CHARACTERIZING AND TESTING SUBDIFFERENTIAL REGULARITY FOR PIECEWISE SMOOTH OBJECTIVE FUNCTIONS

ANDREA WALTHER[1] AND ANDREAS GRIEWANK[2]

Abstract. Functions defined by evaluation programs involving smooth elementals and absolute values as well as the max- and min-operator are piecewise smooth. Using piecewise linearization we derived in [7] for this class of nonsmooth functions φ first and second order conditions for local optimality (MIN). They are necessary and sufficient, respectively. These generalizations of the classical KKT and SSC theory assumed that the given representation of φ satisfies the Linear-Independence-Kink-Qualification (LIKQ). In this paper we relax LIKQ to the Mangasarin-Fromovitz-Kink-Qualification (MFKQ) and develop a constructive condition for a local convexity concept, i.e., the convexity of the local piecewise linearization on a neighborhood. As a consequence we show that this first order convexity (FOC) is always required by subdifferential regularity (REG) as defined in [20], and is even equivalent to it under MFKQ. Whereas it was observed in [7] that testing for MIN is polynomial under LIKQ, we show here 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.

Keywords: Subdifferential Regularity, First-Order-Convexity, Clarke Gradient, Linear-Independence-Kink-Qualification, Mordukhovich Gradient, Mangasarin-Fromovitz-Kink-Qualification, Abs-Normal-Form

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. Like in algorithmic, or automatic, differentiation [6], the key assumption facilitating this process is that the problem 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 like abs, min, and max as part of the mix, we arrive at a subclass of piecewise smooth functions that can still be analyzed by slight extensions of automatic differentiation tools. However, the resulting extended program does not produce *gradients, Jacobians, Hessians, or Taylor coefficients*, but represent a procedure for evaluating $\Delta\varphi(x;\Delta x)$, an (incremental) piecewise linearization of φ developed at x and evaluated at Δx. The construction of this approximation is given in [4], where we also show that

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

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 at $\Delta x = 0$ a derivative that vanishes. 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 [21].

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 in particular true for the properties:

Local minimality (MIN) and Local convexity (CON)

[1]Institut für Mathematik, Universität Paderborn, Paderborn, Germany
[2]School of Mathematical Sciences and Information Technology, Yachay Tech, Urcuquí, Ecuador

Note that the relation between the two sides of (1) and hence the above implications are 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 $\Delta x \neq 0$. In the companion paper [8] we explore these optimality conditions and the resulting rates of convergence for successive piecewise 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) and First Order Convexity (FOC). They are necessary conditions for φ to have the corresponding properties MIN and CON. 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.

As we will see some of these assumptions are kink qualifications that exclude degeneracies of first derivative matrices, and others are curvature conditions that require second derivative matrices to be definite on certain subspaces. Of course the whole point of the exercise is to characterize these extra assumptions and the properties of $\Delta\varphi(x;\Delta x)$ itself constructively by linear algebra tests on the so-called abs-normal form. In its abs-normal form, $\Delta\varphi(x;\Delta x)$ is represented by several real matrices and vectors, which can be analyzed by various linear algebra procedures. In [7] it was shown how for a scalar-valued function $\varphi(x)$ this information can be used to characterize local optimality, i.e., property MIN in terms of generalized Karush-Kuhn-Tucker (KKT) and Positive Curvature Conditions. This optimality analysis was based on a generalization of the Linear-Independence-Constraint-Qualification (LICQ) called Linear-Independence-Kink-Qualification (LIKQ), which is generic [5], i.e., always satisfiable by arbitrary small perturbations of a given abs-normal form. Of course, some of these perturbations may alter inherent structural properties of a given problem, which is why we wish to relax LIKQ, just like LICQ in the smooth, constrained case. There, a popular relaxation is the Mangasarin-Formovitz-Constraint-Qualification (MFCQ), which naturally generalizes to the Mangasarin-Fromovitz-Kink-Qualification (MFKQ).

Our main thrust here is to characterize subdifferential regularity (REG) for our class of nonsmooth functions. This well known nonsmoothness property is a necessary condition for partial smoothness [13], which yields in turn a special case of the $\mathcal{VU}$ decomposition [16]. As it turns out REG at a point x always requires FOC, i.e., convexity of the piecewise linearization $\Delta\varphi(x;\Delta x)$ with respect to Δx near the origin $\Delta x = 0$. Therefore, we refer to REG also as a *convexity property*, although it strictly speaking 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 2.4. 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 f at x as defined in [20]. 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 a point $x \in \mathbb{R}^n$ if and only if the problem*

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

for $\mu_j \in \mathbb{R}$, $\mu_j \geq 0$, and $x_j \in \mathbb{R}^n$, $j \in \{1,\ldots,n\}$, 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 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 [2] 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 target function of Eq. (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., [15, 18]). 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 following implication chain:

$$CON \implies REG \underset{(MFKQ)}{\overset{\Longrightarrow}{\Longleftarrow}} FOC$$

Figure 1. *Relations between convexity properties of φ*

The one-way implication at the beginning follows directly from the definitions. The relation of interest is the near equivalence between regularity (REG) and first order convexity (FOC). Here, we need the Mangasarin-Fromovitz-Kink-Qualification (MFKQ) for the more difficult, converse implication.

The paper is organized as follows. In Section 2, first we 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. Section 4 discusses the computational complexity of testing for convexity. The paper concludes with a summary and outlook in Section 5.

2. Kink Qualifications for Nonsmooth Problems.

For the definition of kink qualifications, we consider the class of objective functions that are defined as compositions of smooth elemental functions and the absolute value function $\mathrm{abs}(x) = |x|$. This includes also $\max(x, y)$, $\min(x, y)$, 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 elementary function would lead to objectives that are still Lipschitz continuous and lexicographically differentiable [19] but no longer piecewise smooth [21].

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 $i = 1 \ldots s$, where we assume throughout that z_j can only influence z_i if $j < i$. Hence, one obtains 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 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 $\overline{\mathcal{D}} \times \overline{\mathbb{R}^s_+} \subset \mathbb{R}^{n+s}$. Sometimes, we write

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

to denote the objective directly in terms of the argument vector x only. In this paper, we are mostly interested in the case where the nonlinear elementals are all at least once continuously differentiable yielding the following function class:

Definition 2.1. *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 $C^d_{abs}(\overline{\mathcal{D}})$. Since F and f are smooth in the respective arguments, the derivatives*

$$\begin{aligned} L &\equiv \frac{\partial}{\partial |z|} F(x, |z|) \in \mathbb{R}^{s\times s}, & Z &\equiv \frac{\partial}{\partial x} F(x, |z|) \in \mathbb{R}^{s\times n} \; , \\ a &\equiv \frac{\partial}{\partial x} f(x, |z|) \in \mathbb{R}^{n}, \quad \textit{and} & b &\equiv \frac{\partial}{\partial |z|} f(x, |z|) \in \mathbb{R}^{s} \; . \end{aligned} \tag{5}$$

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

Due to our assumption on the numbering of the switching variables, the derivative matrix L is strictly lower triangular. Note that a mathematical map $\varphi \in C^d_{abs}(\overline{\mathcal{D}})$ may have different abs-normal decompositions as shown below for the example proposed by Hiriart-Urruty and Lemaréchal. The properties occurring in Fig. 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 $\sigma(x) \equiv \mathbf{sgn}(z(x))$ and the corresponding diagonal matrix $\Sigma(x) = \mathbf{diag}(\sigma(x))$. 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 consider frequently the situation where σ varies over all possibilities $\{-1, 1\}^s$. As observed already in [9] also for the nonlinear case, the limiting gradients as defined later in Def. 3.2 of φ in the vicinity of x are given by

$$g_\sigma^\top \;\equiv\; a^\top + b^\top \Sigma (I - L\Sigma)^{-1} Z = a^\top + b^\top (\Sigma - L)^{-1} Z \; , \tag{6}$$

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

$$S_\sigma = \{x \in \mathbb{R}^n \,|\, \mathbf{sgn}(z(x)) = \sigma\} \tag{7}$$

as a decomposition of the argument space, where one has

$$\varphi(x) = \varphi_\sigma(x) \qquad \text{for all } x \in S_\sigma \tag{8}$$

and φ_σ is one of finitely many differentiable selection functions in the sense of Scholtes [21]. At a given point x, the nonsmoothness of the target function φ is caused by the so-called *active* switching variables $z_i(x) = 0$ for $1 \le i \le s$. We collect them in the active switch set

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

with $|\sigma| \equiv \|\sigma\|_1$ and $|\alpha|$ defined correspondingly. Later on, we will distinguish two different scenarios for the activity pattern α:

Definition 2.2 (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 also the function φ is localized at x. Otherwise, the switching and also the function itself is nonlocalized.

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

$$\nabla z^\sigma \equiv \frac{\partial}{\partial x} z^\sigma \;=\; (I - L\Sigma)^{-1} Z \in \mathbb{R}^{s \times n}, \tag{9}$$

where Z and L are evaluated at $(x, z^\sigma(x))$.

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

Example 2.3 (HUL). Hiriart-Urruty and Lemaréchal highlighted the piecewise linear, convex function $\varphi : \mathbb{R}^2 \to \mathbb{R}$,

$$\varphi(x_1, x_2) = \max\{-100, 3x_1 - 2x_2, 2x_1 - 5x_2, 3x_1 + 2x_2, 2x_1 + 5x_2\}\,. \tag{10}$$

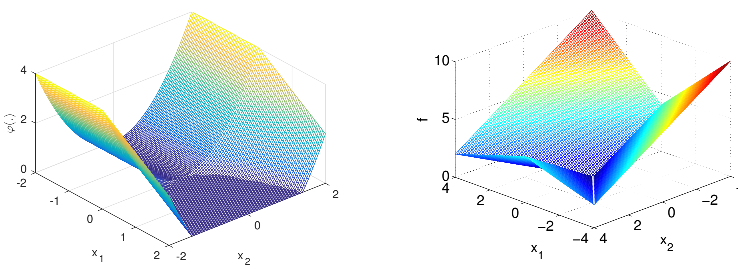
To derive an abs-normal form for this function one could either use the straight forward formulation

$$\begin{aligned} \varphi(x) &= \max\{\max\{\max\{\max\{y_0(x), y_1(x)\}, y_2(x)\}, y_3(x)\}, y_4(x)\} \\ \text{with } & y_0(x) = -100,\; y_1(x) = 3x_1 - 2x_2,\; y_2(x) = 2x_1 - 5x_2, \\ & y_3(x) = 3x_1 + 2x_2,\; y_4(x) = 2x_1 + 5x_2, \end{aligned} \tag{11}$$

yielding four switching variables. Alternatively, one may use the mathematically equivalent description

$$\varphi(x) = \max\{\max\{-100, 2x_1 + 5|x_2|\}, 3x_1 + 2|x_2|\} \tag{12}$$

that requires only three switching variables. As we will see later, the two representations have quite different properties.



(a) $\varphi(x)$ as defined in Eq. (14)

(b) $\varphi(x)$ as defined in Eq. (15) for $\varepsilon = 0.5$

FIGURE 2. *Half Pipe Example and Gradient Cube Example for $n = 2$*

Example 2.4 (Abs-sin). As a simple nonconvex example in one dimension, we will employ

$$\varphi : \mathbb{R} \mapsto \mathbb{R}, \qquad \varphi(x) = |\sin(x)| \ . \tag{13}$$

Example 2.5 (Half pipe). The function

$$\begin{aligned} \varphi : \mathbb{R}^2 \mapsto \mathbb{R}, \qquad \varphi(x_1, x_2) &= \max(x_2^2 - \max(x_1, 0), 0) \\ &= \begin{cases} x_2^2 & \text{if } x_1 \le 0 \\ x_2^2 - x_1 & \text{if } 0 \le x_1 \le x_2^2 \\ 0 & \text{if } 0 \le x_2^2 \le x_1 \end{cases} , \end{aligned} \tag{14}$$

is also nonconvex as illustrated on the left hand side of Fig. 2.

Example 2.6 (Gradient cube). Here, we consider the gradient cube example as introduced in [7] for $n = 2$ defined by

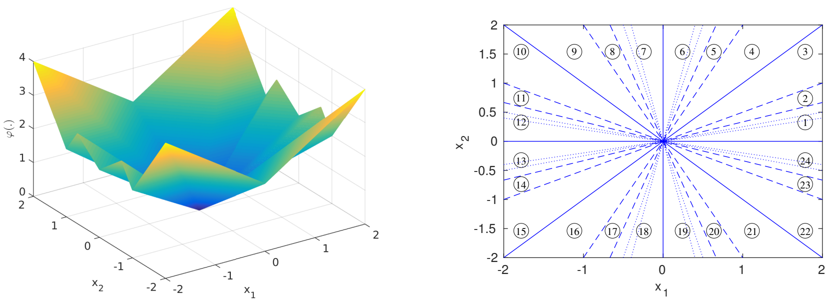
$$\begin{aligned} \varphi : \mathbb{R}^2 \mapsto \mathbb{R}, \quad \varphi(x_1, x_2) &= |x_2 - |x_1|| + \varepsilon |x_1|, \\ &= \begin{cases} \varphi_1(x_1, x_2) = x_2 - x_1 + \varepsilon x_1 & \text{if } x_2 \ge x_1 \ge 0 \\ \varphi_2(x_1, x_2) = x_2 + x_1 - \varepsilon x_1 & \text{if } x_2 \ge -x_1, x_1 < 0 \\ \varphi_3(x_1, x_2) = -x_2 - x_1 - \varepsilon x_1 & \text{if } x_2 < -x_1, x_1 < 0 \\ \varphi_4(x_1, x_2) = -x_2 + x_1 + \varepsilon x_1 & \text{if } x_1 > x_2, x_1 \ge 0 \end{cases} . \end{aligned} \tag{15}$$

This function is illustrated on the right hand side of Fig. 2.

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

$$\begin{aligned} \varphi : \mathbb{R}^2 \mapsto \mathbb{R}, \quad \varphi(x) &= \sum_{i=0}^{q} (y_{2i}(x) + \varepsilon y_{2i+1}(x)) \quad \text{with} \\ y_0(x) &= |x_1| + |x_2|, \quad y_1(x) = |x_1 + x_2| + |x_1 - x_2| \\ y_i(x) &= |x_1 + i\, x_2| + |x_1 - i\, x_2| + |x_1 + x_2/i| + |x_1 - x_2/i|, \ i > 2 \ , \end{aligned} \tag{16}$$



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

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

FIGURE 3. *Function of Eq.* (16) *with* $\varepsilon = -0.5$

which is illustrated on the left hand side of Fig. 3 for $q = 1$. Hence, φ has $s = 8q + 4$ switching variables yielding 2^{8q+4} definite signature vectors. The solid lines on the right hand side of Fig. 3 represent the 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 that have corresponding 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 Eq. (7) satisfy the classical constraint qualifications LICQ or MFCQ in some neighborhood of a given point $\mathring{x}$ with signature $\mathring{\sigma} = \sigma(\mathring{x})$. 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 $\rho > 0$ the intersection $B_\rho \cap S_\sigma$ can only be nonempty if

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

This partial ordering of the signature vectors was already used in [9]. Like in the piecewise linear case we can find that the closure $\bar{S}_\sigma$ of any S_σ is contained in the *extended closure*

$$\hat{S}_\sigma \equiv \{x \in \mathbb{R}^n : \sigma \succeq \sigma(x)\} \supset \bar{S}_\sigma \ . \tag{17}$$

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

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

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 constraints as

$$\hat{S}_\sigma \equiv \{x \in \mathbb{R}^n \;:\; \sigma_i\, z_i^\sigma(x) \;\geq 0 \quad \text{for } i = 1\ldots s\} \; . \tag{18}$$

As observed in [7, Sec. 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*

$$\min f_\sigma(x) \equiv f(x, \Sigma z^\sigma(x)) \quad s.t. \quad x \in \hat{S}_\sigma \quad \text{with} \quad 0 \notin \sigma \succeq \mathring{\sigma} \; . \tag{19}$$

It is natural to look at constraint qualifications for these problems, which will be useful for the regularity analysis in the Section 3. For any such definite σ the constraints that are active at $\mathring{x}$ have the same indices $i \in \mathring{\alpha} = \alpha(\mathring{x})$, but the corresponding constraints $\sigma_i\, z_i^\sigma(x) \geq 0$ are not the same since σ differs. The Jacobian of all constraints is given according to Eq. (9) by

$$\Sigma\nabla z^\sigma \;=\; \Sigma(I - L\Sigma)^{-1}Z \;=\; (\Sigma - L)^{-1}Z \in \mathbb{R}^{s\times n} \; , \tag{20}$$

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

Lemma 2.8 (Jacobian of active constraints). *Consider for a definite signature vector $\sigma \in \{-1, 1\}^s$ the branch problem (19). For $\mathring{x} \in \mathbb{R}^n$, the Jacobian of the constraints that are active at $\mathring{x}$ is given by*

$$J_\sigma \equiv (\sigma_i \nabla z_i^\sigma)_{i\in\mathring{\alpha}} = \check{\Sigma}(I - \check{L}\check{\Sigma})^{-1}\check{Z} \;=\; (\check{\Sigma} - \check{L})^{-1}\check{Z} \in \mathbb{R}^{|\mathring{\alpha}|\times n} \tag{21}$$

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

Proof. Since $\sigma \succeq \mathring{\sigma}$, we have

$$\Sigma = \mathring{\Sigma} + \Gamma \quad \text{with} \quad \mathring{\Sigma}\,\Gamma = 0$$

for a diagonal matrix Γ with $\mathbf{diag}(\Gamma) \in \{-1, 0, 1\}^s$. This yields

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

Defining

$$\mathring{L} \equiv (I - L\mathring{\Sigma})^{-1}L \quad \text{and} \quad \mathring{Z} \equiv (I - L\mathring{\Sigma})^{-1}Z \; , \tag{22}$$

and $\mathring{P} \equiv |\Gamma|$ to zero out the inactive constraints, 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 next to last identity follows from the Neumann series. Extracting the submatrices

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

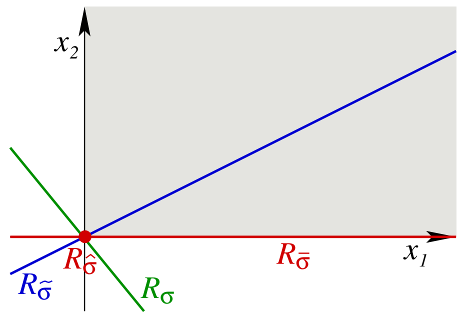


FIGURE 4. *Four different scenarios for MFCQ with $n = 1$ and $|\mathring{\alpha}| = 2$*

one obtains for the Jacobian of the active constraints the reduced identity

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

as stated in the assertion. □

This relation was also derived in [7] 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 2.9 (Uniformity of rank and nullspace). *The active Jacobian J_σ has for all $\sigma \succeq \mathring{\sigma}$ the same rank $r \le \min(|\mathring{\alpha}|, n)$ and the same nullspace 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 Jacobian J_σ have full rank $r = |\mathring{\alpha}| \le n$ if and only if the $|\mathring{\alpha}| \times n$ matrix $\check{Z}$ has full rank $\mathring{\alpha} \le n$. Hence, at $\mathring{x}$ either all branch problems satisfy LICQ or none of them. Otherwise, if the columns of $\check{Z}$ are linearly independent such that $r = n < |\mathring{\alpha}|$ then the nullspace 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 Mangasarin-Fromovitz-Constraint-Qualification [14] 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 $J_\sigma v \ge 0$ 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 Fig. 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

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

since these are the constraints dual to the original MFCQ conditions

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

The convex combination (25) implies for any objective the Fritz John condition

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

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

MFCQ on example. The following simple example shows that MFCQ may hold for one definite σ but not for another. Consider the localized case

(26) $$z_1 = x_1, \quad z_2 = x_2^2 - \tfrac{1}{2}(x_1 + |z_1|), \quad \mathring{x} = 0 \in \mathbb{R}^2 \; ,$$

i.e., $n = 2, s = 2$, and

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

Then we have $\mathring{\sigma} = (0,0)^\top$ and for $\sigma \in \{-1,1\}^2$

$$J_\sigma = \begin{bmatrix} \sigma_1 & 0 \\ \frac{1}{2} & \sigma_2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & 0 \end{bmatrix} = \begin{bmatrix} \sigma_1 & 0 \\ -\frac{1}{2}\sigma_1\sigma_2 & \sigma_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\frac{1}{2} & 0 \end{bmatrix} = \begin{bmatrix} \sigma_1 & 0 \\ -\frac{1}{2}\sigma_2(\sigma_1 + 1) & 0 \end{bmatrix} \; .$$

Hence, we get 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 corresponding 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 to that 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 $v_1 > 0$. It follows that 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 be also 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 Cor. 2.9, one has

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

and all Jacobians have the same nullspace 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 Eq. (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 an extension of the definition of LIKQ in [7]:

Definition 2.10 (LIKQ and MFKQ). *For $\varphi \in C^1_{abs}(\overline{\mathcal{D}})$ according to Definition 2.1 consider the reduced quantities $\check{Z}$ and $\check{L}$ as defined in Lemma 2.8 at a point $\mathring{x}$. Then we say that LIKQ is satisfied 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 $\sigma \succeq \mathring{\sigma}$ the vector inequality $J_\sigma v > 0$ is solvable for some $v \in \mathbb{R}^n$ unless 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 [7] we showed that the two nonsmooth versions of the chained Rosenbrock function suggested according to [10] by Nesterov satisfy LIKQ everywhere, and that it holds for their *natural* abs-normal representation, i.e., without any modification or preprocessing. That allowed 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 can not be tested in polynomial time.

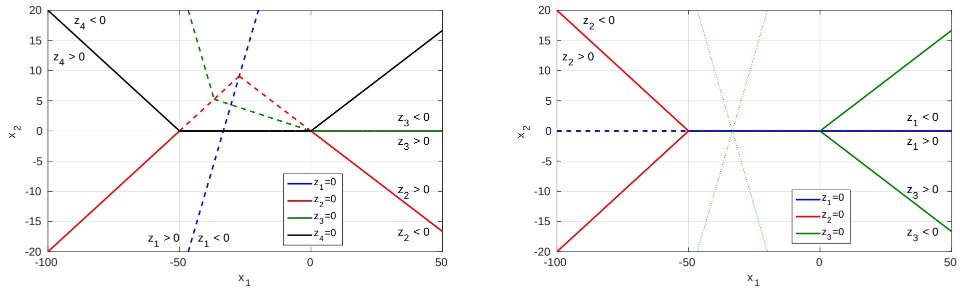
Lemma 2.11 (Kink qualifications for example problems). *With respect to the kink qualifications LIKQ and MFKQ as introduced above, one obtains the following results at $\mathring{x} = 0 \in \mathbb{R}$ and $\mathring{x} = 0 \in \mathbb{R}^2$, respectively:*

	HUL					
	Eq. (11)	*Eq.* (12)	*abs-sin*	*half pipe*	*gradient cube*	*lemon squeezer*
LIKQ	⦸	✓	✓	⦸	✓	⦸
MFKQ	✓	✓	✓	⦸	✓	✓

Proof. For the HUL example and the representation given by Eq. (11), 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 Fig. 5a. In the neighborhood of $\mathring{x}$, there are 6 polyhedra, i.e., less than 2^s, and they are all open. It also follows immediately from Fig. 5a 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.

For the alternative representation given by Eq. (12), one has for $\mathring{x} = 0 \in \mathbb{R}^2$ that $s = 3$ and $\mathring{\alpha} = \{1,3\}$ with

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



(a) Nondifferentiable points for Eq. (11)   (b) Nondifferentiable points for Eq. (12)

FIGURE 5. *Kink structure for different formulations of Eq.* (10)

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

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 also MFKQ.

For the half pipe example, the representation

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

defines the switching variables

$$z_1 = x_1 \quad \text{and} \quad z_2 = x_2^2 - \tfrac{1}{2}(x_1 + |z_1|)$$

as already considered in Eq. (26), which means that at $\mathring{x} = 0$

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

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

For the gradient cube example, Eq. (15) yields the switching variables

$$z_1 = x_1 \quad \text{and} \quad z_2 = x_2 - |z_1| \,,$$

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 also MFKQ.

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

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

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

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

and obviously LIKQ can not hold a $\mathring{x}$. Since $L = 0 \in \mathbb{R}^{s\times s}$, 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

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

Therefore, either one can find for a given Σ values of v_1 and v_2 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 $q \in \mathbb{N}$. □

As can be seen from the HUL example, the representation of a function may have a considerable influence on the kink qualifications. To avoid that LIKQ is violated one should try to introduce as few kinks as possible.

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 in the vicinity of 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 like multiphase equilibria of mixed fluids lack of convexity may lead to the instability of single phase equilibria as discussed in the introduction. Therefore we will examine conditions for convexity in the vicinity of a given point, irrespective of whether the point is even stationary or not, later in this section in more detail. Such a verification of convexity is of interest not only for optimality but for example also for computer graphics. For a different class of piecewise defined functions, such convexity tests were defined for example in [1, 3]. Like for optimality, see [7] and [8], 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 subdifferentials according to [17, 20]:

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 $d\varphi(x)(.) : \mathbb{R}^n \mapsto \mathbb{R}$ 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

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

This allows to define the set of (general) subgradients as the outer semi-continuous envelop

$$\partial^M \varphi(x) = \{ g \in \mathbb{R}^n \,\Big|\, \exists \{x_k\}_{k \in \mathbb{N}} : x_k \to x, \varphi(x_k) \to \varphi(x), \quad g_k \in \widehat{\partial}^M \varphi(x_k), g_k \to g \Big\} . \tag{27}$$

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

Since we consider C^d_{abs} functions $\varphi(.)$ throughout the whole paper, all $\varphi(.)$ are lower semi-continuous 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 $\varphi(.)$ in a given point x.

Another widely used derivative concept is based on limits of classical gradients. For this purpose, one exploits Rademacher's theorem that guarantees that Lipschitz continuous functions like the C^d_{abs} functions considered in this paper are almost everywhere differentiable. Let $D_\varphi \subset \overline{\mathcal{D}}$ denote the set where the C^d_{abs} 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:

Definition 3.2 (Limiting gradients and Clark 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*

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

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

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

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

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

These conical gradients and their generalization conical Jacobians are for example considered in [11, 12]. 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 Eq. (6) such that the tangent cone of the coincidence set $\{x \in \mathbb{R}^n \,|\, \varphi(x) = \varphi_\sigma(x)\}$ at $\mathring{x}$ has a nonempty interior, see [4].

Example 3.4 (Generalized derivatives for the abs-sin example). The function introduced in Eq. (13) 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 linearization of $\varphi(.)$ at $\mathring{x} = 0$ is given by $\Delta\varphi(0;\Delta x) = |\Delta x|$, which is a convex function despite the fact that $\varphi(.)$ itself is not convex at $\mathring{x} = 0$. It follows from this linearization 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 Eq. (14), one can check that $\varphi(.)$ 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

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

Hence, in this case $\widehat{\partial}^M\varphi(0)$ is a proper subset of $\partial^M\varphi(0)$, such that $\varphi(.)$ 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

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

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 $\varphi_i(.), 1 \le i \le 4$, i.e.,

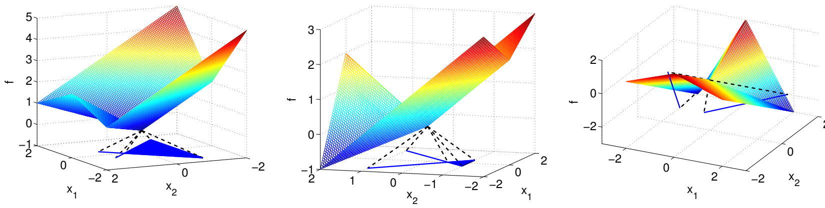
$$g_1 = (-1+\varepsilon,1), \; g_2 = (1-\varepsilon,1), \; g_3 = (-1-\varepsilon,-1), \text{ and } g_4 = (1+\varepsilon,-1) \, .$$

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

$$\liminf_{x \to \mathring{x}, x \neq \mathring{x}} \frac{\varphi(x) - \varphi(\mathring{x}) - \langle g, x - \mathring{x}\rangle}{\|x - \mathring{x}\|} \ge 0 \, . \tag{28}$$

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

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



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

FIGURE 6. *Mordukovich subdifferentials for Gradient Cube problem*

since $\varphi(.)$ is a convex function for $\varepsilon \geq 1$ and then the second equality is given by Prop. 8.12 of [20]. For $\varepsilon \in [-1, 1)$ one has to examine the situation more closely, since $\varphi(.)$ is no longer convex. As can be seen from Fig. 2, in this case the gradients of $\varphi_3(.)$ and $\varphi_4(.)$ define supporting hyperplanes for $\varphi(.)$. 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 Eq. (28) holds if $\varepsilon \in [-1, 1]$. It follows that

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

whereas Eq. (27) yields with $\widehat{\partial}^M \varphi(0) \subset \partial^M \varphi(0)$ according to [20, Theo. 8.6]

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

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

$$\widehat{\partial}^M \varphi(0) = \emptyset \quad \text{and} \quad \partial^M \varphi(0) = \operatorname{conv}\{\mathrm{g}_1, \mathrm{g}_4\} \cup \operatorname{conv}\{\mathrm{g}_2, \mathrm{g}_3\} \ .$$

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

$$\partial^K \varphi(0) = \partial^L \varphi(0) = \{g_1, g_2, g_3, g_4\} \quad \text{and} \quad \partial^C \varphi(0) = \operatorname{conv}\{\mathrm{g}_1, \mathrm{g}_2, \mathrm{g}_3, \mathrm{g}_4\}$$

for all values of ε.

Example 3.7 (Generalized derivatives for the lemon squeezer example). The corresponding $\varphi(.)$ is not differentiable at $\mathring{x} = (0, 0)^\top \in \mathbb{R}^2$. For notational simplicity, we only consider the case $q = 1$ here. Possible candidates for a regular subgradient are given by the gradients of the linear selection functions $\varphi_i(.), 1 \leq i \leq 24$, 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 $\varepsilon \geq 0$, $\varphi(.)$ 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 also the right hand side of Fig. 3. 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. This yields for $0 > \varepsilon > -9/13$

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

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 $\varphi_6(.)$ and $\varphi_7(.)$ as well as $\varphi_{18}(.)$ and $\varphi_{19}(.)$ attain negative values. Hence, no supporting hyperplane exists at $\mathring{x} = 0$ yielding

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

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

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

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 examine these relations now more closely.

Relations Between Generalized Derivatives. Exploiting the C^d_{abs} structure, one obtains the following results:

Proposition 3.8 (Limiting, Mordukovich, and Clark subdifferentials). *For the C^d_{abs} function $\varphi : \overline{\mathcal{D}} \to \mathbb{R}$, $\overline{\mathcal{D}} \subset \mathbb{R}^n$, 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 $\varphi(.)$ is regular in $x \in \mathbb{R}^n$ if and only if

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

Proof. It follows from Def. 3.2, that for an element $g \in \partial^L \varphi(x)$ there exists a sequence $\{x_k\}_{k\in\mathbb{N}}$ such that $\varphi(.)$ is differentiable at x_k and $\nabla\varphi(x_k) \to g$. According to [20, Ex. 8.8a], then one 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 $\varphi(.)$ is smooth in a neighborhood of x, then one has according to [20, Ex. 8.8b]

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

If $\varphi(.)$ 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 $\varphi(x) = \varphi_{\sigma_1}(x) = \varphi_{\sigma_2}(x)$ 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 Eq. (28) 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 also Exam. 3.5 and Exam. 3.4, respectively, for an illustration. Therefore, it follows that

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

If more switching variables vanish at x, if 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 Eq. (28), also convex combinations of these gradients or of a proper subset of these gradients are elements of $\partial^M \varphi(x)$, see Exam. 3.6 with $\varepsilon < 1$ for an illustration. This yields

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

For proving the second assertion, first assume that $\varphi(.)$ is regular in x. Then one has that $\widehat{\partial}^M \varphi(x) = \partial^M \varphi(x)$ is convex, see [20, Theo. 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

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

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

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

Then, if follows from the C^d_{abs} structure of the functions considered here and the definition of the regular Mordukovich subgradient that $\partial^M \varphi(x) = \widehat{\partial}^M \varphi(x)$ must hold. Hence, $\varphi(.)$ 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

$$\partial^K \varphi(\mathring{x}) = \partial^L \varphi(\mathring{x})\ . \tag{29}$$

Proof. The first inclusion was already shown in [4, 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 [4, Cor. 1], which states that the gradient of φ_σ must then be conical. ∎

Hence, since MFKQ holds at $\mathring{x} = 0$ for the Examples 2.4 and 2.6, the equality of the conical and limiting gradients at $\mathring{x} = 0$ derived in the 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 2.5. 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 linearization Eq. (1) given above, 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 piecewise 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 $\varphi(.)$ is called first order convex in some ball about $\mathring{x}$ if its linearization, i.e., first order model, in $\mathring{x}$ is convex. For this new concept, one has

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 $\varphi(.)$ is regular at $\mathring{x}$. Furthermore, if MFKQ holds at $\mathring{x} \in \mathbb{R}^n$, then $\varphi(.)$ is first order convex in some ball about $\mathring{x}$ if and only if $\varphi(.)$ is regular at $\mathring{x}$.*

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

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

Furthermore, one obtains from Prop. 3.9 that

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

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

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

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

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

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

yielding

$$\widehat{\partial}^M \varphi(\mathring{x}) = \partial^M \varphi(\mathring{x})$$

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

As can be seen, FOC implies the inclusion $\partial^L \varphi(x) \subset \partial^M \varphi(x)$. One might assume that this inclusion on the other hand also implies FOC and therefore regularity. However, Exam. 2.6 for $\varepsilon \in [-1, 1)$ shows that this is not the case.

The findings so far can be summarized in the following way. For a C^d_{abs} function $\varphi(.)$, we have

- in the general case

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

  as well as

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

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

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

As an immediate corollary the implication chain CON $\implies$ REG $\implies$ FOC, that is, convexity of C^d_{abs} functions is inherited by their piecewise linearizations, holds as we already 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 piecewise linearization $\Delta\varphi(\mathring{x}; .) : \mathbb{R}^n \to \mathbb{R}$ is convex for Δx in some ball about the origin, i.e., if $\varphi(.)$ 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 $\tau \in (0,1)$

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

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. ☐

One has to note that the convexity of $\Delta\varphi(\mathring{x}; .)$ is only necessary but not sufficient for the convexity of $\varphi(.)$ as illustrated by Exam. 2.4.

Now the question is of course, how we can determine whether the piecewise 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., $\mathring{\alpha} = \alpha(\mathring{x}) = \{1, \dots s\}$. Furthermore, assume that LIKQ holds at $\mathring{x}$ so that $s \leq n$. Then the piecewise linearization of φ at $\mathring{x}$ is locally convex if and only if componentwise*

$$b^\top (I - DL)^{-1} \;\geq\; 0 \quad \textit{if} \quad |D| \leq 1 \, , \tag{30}$$

where D ranges over all diagonal matrices. In the nonlocalized case, L and b in Eq. (30) must be replaced by $\check{L}$ as defined in Eq. (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 $\Sigma = \mathbf{diag}(\sigma)$. Then P_σ is a convex cone in which the function $-\Sigma z^\sigma$ has by Eq. (9) the Jacobian $(I - L\Sigma)^{-1} Z$ yielding

$$N \;\equiv\; \Sigma \nabla z^\sigma \;=\; \Sigma(I - L\Sigma)^{-1} Z \;=\; (\Sigma - L)^{-1} Z \in \mathbb{R}^{s \times n} \, . \tag{31}$$

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

$$\sigma^i_j \;=\; \begin{cases} \sigma_j & \text{if } \; j \neq i \\ -\sigma_i & \text{if } \; j = i \end{cases} \, . \tag{32}$$

Therefore, the corresponding diagonal matrices satisfy

$$\Sigma^i = \mathbf{diag}(\sigma^i) \; = \; \Sigma - 2\sigma_i e_i e_i^\top \; . \tag{33}$$

By Eq. (6) the gradients of φ in P_σ and P_{σ^i} are given by

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

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

$$g_{\sigma^i} - g_\sigma = -2b^\top(\Sigma - L)^{-1}e_i\sigma_i e_i^\top(\Sigma - L)^{-1}Z = 2b^\top(\Sigma - L)^{-1}\sigma_i e_i \nu^i \; . \tag{34}$$

Hence, we see that convexity requires $b^\top(\Sigma - L)^{-1}\sigma_i e_i \geq 0$ for all i and thus $b^\top(\Sigma - L)^{-1}\Sigma \geq 0$ componentwise. Moreover this vector inequality must hold for all definite signatures $\sigma \in \{-1, 1\}^s$. Since $b^\top(D-L)^{-1} \geq 0$ 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 all we still have to show is that it is also sufficient. Take again 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., bend upwards so that the whole piecewise linear function is convex. By continuity that follows then 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 with its neighbor cones $\mathcal{P}_{\sigma^i}$ with the signature matrices $\Sigma^i, i \in \mathring{\alpha}(\mathring{x})$, where Σ^i is defined as in Eq. (33). Exactly along 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. Complexity Analysis. We proved in [7] that when LIKQ holds then stationarity and normal growth, i.e., first order optimality, can be tested in polynomial time. So far, we did not succeed in deriving a test that is polynomial in time for testing stationarity under MFQK. Even testing for MFQK at $\mathring{x}$ seems to be expensive it that it is not polynomial in time since the existence of a vector v with $J_\sigma v > 0$ for all $\sigma \succeq \mathring{\sigma}$ has to be verified.

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

DEFINITION 4.1 (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*

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

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 $\mathring{x}$. Using again $\Sigma = \mathbf{diag}(\sigma)$, the convexity test can be rewritten as

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

Its complement is given by

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

LEMMA 4.2. *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 4.3 (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 4.4. *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 into 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 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 for each clause $\psi_i(x)$ involving three variables x_{i1}, x_{i2}, x_{i3} corresponding $\sigma_{ij}, j = 0, \dots, 3$ 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 $\tilde{L} = \mathbf{diag}((L_i)_{i=1,\dots,m})$ in the following way. The clause $\psi_i(x)$ involves three variables x_{i1}, x_{i2}, x_{i3}. Define $\Sigma_i = \mathbf{diag}(\sigma_{i0}, \sigma_{i1}, \sigma_{i2}, \sigma_{i3})$, and $b_i = (1,1,1,1)^\top$. Then, one obtains with

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

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

Setting

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

one obtains already one important part of an 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 Eq. (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, one might have for one $\Sigma_\psi = \mathbf{diag}(\sigma_{10}, \dots, \sigma_{m3})$ 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 in addition 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 for each component of x how often it 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 will 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 the next appearance of x_l, i.e., the $(k+1)$th 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

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

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 of one. 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 Exam. 4.5. 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 already 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 $(k+1)$th 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. Setting now $b = (0_M, 1_{\hat{m}})^\top)$ the coupling conditions yield 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 condition

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

requiring that $\sigma_{ij_i} = \sigma_{\hat{\imath}j_{\hat{\imath}}}$. If x_l occurs in ψ_i and its negation in $\psi_{\hat{\imath}}$ or if the negation of x_l occurs in ψ_i and x_l 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 that way one can reconstruct from all Σ that fulfill 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 serves to illustrate this rather involved reduction from a instance of TAUTOLOGY to an instance of CONV:

Example 4.5 (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

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

Furthermore, using the polynomial reduction introduced in the proof of Theo. 4.4, one has that x_1 is represented by σ_{11}, σ_{21}, and σ_{31}, x_2 by σ_{12} and σ_{32}, x_3 is represented by σ_{13}, σ_{22}, and σ_{33}, and x_4 by σ_{23}. The coupling matrix is given by

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

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 Eq. (36). Using now $b = (0_M, 1_{\hat{m}})^\top$, one obtains from $b^\top(\Sigma - L)^{-1}\Sigma$ that

$$\begin{array}{lll}
\sigma_{11} - \sigma_{21} \geq 0 & -\sigma_{11} + \sigma_{21} \geq 0 & \Rightarrow \sigma_{11} = \sigma_{21} \\
\sigma_{21} - \sigma_{31} \geq 0 & -\sigma_{21} + \sigma_{31} \geq 0 & \Rightarrow \sigma_{21} = \sigma_{31} \\
\sigma_{12} + \sigma_{32} \geq 0 & -\sigma_{12} - \sigma_{32} \geq 0 & \Rightarrow \sigma_{12} = -\sigma_{32} \\
\sigma_{13} + \sigma_{22} \geq 0 & -\sigma_{13} - \sigma_{22} \geq 0 & \Rightarrow \sigma_{13} = -\sigma_{22} \\
\sigma_{22} + \sigma_{33} \geq 0 & -\sigma_{22} - \sigma_{33} \geq 0 & \Rightarrow \sigma_{22} = -\sigma_{33}
\end{array}$$

and therefore consistent values for the components of x.

5. Summary and Outlook. In this paper, we first generalize the kink qualifications for C^d_{abs} functions under LIKQ to the more general case of the new kink qualification MFKQ. In contrast to LIKQ, it is so far not possible to verify MFKQ in polynomial time. Next we studied convexity conditions for C^d_{abs} functions and their piecewise linearization under LIKQ and MFKQ. This includes also an analysis of the relation of different derivative concepts and regularity. Here, the main result is that under MFKQ first order convexity and regularity are equivalent. Furthermore, we proved that testing for convexity co-NP complete even under LICQ and thus certainly under MFKQ. Thus we conclude that on the class of C^d_{abs} functions regularity is a rather theoretical, nonconstructive concept. The complexity analysis for the test whether MFKQ holds at a given point is still open and subject of future research. Generally speaking, it seems that the combinatorial aspect and its computational complexity deserves more attention in nonsmooth analysis.

REFERENCES

[1] H. Bauschke, Y. Lucet, and H. Phan. On the convexity of piecewise-defined functions. *ESAIM: Control, Optimisation and Calculus of Variations*, 22(3):728–742, 2016.

[2] L. Eneya, T. Bosse, and A. Griewank. A method for pointwise evaluation of polyconvex envelopes. *Afrika Matematika*, 24:1–24, 2013.

[3] B. Gardiner and Y. Lucet. Computing the conjugate of convex piecewise linear-quadratic bivariate functions. *Mathematical Programming*, 139(1–2):161–184, 2013.

[4] A. Griewank. On stable piecewise linearization and generalized algorithmic differentiation. *Optimization Methods and Software*, 28(6):1139–1178, 2013.

[5] A. Griewank, J.-U. Bernt, M. Randons, and T. Streubel. Solving piecewise linear equations in abs-normal form. *Lin. Algebra and its Appl.*, 471:500–530, 2015.

[6] A. Griewank and A. Walther. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. SIAM, 2008.

[7] A. Griewank and A. Walther. First and second order optimality conditions for piecewise smooth objective functions. *Optimization Methods and Software*, 31(5):904–930, 2016.

[8] A. Griewank and A. Walther. Relaxed kink qualifications in optimality conditions for piecewise smooth objective functions. Technical report, Yachay Tech University, 2017.

[9] A. Griewank, A. Walther, S. Fiege, and T. Bosse. On Lipschitz optimization based on gray-box piecewise linearization. *Mathematical Programming Series A*, 158(1-2):383–415, 2016.

[10] M. Gürbüzbalaban and M.L. Overton. On Nesterov's nonsmooth Chebyshev-Rosenbrock functions. *Nonlinear Anal: Theory, Methods & Appl.*, 75(3):1282–1289, 2012.

[11] K.A. Khan and P.I. Barton. Evaluating an element of the Clarke generalized Jacobian of a composite piecewise differentiable function. *ACM Transactions on Mathematical Software*, 39(4):28, 2013.

[12] K.A. Khan and P.I. Barton. A vector forward mode of automatic differentiation for generalized derivative evaluation. *Optimization Methods and Software*, 30(6):1185–1212, 2015.

[13] A.S. Lewis. Active sets, nonsmoothness, and sensitivity. *SIAM Journal on Optimization*, 13(3):702–725, 2002.

[14] O.L. Mangasarian and S. Fromovitz. The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *Journal of Mathematical Analysis and Applications*, 17:37–47, 1967.

[15] M. Michelsen. The isothermal flash problem. part i. stability. *Fluid Phase Equilibria*, 9(1):1–19, 1982.

[16] R. Mifflin and C. Sagastizábal. Primal-dual gradient structured functions: second-order results; links to epi-derivatives and partly smooth functions. *SIAM Journal on Optimization*, 13(4):1174–1194, 2003.

[17] B. Mordukhovich. *Variational Analysis and Generalized Differentiation I: Basic Theory.* Springer, 2006.

[18] N.R. Nagarajan, A.S. 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*, 62:191–210, 1991.

[19] Y. Nesterov. Lexicographic differentiation of nonsmooth functions. *Mathematical Programming Series A*, 104(2-3):669–700, 2005.

[20] R.T. Rockafellar and R.J.-B. Wets. *Variational analysis.* Springer, 1998.

[21] S. Scholtes. *Introduction to Piecewise Differentiable Functions.* Springer, 2012.