational simplicity let us assume that φ itself is piecewise linear, that x is the origin, and $\varphi(0) = 0$ so that $\Delta\varphi(0;x) = \varphi(x)$. Consider any definite signature $\sigma \in \{-1,1\}^s$ and the corresponding $\Sigma = \text{diag}(\sigma)$. Then P_σ is a convex cone in which, by (13), the function $-\Sigma z^\sigma$ has the Jacobian $(I - L\Sigma)^{-1}Z$, yielding

$$N \equiv \Sigma\nabla z^\sigma = \Sigma(I - L\Sigma)^{-1}Z = (\Sigma - L)^{-1}Z \in \mathbb{R}^{s\times n}.$$

The ith row $\nu^i \equiv e_i^\top N$ represents the outward normal of P_σ on the $(n-1)$-dimensional interface with its neighbor cone P_{σ^i} defined by

$$\sigma^i_j = \begin{cases} \sigma_j & \text{if } j \ne i, \\ -\sigma_i & \text{if } j = i. \end{cases}$$

Therefore, the corresponding diagonal matrices satisfy

$$\Sigma^i = \text{diag}(\sigma^i) = \Sigma - 2\sigma_i e_i e_i^\top. \tag{29}$$

By (10) the gradients of φ in P_σ and P_{σ^i} are given by

$$g_\sigma = a^\top + b^\top(\Sigma - L)^{-1}Z \qquad \text{and} \qquad g_{\sigma^i} = a^\top + b^\top(\Sigma^i - L)^{-1}Z.$$

Convexity across the boundary between P_σ and P_{σ^i} is given if and only if we have $(g_{\sigma^i} - g_\sigma)^\top \nu^i \ge 0$. The difference between the two gradients can be worked out by the Sherman–Morrison formula as

$$\begin{aligned} g_{\sigma^i} - g_\sigma &= b^\top(\Sigma - 2\sigma_i e_i e_i^\top - L)^{-1}Z - b^\top(\Sigma - L)^{-1}Z \\ &= b^\top\left[(\Sigma - L)^{-1} + 2\sigma_i \frac{(\Sigma - L)^{-1}e_i e_i^\top(\Sigma - L)^{-1}}{1 - 2\sigma_i e_i^\top(\Sigma - L)^{-1}e_i} - (\Sigma - L)^{-1}\right] Z. \end{aligned}$$

Because of the triangular nature of L, the denominator reduces to $1 - 2\sigma_i^2 = -1$ and we obtain

$$g_{\sigma^i} - g_\sigma = -2b^\top(\Sigma - L)^{-1}e_i\sigma_i e_i^\top(\Sigma - L)^{-1}Z = 2b^\top(\Sigma - L)^{-1}\sigma_i e_i \nu^i.$$

Hence, we see that convexity requires $b^\top(\Sigma - L)^{-1}\sigma_i e_i \ge 0$ for all i and thus requires $b^\top(\Sigma - L)^{-1}\Sigma \ge 0$ componentwise. Moreover, this vector inequality must hold for all definite signatures $\sigma \in \{-1,1\}^s$. Since $b^\top(D - L)^{-1} \ge 0$ is multilinear in the components of the entries of a diagonal matrix D we conclude that the condition given in the assertion is indeed necessary for local convexity of the piecewise linear function $\varphi(x) = \Delta\varphi(0; x)$. So it only remains to show that it is also sufficient. Again, take any two points $u \neq v$ in the homogeneous vicinity of 0. We have to show that $\varphi((u + v)/2) \leq (\varphi(u) + \varphi(v))/2$. Clearly, $\varphi(u(1 - \tau) + \tau v)$ is a piecewise linear function of $\tau \in (0, 1)$. By an arbitrary small perturbation of u and v we can ensure that the line segment between them moves repeatedly from one cone P_σ to one of its neighbors penetrating the interface transversally. Any one of these finitely many kinks is convex, i.e., bends upwards so that the whole piecewise linear function is convex. By continuity, this then follows for the original, unperturbed pair (u, v) as well.

If the abs-normal form of φ is nonlocalized at $\mathring{x}$, the analysis is very similar. As above, one obtains that $\nu^i \equiv e_i^\top N$, $i \in \mathring{\alpha}(\mathring{x})$, represents the outward normal of $\mathcal{P}_\sigma$ on the interface made with its neighbor cones $\mathcal{P}_{\sigma^i}$ with the signature matrices Σ^i, $i \in \mathring{\alpha}(\mathring{x})$, where Σ^i is defined as in (29). Along exactly the same lines as above, one obtains that the inequalities

$$b^\top (D - L)^{-1} e_i \geq 0$$

must hold for all $i \in \mathring{\alpha}(\mathring{x})$ and all

$$D = \Sigma(\mathring{x}) + \sum_{i \in \mathring{\alpha}(\mathring{x})} d_i e_i e_i^\top \quad \text{with} \quad |D| \leq 1.$$

Then, one can argue again that this property is also sufficient. ∎

4. Support conditions. In this section we establish the implications on the right-hand side of Figure 1. The implication FOC $\Longrightarrow$ FOS simply states that if the piecewise linearization is convex near the origin, it has a subgradient g at the origin. The converse need not be true, as is demonstrated by the gradient cube example, which always has the subgradient $g = 0$ but need not be convex in any neighborhood of the origin. It is similarly obvious that any subgradient of the piecewise linearization will be a regular subgradient of φ itself and that at a local minimum $g = 0$ is a subgradient. Hence, we certainly have MIN $\Longrightarrow$ SUP $\Longrightarrow$ FOS, and all that is left to prove is the converse implication that FOS implies SUP under MFKQ and additional curvature conditions. If FOS holds, then the shifted function $(\Delta\varphi(x; \Delta x) - g^\top \Delta x)$ has 0 as a local minimizer. As shown in [10, Thms. 4.1 and 4.3], this yields SUP under MFKQ if certain second-order conditions are satisfied by the pair (f, F). The linear shift by g makes no difference, which is why we have not restated the curvature condition (POSC) in detail. As of now, it also looks combinatorially difficult to falsify, except in its strong form.

So now let us briefly examine the conditions for FOS to be satisfied, i.e., whether one can find for $\mathring{x}$ fixed a vector $g \in \mathbb{R}^n$ such that

$$\Delta\varphi(\mathring{x}, \Delta x) \geq g^\top \Delta x.$$

If such a vector g exists, the piecewise linear function

$$\Delta\varphi(\mathring{x}, \Delta x) - g^\top \Delta x \tag{30}$$

has a local minimum at $\Delta x = 0$. Assuming that φ and thus $\Delta\varphi(\mathring{x}; \Delta x)$ are localized at $\mathring{x}$, i.e., $\alpha(\mathring{x}) = \{1, \ldots, s\}$, and that LIKQ holds, it follows from the necessary optimality conditions shown in [8] that a vector $\mu \in \mathbb{R}^s$ of Lagrange multipliers exists such that tangential stationarity is fulfilled. Hence, with a as defined in (8),

$$a^\top - g^\top + \mu^\top Z = 0 \quad \Leftrightarrow \quad g^\top = a^\top + \mu^\top Z \tag{31}$$

must hold. Furthermore, the Lagrange vector μ must fulfill the normal growth condition, yielding

$$b^\top \geq |\mu|^\top - L^\top \mu, \tag{32}$$

where b and L are given by the abs-linearization. If such a $\mu \in \mathbb{R}^s$ exists, one can determine a g that defines a supporting hyperplane using (31) and thus verifies FOS. System (32) can be interpreted as a feasibility problem in linear programming that is comparatively easy to solve.

Obviously, the requirement of φ being localized at $\mathring{x}$ is rather strong. Therefore, now we derive conditions for FOS in the nonlocalized situation, where $|\alpha(\mathring{x})| \neq s$. As proposed in [8, sect. 4], one can then split z into two parts such that

$$\begin{aligned} \hat{z} &\equiv (\sigma_i z_i)_{i \notin \alpha} \equiv (|z_i|)_{i \notin \alpha} \in \mathbb{R}^{|\sigma|}, \\ \check{z} &= (z_i)_{i \in \alpha} \in \mathbb{R}^{|\alpha|} \end{aligned}$$

and the components of $\hat{z}$ will keep their positive sign in a neighborhood B of $\mathring{x}$. Similarly, one can define the two functions

$$\begin{aligned} \hat{F} &= (\sigma_i e_i^\top F)_{i \notin \alpha} : \mathbb{R}^{n+|\alpha|+|\sigma|} \to \mathbb{R}^{|\sigma|}, \quad \hat{z} = \hat{F}(x, \bar{z}, \hat{z}) \quad \text{with} \quad \bar{z} = |\check{z}| \in \mathbb{R}^{\alpha}, \\ \check{F} &= (e_i^\top F)_{i \in \alpha} : \mathbb{R}^{n+|\alpha|+|\sigma|} \to \mathbb{R}^{|\alpha|}, \qquad \check{z} = \check{F}(x, \bar{z}, \hat{z}), \end{aligned}$$

and write $f(x, \bar{z}, \hat{z})$ instead of $f(x, |z|)$. If LIKQ holds at $\mathring{x}$ and φ is nonlocalized at $\mathring{x}$, it follows from the necessary optimality conditions shown in [8, Prop. 4] that, for $\tilde{\varphi}$ as defined in (30), $\mathring{x}$ can only be a local minimizer of $\Delta\varphi(\mathring{x}, \Delta x) - g^\top \Delta x$ if there exist two unique multiplier vectors $\check{\mu} \in \mathbb{R}^{|\alpha|}$ and $\hat{\mu} \in \mathbb{R}^{|\sigma|}$ satisfying tangential stationarity, i.e.,

$$\begin{bmatrix} f_x - g & f_{\hat{z}} \end{bmatrix} = -\begin{bmatrix} \check{\mu}^\top & \hat{\mu}^\top \end{bmatrix} \begin{bmatrix} \check{F}_x & \check{F}_{\hat{z}} \\ \hat{F}_x & \hat{F}_{\hat{z}} - I \end{bmatrix} \in \mathbb{R}^{n+|\sigma|}, \tag{33}$$

and the normal growth condition

$$f_{\bar{z}} \geq |\check{\mu}^\top| - \begin{bmatrix} \check{\mu}^\top & \hat{\mu}^\top \end{bmatrix} \begin{bmatrix} \check{F}_{\bar{z}} \\ \hat{F}_{\bar{z}} \end{bmatrix} \in \mathbb{R}^{|\alpha|}, \tag{34}$$

yielding

$$g = -f_x + \check{\mu}^\top \check{F}_x + \hat{\mu}^\top \check{F}_{\hat{z}}.$$

Hence, one can easily derive a supporting vector g if the Lagrange multipliers $\check{\mu}^\top$ and $\hat{\mu}^\top$ satisfying (34) exist, which is again equivalent to a linear optimization problem.

5. Complexity analysis for FOC. We proved in [8] that, when LIKQ holds, stationarity and normal growth, i.e., first-order optimality, can be tested in polynomial time. So far, we have not succeeded in deriving a test that is polynomial in time for testing stationarity under MFKQ. Even testing for MFKQ at $\mathring{x}$ seems to be expensive because it is not polynomial in time, due to the existence of a vector v with $J_\sigma v > 0$ for all $\sigma \succeq \mathring{\sigma}$ that has to be verified.

While these statements remain conjectures for the time being, we will prove here that the convexity test derived in the proof of Theorem 3.13 is co-NP-complete. For this purpose, it can be stated as follows.

DEFINITION 5.1 (CONV). *For a given pair (b, L) with $b \in \mathbb{R}^s$ and $L \in \mathbb{R}^{s\times s}$ strictly lower triangular, the verification of*

$$b^\top(\Sigma - L)^{-1}\Sigma \geq 0 \quad \forall \Sigma = \mathrm{diag}(\sigma) \quad \textit{with} \quad \sigma \in \{-1, 1\}^s$$

is equal to the convexity test for a corresponding C^d_{abs} function $\varphi : \mathbb{R}^n \to \mathbb{R}$ having the vector b and the matrix L as parts of its abs-normal form at a point $\mathring{x}$. Again using $\Sigma = \mathrm{diag}(\sigma)$, the convexity test can be rewritten as

$$\text{(CONV)} \qquad \forall\sigma \in \{-1, 1\}^s,\ \forall i \in \{1, \ldots, s\} : \left(yb^\top(\Sigma - L)^{-1}\Sigma\right)_i \geq 0.$$

Its complement is given by

$$(\overline{\text{CONV}}) \qquad \exists\sigma \in \{-1, 1\}^s,\ \exists i \in \{1, \ldots, s\} : \left(b^\top(\Sigma - L)^{-1}\Sigma\right)_i < 0.$$

LEMMA 5.2. *The problem* (CONV) *is an element of the complexity class co-NP.*

Proof. We have to show that $(\overline{\text{CONV}})$ is an element of the complexity class NP. Then its complement, i.e., (CONV) is an element of co-NP. For any given $\sigma \in \{-1, 1\}^s$, one can compute the whole vector $v \equiv b^\top(\Sigma - L)^{-1}\Sigma$ by one forward substitution using $\mathcal{O}(s^2)$ operations. Then, one needs at most additional s operations to check whether there exists an index i such that one component of the vector v is less than zero. It follows that given an instance $\sigma \in \{-1, 1\}^s$ one can decide in polynomial time whether $(\overline{\text{CONV}})$ holds for this particular σ or not. Therefore, $(\overline{\text{CONV}})$ is an element of the complexity class NP. ∎

To show, that (CONV) is also co-NP-complete we will reduce a co-NP-complete decision problem in polynomial time to the decision problem (CONV).

DEFINITION 5.3 (TAUTOLOGY). *The decision problem*

$$\forall x \in \{0, 1\}^N : \psi(x) = \bigwedge_{i=1}^{m} \psi_i(x) = 1,$$

where each clause $\psi_i(x)$, $i = 1, \ldots, m$, is limited to a disjunction of at most three literals and a literal is either a variable or the negation of a variable is called TAUTOLOGY.

Usually, TAUTOLOGY is not restricted to this specific form of conjunctions of special clauses but similar to the general SAT problem and its restriction to 3-SAT, one can consider this special form of TAUTOLOGY.

THEOREM 5.4. *The decision problem* (CONV) *is co-NP-complete.*

Proof. The proof comprises two parts. First, we construct a polynomial reduction algorithm f that maps a given instance ψ of a TAUTOLOGY decision problem to one specific instance $f(\psi)$ of (CONV). Then we show that ψ is a tautology if and only if the convexity test holds for the instance $f(\psi)$ of (CONV).

To derive the reduction algorithm, we exploit the following fact: for a given instance

$$\forall x \in \{0,1\}^N : \psi(x) = \bigwedge_{i=1}^{m} \psi_i(x) = 1$$

of TAUTOLOGY, one can define σ_{ij}, $j = 0, \dots, 3$, corresponding to each clause $\psi_i(x)$ involving the three variables x_{i1}, x_{i2}, x_{i3} by

$$\sigma_{i0} \in \{-1, 1\} \text{ arbitrarily,}$$
$$\sigma_{ij} = \begin{cases} -1 + 2x_{ij} & \text{if } x_{ij} \text{ occurs in } \psi_i(x), \\ 1 - 2x_{ij} & \text{if the negation of } x_{ij} \text{ occurs in } \psi_i(x), \end{cases} \qquad j = 1, 2, 3. \tag{35}$$

Then, one can show by a simple truth table that

$$\psi_i(x) = 1 \ \Leftrightarrow\ \sigma_{i1} + \sigma_{i2} + \sigma_{i3} \geq -1 \qquad \text{and} \qquad \psi_i(x) = 0 \ \Leftrightarrow\ \sigma_{i1} + \sigma_{i2} + \sigma_{i3} = -3.$$

This observation will be used to construct f. First, we define for each clause $\psi_i(x)$ a corresponding block L_i of a strictly lower triangular matrix $\tilde{L} = \operatorname{diag}((L_i)_{i=1,\dots,m})$ in the following way. The clause $\psi_i(x)$ involves three variables: x_{i1}, x_{i2}, x_{i3}. Define $\Sigma_i = \operatorname{diag}(\sigma_{i0}, \sigma_{i1}, \sigma_{i2}, \sigma_{i3})$ and $b_i = (1, 1, 1, 1)^\top$. Then, one obtains that

$$(\Sigma_i - L_i)^{-1}\Sigma_i = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \sigma_{i1} & 1 & 0 & 0 \\ \sigma_{i2} & 0 & 1 & 0 \\ \sigma_{i3} & 0 & 0 & 1 \end{pmatrix} \quad \text{with} \quad L_i = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \tag{36}$$

$$b_i^\top (\Sigma_i - L_i)^{-1}\Sigma_i = (1 + \sigma_{i1} + \sigma_{i2} + \sigma_{i3}, 1, 1, 1). \tag{37}$$

Setting

$$b_\psi = (b_i)_{i=1,\dots,m} \qquad \text{and} \qquad \tilde{L} = \operatorname{diag}((L_i)_{i=1,\dots,m}),$$

one obtains an important part of instance $f(\psi)$. It follows that

$$\begin{aligned} \psi_i(x) = 1 \quad &\Rightarrow \quad 1 + \sigma_{i1} + \sigma_{i2} + \sigma_{i3} \geq 0 \quad \text{and} \\ \psi_i(x) = 0 \quad &\Rightarrow \quad 1 + \sigma_{i1} + \sigma_{i2} + \sigma_{i3} = -2 < 0. \end{aligned}$$

Hence, if $\psi_i(x) = 1$ for all x, one obtains that $1 + \sigma_{i1} + \sigma_{i2} + \sigma_{i3} \geq 0$ holds for the corresponding entry of the vector $b_\psi^\top(\Sigma - \tilde{L})^{-1}\Sigma$ when defining Σ according to (35). All remaining entries of this vector are greater than zero by definition. However, so far the different σ_{ij} representing the same component x_l, $1 \leq l \leq N$, of x are not coupled. Hence, we might have for one $\Sigma_\psi = \operatorname{diag}(\sigma_{10}, \dots, \sigma_{m3})$ that

$$b_\psi^\top(\Sigma_\psi - L_\psi)^{-1}\Sigma_\psi \geq 0,$$

but it is not possible to reconstruct a corresponding x such that $\psi(x) = 1$, since Σ_ψ might contain contradicting values for one component of x. Therefore, we introduce additional coupling conditions such that all x_{ij} corresponding to the same component x_l of x have a consistent value. For this purpose, we count how often each component of x occurs in $\psi(x)$. This number can be determined in polynomial time and is denoted by c_l, $l = 1, \dots, N$. Setting

$$M \equiv \sum_{l=1}^{N} 2(c_l - 1), \quad M_l \equiv \sum_{k=1}^{l-1} 2(c_k - 1), \quad \text{and} \quad \hat{m} \equiv 4m,$$

we construct a strictly lower triangular matrix

$$L = \begin{pmatrix} 0_{M\times M} & 0_{M\times \hat{m}} \\ C & \tilde{L} \end{pmatrix},$$

where C defines the coupling of the x_{ij} corresponding to the same component x_l. This coupling will be done in the following way. If x_l appears only once in $\psi(x)$, nothing has to be done. For $c_l > 1$ assume that x_l occurs for the kth time, $k \in \{1, \ldots, c_l - 1\}$, in the clause ψ_i at place j_i and that the next appearance of x_l, i.e., the $(k+1)$st one, is in $\psi_{\hat{\imath}}$ at place $j_{\hat{\imath}}$. Then, one places a 1 in the entry $(M_j + 2(k-1) + 1, 4(i-1) + 1 + j_i)$, i.e., in the column $M_j + 2(k-1) + 1$ and row $4(i-1) + 1 + j_i$. The entry in row $j_{\hat{\imath}}$ in the same column depends on the specific appearance of x_l in ψ_i and $\psi_{\hat{\imath}}$. It is set to

$$\begin{array}{rl} 1 & \text{if } x_l \text{ occurs in } \psi_i \text{ and its negation in } \psi_{\hat{\imath}}, \\ 1 & \text{if the negation of } x_l \text{ occurs in } \psi_i \text{ and } x_l \text{ occurs in } \psi_{\hat{\imath}}, \\ -1 & \text{if } x_l \text{ occurs in } \psi_i \text{ and also in } \psi_{\hat{\imath}}, \\ -1 & \text{if the negation of } x_l \text{ occurs in } \psi_i \text{ and also in } \psi_{\hat{\imath}}. \end{array}$$

all remaining entries of this column in C are set to zero. Hence, there are only two nonzero entries in the column $M_j + 2(k-1) + 1$ of C, each with the absolute value 1. The next column $M_j + 2(k-1) + 2$ of C is set to the column $M_j + 2(k-1) + 1$ of C times -1. For an illustration of this construction see Example 5.5. Now, we have to examine the matrix $(\Sigma - L)^{-1}\Sigma$ for which one has

$$(\Sigma - L)^{-1}\Sigma = \begin{pmatrix} I_{M\times M} & 0_{M\times \hat{m}} \\ \tilde{C} & (\Sigma_\psi - L_\psi)^{-1}\Sigma_\psi \end{pmatrix},$$

where the upper part follows immediately from the structure of L, the lower right part was examined above, and only the lower left part $\tilde{C}$ has to be derived. If $c_l > 1$ and x_l occurs for the kth time, $k \in \{1, \ldots, c_l - 1\}$, in clause ψ_i at place j_i and the next appearance of x_l, i.e., the $(k+1)$st one, is in $\psi_{\hat{\imath}}$ at place $j_{\hat{\imath}}$, then the entry $(M_j+2(k-1)+1, 4(i-1)+1+j_i)$, i.e., in column $M_j+2(k-1)+1$ and row $4(i-1)+1+j_i$ is σ_{ij_i}. The entry in the row $j_{\hat{\imath}}$ in the same column is the corresponding entry in C multiplied with $\sigma_{\hat{\imath}j_{\hat{\imath}}}$. Once more, the column $M_j + 2(k-1) + 2$ of $\tilde{C}$ equals column $M_j + 2(k-1) + 1$ multiplied by -1. Now, setting $b = (0_M, 1_{\hat{m}})^\top$, if x_l occurs in ψ_i and $\psi_{\hat{\imath}}$ or if the negation of x_l occurs in ψ_i and also in $\psi_{\hat{\imath}}$, the coupling conditions yield

$$\begin{array}{rl} \sigma_{ij_i} - \sigma_{\hat{\imath}j_{\hat{\imath}}} \geq 0 & \text{in column } M_j + 2(k-1) + 1, \\ -\sigma_{ij_i} + \sigma_{\hat{\imath}j_{\hat{\imath}}} \geq 0 & \text{in column } M_j + 2(k-1) + 2, \end{array}$$

requiring that $\sigma_{ij_i} = \sigma_{\hat{\imath}j_{\hat{\imath}}}$. If x_l occurs in ψ_i and its negation occurs in $\psi_{\hat{\imath}}$, or if the negation of x_l occurs in ψ_i and x_l occurs in $\psi_{\hat{\imath}}$, one gets

$$\begin{array}{rl} \sigma_{ij_i} + \sigma_{\hat{\imath}j_{\hat{\imath}}} \geq 0 & \text{in column } M_j + 2(k-1) + 1, \\ -\sigma_{ij_i} - \sigma_{\hat{\imath}j_{\hat{\imath}}} \geq 0 & \text{in column } M_j + 2(k-1) + 2, \end{array}$$

requiring that $\sigma_{ij_i} = -\sigma_{\hat{\imath}j_{\hat{\imath}}}$. Since all appearances of x_l are covered in that way, one can reconstruct, from all Σ satisfying the convexity condition for $f(\psi)$, a corresponding x that fulfills ψ.

It also follows immediately that all x with $\psi(x) = 1$ define a corresponding Σ such that the convexity condition holds. This yields the assertion that (CONV) is co-NP-complete. ∎

The following example illustrates this rather involved reduction from an instance of TAUTOLOGY to an instance of (CONV).

Example 5.5 (polynomial reduction). Consider for $x \in \{0,1\}^4$ the instance

$$\psi(x) = (x_1 \vee \neg x_2 \vee x_3) \wedge (x_1 \vee \neg x_3 \vee x_4) \wedge (x_1 \vee x_2 \vee x_3)$$

yielding

$$\begin{gathered} c_1 = 3, \quad c_2 = 2, \quad c_3 = 3, \quad c_4 = 1, \\ M = 10, \quad M_1 = 0, \quad M_2 = 4, \quad M_3 = 6, \quad \hat{m} = 12. \end{gathered}$$

Furthermore, using the polynomial reduction introduced in the proof of Theorem 5.4, one has that x_1 is represented by σ_{11}, σ_{21}, and σ_{31}; x_2 is represented by σ_{12} and σ_{32}; x_3 is represented by σ_{13}, σ_{22}, and σ_{33}; and x_4 is represented by σ_{23}. The coupling matrix is given by

$$\begin{aligned} &C = (c_1, -c_1, c_2, -c_2, c_3, -c_3, c_4, -c_4, c_5, -c_5) \in \mathbb{R}^{12\times 10} \qquad \text{with} \\ &c_1 = e_2 - e_6, \;\; c_2 = e_6 - e_{10}, \;\; c_3 = e_3 - e_{11}, \;\; c_4 = e_4 + e_7, \;\; c_5 = e_7 + e_{12}, \end{aligned}$$

where e_i denotes the ith unit vector in $\mathbb{R}^{12}$. Then one has

$$(\Sigma - L)^{-1}\Sigma = \begin{pmatrix} I_{M\times M} & 0_{M\times\hat{m}} \\ \tilde{C} & \tilde{L}_\psi \end{pmatrix} \quad \text{with}$$

$$\tilde{C} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \sigma_{11} & -\sigma_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{12} & -\sigma_{12} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{13} & -\sigma_{13} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\sigma_{21} & \sigma_{21} & \sigma_{21} & -\sigma_{21} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{22} & -\sigma_{22} & \sigma_{22} & -\sigma_{22} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\sigma_{31} & +\sigma_{31} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{32} & -\sigma_{32} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{33} & -\sigma_{33} \end{pmatrix},$$

and $\tilde{L}_\psi$ is a block diagonal matrix according to (36). Now, using $b = (0_M, 1_{\hat{m}})^\top$, one obtains consistent values for the components of x from $b^\top(\Sigma - L)^{-1}\Sigma$ since

$$\begin{aligned} \sigma_{11} - \sigma_{21} \geq 0, \quad -\sigma_{11} + \sigma_{21} \geq 0 \quad &\Rightarrow \quad \sigma_{11} = \sigma_{21}, \\ \sigma_{21} - \sigma_{31} \geq 0, \quad -\sigma_{21} + \sigma_{31} \geq 0 \quad &\Rightarrow \quad \sigma_{21} = \sigma_{31}, \\ \sigma_{12} + \sigma_{32} \geq 0, \quad -\sigma_{12} - \sigma_{32} \geq 0 \quad &\Rightarrow \quad \sigma_{12} = -\sigma_{32}, \\ \sigma_{13} + \sigma_{22} \geq 0, \quad -\sigma_{13} - \sigma_{22} \geq 0 \quad &\Rightarrow \quad \sigma_{13} = -\sigma_{22}, \\ \sigma_{22} + \sigma_{33} \geq 0, \quad -\sigma_{22} - \sigma_{33} \geq 0 \quad &\Rightarrow \quad \sigma_{22} = -\sigma_{33}. \end{aligned}$$

6. Summary and outlook. In this paper we examined the relation between several generalized derivative concepts for piecewise smooth functions that can be represented in abs-normal form and thus belong to the class C^d_{abs}. In particular, we studied the chain of conical, limiting, and Clarke generalized derivatives. We also studied the computational complexity of testing particular properties under the certain kink qualifications, namely LIKQ or the more general kink qualification MFKQ introduced in [10]. In contrast to LIKQ, it has not so far been possible to verify MFKQ itself in polynomial time. We give five examples of piecewise differentiable functions in abs-normal form that satisfy LIKQ or MFKQ or neither. Next, we studied convexity conditions for $C^d_{\rm abs}$ functions and their abs-linearization under LIKQ and MFKQ. The main result is that first-order convexity and regularity are equivalent under MFKQ. Moreover, Fréchet subdifferentiability is equivalent to FOS. The implications of the various properties are displayed in Figure 1. Furthermore, we proved that testing for convexity is co-NP-complete even under LICQ and thus certainly under MFKQ. Thus, we conclude that, at least on the class of $C^d_{\rm abs}$ functions, subdifferential regularity is a rather theoretical, nonconstructive concept. The complexity analysis for the test of whether MFKQ holds at a given point is still an open problem and the subject of future research. Generally speaking, it seems that the combinatorial aspect and its computational complexity deserve more attention in nonsmooth analysis. Moreover, it is not clear whether some other generalization of MFCQ is not better suited for the piecewise smooth case. Recently, the process of piecewise linearization has been extended to the class $C^d_{\rm euc} \supset C^d_{\rm abs}$ of functions representable by smooth elementals and the Euclidean norm [9]. The implications in Figure 1 on this more general class of Lipschitzian but not piecewise differentiable functions remain to be seen.

Acknowledgments. The authors would like to thank Johannes Blömer for discussions on complexity analysis. The authors acknowledge the advice of Manuel Radons concerning the NP-co-completeness result. Furthermore, the authors appreciate the careful reading and suggestions of two anonymous reviewers.

REFERENCES

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

[2] H. Bauschke, Y. Lucet, and H. Phan, *On the convexity of piecewise-defined functions*, ESAIM Control Optim. Calc. Var., 22 (2016), pp. 728–742.

[3] L. Eneya, T. Bosse, and A. Griewank, *A method for pointwise evaluation of polyconvex envelopes*, Afr. Mat., 24 (2013), pp. 1–24.

[4] B. Gardiner and Y. Lucet, *Computing the conjugate of convex piecewise linear-quadratic bivariate functions*, Math. Program., 139 (2013), pp. 161–184.

[5] A. Griewank, *On stable piecewise linearization and generalized algorithmic differentiation*, Optim. Methods Softw., 28 (2013), pp. 1139–1178.

[6] A. Griewank, J.-U. Bernt, M. Randons, and T. Streubel, *Solving piecewise linear equations in abs-normal form*, Linear Algebra Appl., 471 (2015), pp. 500–530.

[7] A. Griewank and A. Walther, *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, 2nd ed., SIAM, Philadelphia, PA, 2008.

[8] A. Griewank and A. Walther, *First and second order optimality conditions for piecewise smooth objective functions*, Optim. Methods Softw., 31 (2016), pp. 904–930.

[9] A. Griewank and A. Walther, *Beyond the Oracle: Opportunities of Piecewise Differentiation*, Preprint SPP1962-086, Weierstrass Institute for Applied Analysis and Stochastics, Berlin, 2018.

[10] A. Griewank and A. Walther, *Relaxing kink qualifications and proving convergence rates in piecewise smooth optimization*, SIAM J. Optim., 29 (2019), 262—289.

[11] A. Griewank, A. Walther, S. Fiege, and T. Bosse, *On Lipschitz optimization based on gray-box piecewise linearization*, Math. Program., 158 (2016), pp. 383–415.

[12] M. Gürbüzbalaban and M. Overton, *On Nesterov's nonsmooth Chebyshev–Rosenbrock functions*, Nonlinear Anal., 75 (2012), pp. 1282–1289.

[13] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms I: Fundamentals*, Grundlehren Math. Wiss. 305, Springer, Berlin, 1993.

[14] K. Khan and P. Barton, *Evaluating an element of the Clarke generalized Jacobian of a composite piecewise differentiable function*, ACM Trans. Math. Software, 39 (2013), No. 23.

[15] K. Khan and P. Barton, *A vector forward mode of automatic differentiation for generalized derivative evaluation*, Optim. Methods Softw., 30 (2015), pp. 1185–1212.

[16] A. Kruger, *On Fréchet subdifferentials*, J. Math. Sci. (N.Y.), 116 (2003), pp. 3325–3358.

[17] A. Lewis, *Active sets, nonsmoothness, and sensitivity*, SIAM J. Optim., 13 (2002), pp. 702–725.

[18] O. Mangasarian and S. Fromovitz, *The Fritz John necessary optimality conditions in the presence of equality and inequality constraints*, J. Math. Anal. Appl., 17 (1967), pp. 37–47.

[19] M. Michelsen, *The isothermal flash problem, part* I*: Stability*, Fluid Phase Equilibria, 9 (1982), pp. 1–19.

[20] R. Mifflin and C. Sagastizábal, *Primal-dual gradient structured functions: Second-order results; links to epi-derivatives and partly smooth functions*, SIAM J. Optim., 13 (2003), pp. 1174–1194.

[21] B. Mordukhovich, *Variational Analysis and Generalized Differentiation* I*: Basic Theory*, Grundlehren Math. Wiss. 330, Springer, Berlin, 2006.

[22] N. Nagarajan, A. Cullick, and A. Griewank, *New strategy for phase equilibrium and critical point calculations by thermodynamic energy analysis, part* I*: stability analysis and flash*, Fluid Phase Equilibria, 62 (1991), pp. 191–210.

[23] Y. Nesterov, *Lexicographic differentiation of nonsmooth functions*, Math. Program., 104 (2005), pp. 669–700.

[24] R. Rockafellar and R.-B. Wets, *Variational Analysis*, Grundlehren Math. Wiss. 317, Springer, Cham, 1998.

[25] S. Scholtes, *Introduction to Piecewise Differentiable Functions*, Springer, New York, 2012.