RELAXING KINK QUALIFICATIONS AND PROVING CONVERGENCE RATES IN PIECEWISE SMOOTH OPTIMIZATION

ANDREAS GRIEWANK[1] AND ANDREA WALTHER[2]

Abstract. In the paper [9] we derived first order (KKT) and second order (SSC) optimality conditions for functions defined by evaluation programs involving smooth elementals and absolute values. In that analysis, a key assumption on the local piecewise linearization was the Linear Independence Kink Qualification (LIKQ), a generalization of the Linear Independence Constraint Qualification (LICQ) known from smooth nonlinear optimization. We show here first that under LIKQ and SSC with strict complementarity, the natural algorithm of successive piecewise linear optimization with a proximal term (SPLOP) achieves a linear rate of convergence. A version of SPLOP called LiPsMin has already been implemented and tested in [10, 4]. Secondly, we observe that, even without any kink qualifications, local optimality of the nonlinear objective always requires local optimality of its piecewise linearization, and strict minimality of the latter is in fact equivalent to sharp minimality of the former. Moreover, we show that SPLOP will converge quadratically to such sharp minimizers, where the function exhibits linear growth. These results are independent of the particular function representation, and allow in particular duplications of switching variables and other intermediates. Furthermore, we derive for nonsharp minimizers necessary and sufficient optimality conditions without any kink qualification. Finally, we present numerical results to illustrate the obtained convergence results.

Keywords: Abs-Normal Form, Sharp Minimizer, Linear Independence Kink Qualification (LIKQ), Karush Kuhn Tucker (KKT), Second Order Sufficiency Condition (SSC), Strict Complementary, Linear Convergence, Quadratic Convergence.

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 [8], 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.

Piecewise Linearization. In contrast to the classical smooth case, where all individual instructions are assumed to be differentiable near the evaluation points of interest, we allow now also piecewise linear elemental functions like abs, min, and max as part of the mix. By this means, we arrive at a subclass of piecewise smooth functions [16] that can still be analyzed by slight extensions of algorithmic 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 [6], 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

[1]School of Mathematical Sciences and Information Technology, Yachay Tech, Urcuquí, Imbabura, Ecuador

[2]Institut für Mathematik, Universität Paderborn, Paderborn, Germany

not differentiable with respect to Δx in a neighborhood of the origin, in other words we do not have strong Bouligand differentiability as defined in [16].

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 $\Delta x = 0$. 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, where $\Delta\varphi(x;\Delta x)$ is represented by several real matrices and vectors, which can be analyzed by various linear algebra procedures.

Optimality conditions. In [9] it was shown how for a scalar-valued function $\varphi(x)$ this information can be used to characterize local minimality (MIN) in terms of generalized Karush-Kuhn-Tucker (KKT) and Positive Curvature Conditions (POSC, see Theorem 4.3). This optimality analysis was based on a generalization of the Linear Independence Constraint Qualification (LICQ) called Linear Independence Kink Qualification (LIKQ), which is generic [7], 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 avoid any constraint qualification or relax LIKQ somewhat, 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) in our nonsmooth setting. The corresponding, more general optimality conditions are developed to obtain the implications

$$MIN \underset{(POSC)}{\overset{\Longrightarrow}{\Longleftarrow}} FOM$$

where the upper implication is fairly obvious and has been proven formally in [4, Lemma 1]. In other words we note that without any additional geometric assumptions (like constraint or kink qualifications) the (piecewise) linearized problem must have a local minimizer, where the underlying function has one. It is interesting to note that the implication MIN $\Rightarrow$ FOM does not always hold for smooth Nonlinear Optimization Problems (NLOPs) and their linearizations, namely the corresponding Linear Optimization Problems (LOPs) or Quadratic Optimization Problems (QOPs). This is illustrated by the following example taken from [11, Chap. 9]:

Example 1.1. Consider the smooth optimization problem

$$\varphi : \mathbb{R}^2 \mapsto \mathbb{R}, \quad \varphi(x) = x_1 + x_2, \quad \text{such that} \quad x_2 \geq 0,\ x_2 \leq \tfrac{1}{3}x_1^3 \,.$$

Then, $x_* = 0 \in \mathbb{R}^2$ is a minimizer of the original nonlinear problem but not of the corresponding linearization at x_* given by

$$\varphi(x) = x_1 + x_2 \quad \text{such that} \quad x_2 \geq 0,\ x_2 \leq 0 \,,$$

yielding $\mathbb{R} \times \{0\}$ as feasible set. Therefore, the linearized problem is unbounded and has no local minima. This is possible because not even the Abadie constraint

qualification [1] is satisfied so that the tangent cone of the nonlinear problem at its minimizer does not coincide with the tangent cone of linearization at that point.

One might wonder what happens if we convert the constrained smooth problem to an unconstrained nonsmooth one by adding the penalty terms

$$(1+\rho^2)[\max(0,-x_2)+\max(0,x_2-\tfrac{1}{3}x_1^3)]$$

with some large ρ to the original objective. The particular form of the penalty parameter was chosen to simplify the expressions in the following lemma.

LEMMA 1.2. *The composite objective*

$$\varphi_\rho(x) = x_1 + x_2 + (1+\rho^2)\left[\max(0,-x_2)+\max(0,x_2-\tfrac{1}{3}x_1^3)\right]$$

attains its global minimum at

$$\mathbf{x}_\rho \;=\; \left(-\tfrac{1}{\rho}, -\tfrac{1}{3\rho^3}\right) \quad \textit{with} \quad \varphi_\rho(\mathbf{x}_\rho) = -\tfrac{2}{3\rho}\;,$$

where the piecewise linearization is given by

$$\Delta\varphi_\rho(\mathbf{x}_\rho;\Delta x) = \Delta x_1 + \Delta x_2 + (1+\rho^2)\left[\min(0,-\Delta x_2) + \tfrac{1}{3\rho^3} + \max(0,\Delta x_2 + \tfrac{\Delta x_1}{\rho^2})\right]\;,$$

which has an isolated minimizer at $\Delta x = (\Delta x_1, \Delta x_2) = (0,0)$.

Proof. Clearly the optimal value of x_1 must be negative and fixing any such x_1 we get a convex piecewise linear function in x_2. Its value for large x_2 is close to $x_1 + x_2 + (1+\rho^2)|x_2|$ so its minimum must be obtained at one of the kinks $x_2 = 0$ and $x_2 = \frac{1}{3}x_1^3$. Given $x_1 < 0$ the directional derivatives of φ_ρ with respect to $x_2 \approx 0$ are 1 for $x_2 < 0$ and $2+\rho^2$ for $x_2 > 0$. Hence the first kink cannot be the minimizer and we must have $x_2 = \frac{1}{3}x_1^3$. Substituting this into the objective we get

$$\varphi_\rho(x) \;=\; x_1 + \tfrac{1}{3}x_1^3 - (1+\rho^2)\tfrac{1}{3}x_1^3 = x_1 - \tfrac{1}{3}\rho^2 x_1^3\;.$$

Now minimizing with respect to x_1 we get the asserted minimizer and the corresponding minimal value. The incremental piecewise linearization is obtained by replacing nonlinear smooth elementals by their tangent and all piecewise smooth elementals by their piecewise linearization as specified in [6]. ∎

The example shows that due to the lack of a constraint qualification the penalty function is not exact and has a minimizer that converges to the minimizer of the original NLOP as ρ tends to infinity.

A brief look at sharp minima. Another immediate consequence of the second order contact described by Eq. (1) is that φ has a sharp minimum at x_* in that

$$\varphi(x) - \varphi(x_*) \;\geq\; c\|x - x_*\| \quad \text{for some } c > 0 \text{ and all } \; x \approx x_*$$

if and only if this is true for its piecewise linearization, i.e.,

$$\Delta\varphi(x_*; x - x_*) \;\geq\; c\|x - x_*\| \quad \text{for some } c > 0 \text{ and all } \; x \approx x_*$$

which is in turn equivalent to $\Delta x = 0$ being an isolated minimizer of $\Delta\varphi(x_*;\Delta x)$, i.e.,

$$\Delta\varphi(x_*;\Delta x) \;>\; 0 \quad \text{for all small} \quad \Delta x \neq 0\;.$$

In our context all strict minimizers are isolated so that the properties of strictness and isolation are for us one and the same.

It should also be noted here that at any given $\mathring{x}$ and sufficiently small Δx the piecewise linearization $\Delta\varphi(\mathring{x};\Delta x)$ is homogeneous and coincides with the so-called Bouligand derivative $\varphi'(\mathring{x};\Delta x)$ of piecewise smooth functions [16]. For larger Δx the piecewise linearization $\Delta(\mathring{x};\Delta x)$ is in general nonhomogeneous. Indicating Sharpness by a leading S and isolation by an leading I, we obtain the equivalences

$$\text{SMIN} \iff \text{SFOM} \iff \text{IFOM}$$

In Proposition 4.1 we will give a necessary and sufficient condition for sharp minimality, whose falsification is polynomial under LIKQ but may require a combinatorial effort otherwise. In the case of first order minimality (FOM) we have a gap between necessary and sufficient conditions, whose falsification is again simple under LIKQ but potentially combinatorial otherwise.

Successive piecewise linear optimization. In view of Eq. (1) a natural algorithmic strategy, to be called SPLOP for Successive Piecewise Linear Optimization with a proximal term, is to successively update

$$x_{k+1} = x_k + \operatorname*{arg\,min}_{\Delta x} \left\{ \Delta\varphi(x_k;\Delta x) + \tfrac{q}{2}\|\Delta x\|^2 \right\} . \tag{2}$$

Here the penalty factor q of the quadratic term is an estimated bound on the discrepancy between φ and its piecewise linearization. This method was originally proposed in [6] and shown to generate a sequence of iterates $\{x_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ whose cluster points are FOM. If the inner problem of minimizing the regularized piecewise linear model is not solved exactly, but we accept increments Δx that are simply Clarke stationary for $\Delta\varphi$, then the cluster points are guaranteed to be also Clarke stationary as shown in [4]. These results apply without any kink qualification and contain no information with regards to the speed of convergence, except that the step lengths are square summable.

Assuming again that x_{k+1} is computed as a global minimizer of the regularized piecewise linear model we obtain for any other point $x_* \in \mathbb{R}^n$

$$\Delta\varphi(x_k;x_{k+1}-x_k) - \Delta\varphi(x_k;x_*-x_k) \;\le\; \tfrac{q}{2}(\|x_k-x_*\|^2 - \|x_{k+1}-x_k\|^2) . \tag{3}$$

This leads immediately to our first rate of convergence result.

PROPOSITION 1.3. *If x_* is a sharp minimizer in a level set where the right hand side of Eq. (1) is bounded by $\frac{\gamma}{2}\|\Delta x\|^2$ then the SPLOP method as described above with $q \ge \gamma$ converges quadratically to x_* from all x_0 in some ball $B_\rho(x_*)$.*

Proof. Starting with the sharpness we derive from Eq. (3)

$$\begin{aligned} c\|x_{k+1}-x_*\| &\le \varphi(x_{k+1}) - \varphi(x_*) = \varphi(x_{k+1}) - \varphi(x_k) - (\varphi(x_*) - \varphi(x_k)) \\ &\le \Delta\varphi(x_k;x_{k+1}-x_k) - \Delta\varphi(x_k;x_*-x_k) \\ &\quad + \tfrac{\gamma}{2}(\|x_{k+1}-x_k\|^2 + \|x_*-x_k\|^2) \\ &\le \tfrac{\gamma-q}{2}\|x_{k+1}-x_k\|^2 + \tfrac{\gamma+q}{2}\|x_k-x_*\|^2 \le\; q\|x_k-x_*\|^2 . \end{aligned}$$

∎

This gratifying result depends critically on the uniform approximation property (1), which is not guaranteed by the positively homogeneous Bouligand derivative $\varphi'(x;\Delta x)$.

The question of how fast SPLOP converges in the nonsharp case is a little harder to answer. Since the location of a nonsharp minimizer depends critically on curvatures, and our method makes no effort to approximate them, we can expect linear convergence at best. Even that is an achievement since subgradient, proximal, and bundle methods generally only obtain a convergence rate close to $1/\sqrt{k}$ or $1/k$ [17]. However, for structured problems and in the strictly convex case a linear rate has been established for some of these methods, e.g., in [3]. We will do the same under the assumption of SSC with strict complementarity under LIKQ. While these combined assumptions are quite strong, they do not require φ itself to be convex, even locally. Essentially, they ensure that eventually all active kinks causing the nonsmoothness at the minimizers are identified so that SPLOP reduces locally to sequential quadratic optimization (SQOP) with equality constraints. Here the Hessian of the Lagrangian is approximated by the multiple of the identity qI.

Paper organization. The paper is organized as follows. In Section 2 we discuss the representation of piecewise smooth functions in abs-normal form. Furthermore, we give six different example functions that will be used to illustrate the concepts and results. We introduce two kink qualifications, LIKQ and the weaker MFKQ. In Section 3 we study convergence orders for SPLOP under LIKQ. In Section 4 we generalize the optimality conditions of [9] to the more general situations, where no kink qualifications hold. Section 5 contains numerical experiments and the paper concludes with a summary and outlook in Section 6.

2. Objectives in Abs-normal Form. We consider the class of objective functions that are defined as compositions of smooth elemental functions and the absolute value function $\mathrm{abs}(x) = |x|$. Hence they may also include $\max(x, y), \min(x, y)$, and the positive part function $\max(0, x)$, which can be easily cast 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 [13] but no longer piecewise smooth.

By successively numbering all arguments of absolute value evaluations as *switching variables* z_i for $i = 1 \ldots s$, we obtain a piecewise smooth representation of $y = \varphi(x)$ in abs-normal form

$$z \;=\; F(x, |z|)\;, \tag{4}$$

$$y \;=\; f(x, |z|), \tag{5}$$

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 $\overline{\mathcal{D}} \times \overline{\mathbb{R}^s_+} \subset \mathbb{R}^{n+s}$. Here, z_j can only influence z_i if $j < i$ so that when interpreting F as a function of $|z|$, its Jacobian with respect to $|z|$ is strictly lower triangular.

Hence, we state the calculation of all switching variables as equality constraints and handle the vector of the absolute values of the switching variables as extra argument of the then smooth target function f. 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 once or twice continuously differentiable. The resulting function class was first considered in [5] and is specified as follows:

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 (4)-(5) with $f, F \in C^d(\overline{\mathcal{D}} \times \overline{\mathbb{R}^s_+})$ is denoted by $C^d_{abs}(\overline{\mathcal{D}})$.*

Recall that $C^d(\overline{\Omega})$ is the set of functions that possess continuous d-th derivatives in the open set Ω that can be continuously extended to the boundary $\partial\Omega = \overline{\Omega}\backslash\Omega$. In the usual case where F and f are themselves compositions of smooth elemental functions φ_i these are assumed to be $C^d(\overline{\mathcal{D}}_i)$ functions on their respective domains $\mathcal{D}_i$ reachable from $x \in \mathcal{D}$. Note that a mathematical map $\varphi \in C^d_{abs}(\overline{\mathcal{D}})$ may have many different abs-normal decompositions.

Throughout the paper the symbol $|z| \in \overline{\mathbb{R}^s_+}$ is sometimes viewed as a (nonnegative) variable vector in its own right. Since F and f are smooth in the respective arguments, the derivative $L \equiv F_{|z|} \equiv \partial F(x,|z|)\,/\partial|z|$ is well defined on $\overline{\mathcal{D}} \times \overline{\mathbb{R}^s_+}$. Furthermore, it is strictly lower triangular so that one obtains the components of $z = z(x)$ one by one as piecewise smooth Lipschitz continuous functions of x. Accordingly, the other partial derivatives of $F(x,|z|)$ and $f(x,|z|)$ defined on $\overline{\mathcal{D}} \times \overline{\mathbb{R}^s_+}$ will be denoted by

$$Z \equiv \frac{\partial}{\partial x} F(x,|z|) \in \mathbb{R}^{s\times n},\ a = \frac{\partial}{\partial x} f(x,|z|) \in \mathbb{R}^n,\ b = \frac{\partial}{\partial |z|} f(x,|z|) \in \mathbb{R}^s\ . \tag{6}$$

When φ itself is piecewise linear, $y = \varphi(x)$ can be written in the abs-normal form

$$y = y_0 + a^\top(x - x_0) + b^\top(|z| - |z_0|) \quad \text{with} \quad z = z_0 + Z(x - x_0) + L(|z| - |z_0|)\ , \tag{7}$$

where $z_0 = z(x_0)$ and $y_0 = z(x_0)$. In some cases we can simplify matters by assuming that $z_0 = 0$, which means that all kinks are active at the reference point x_0.

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 $\{-1,0,1\}^s$. As observed already in [10] for the nonlinear case, the limiting gradients of φ in the vicinity of x are given by

$$g_\sigma^\top \ \equiv\ a^\top + b^\top \Sigma (I - L\Sigma)^{-1} Z = a^\top + b^\top (\Sigma - L)^{-1} Z\ , \tag{8}$$

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

$$S_\sigma = \{x \in \mathbb{R}^n \,|\, \mathbf{sgn}(z(x)) = \sigma\} \tag{9}$$

as a decomposition of the argument space, where one has

$$\varphi(x) = \varphi_\sigma(x) \qquad \text{for all } x \in S_\sigma \tag{10}$$

and φ_σ is one of finitely many differentiable selection functions in the sense of Scholtes [16]. 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

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

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 $\sigma \in \{-1, 0, 1\}^s$ 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

$$\nabla z^\sigma \equiv \frac{\partial}{\partial x} z^\sigma \;=\; (I - L\Sigma)^{-1} Z \in \mathbb{R}^{s \times n}, \tag{11}$$

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

Kink Qualifications. As illustrated later, the sets S_σ do not necessarily fulfill the Abadie constraint qualification. Therefore, we will now examine under what conditions the sets S_σ as defined in Eq. (9) satisfy the other 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

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

This partial ordering of the signature vectors was already used in [10]. 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*

$$\hat{S}_\sigma \equiv \{x \in \mathbb{R}^n \,|\, \sigma \succeq \sigma(x)\} \;\supset\; \bar{S}_\sigma \; . \tag{12}$$

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

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

For the examination of optimality we can from now on consider only maximal $\hat{S}_\sigma$, which are characterized by σ being a definite signature vector, i.e., $0 \neq \sigma_i$ for all $i = 1, \ldots, s$. This is sufficient since for any indefinite $\hat{\sigma}$, one has according to the monotonicity property that $\hat{S}_{\hat{\sigma}} \subset \hat{S}_\sigma$ for a σ being definite. We will abbreviate the definiteness of the signature vector σ by $0 \notin \sigma$ and note that then $\Sigma = \Sigma^{-1}$ is an involutory matrix. It follows that we have near $\mathring{x}$ the local decomposition property

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

As observed in [9, Sec. 3.2], the point x_* is a local minimizer of $\varphi(x)$ if and only if it is a local minimizer of each one of the *branch problems*

$$\min f(x, \bar{z}) \quad \text{s.t.} \quad \Sigma\, \bar{z} = F(x, \bar{z}) \quad \text{and} \quad \bar{z} \geq 0 \quad \text{for} \quad 0 \notin \sigma \succeq \sigma(x_*) \equiv \sigma_* \; , \tag{13}$$

where $\bar{z} \equiv \Sigma z$ irrespectively whether x_* is localized or not. For the nonlocalized case, i.e., if $|\alpha(x_*)| \neq s$, exploiting again the continuity one obtains that the components of the vector $\hat{z} \equiv (\sigma_i z_i)_{i \notin \alpha(x_*)} \equiv (|z_i|)_{i \notin \alpha(x_*)} \in \mathbb{R}^{|\sigma|}$ will keep their positive sign

in some neighborhood $\bar{B}_\rho \ni x_*$. Later on, we will use $\check{z} = (z_i)_{i\in\alpha(x_*)} \in \mathbb{R}^{|\alpha|}$ or $\bar{z} \equiv (|z_i|)_{i\in\alpha(x_*)}$ for the remaining switching variables depending on the context. Now, we could partition F and the arguments of all functions accordingly to obtain in the nonlocalized case the representation of the branch problems that is given by

$$\min f(x,\bar{z},\hat{z}) \quad \text{s.t.} \quad \check{\Sigma}\bar{z} = \check{F}(x,\bar{z},\hat{z}), \quad \hat{z} = \hat{F}(x,\bar{z},\hat{z}), \quad \bar{z} \geq 0\,, \tag{14}$$

where $\check{\Sigma} \in \{-1,1\}^{|\alpha(x_*)|}$. Given the smooth vector function z^σ as defined above for definite σ, we can describe the feasible set $\hat{S}_\sigma$ based on inequality constraints, i.e.,

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

This yields the third formulation of the branch problems given by

$$\min f_\sigma(x) \equiv f(x, \Sigma z^\sigma(x)) \quad \text{s.t.} \quad x \in \hat{S}_\sigma \quad \text{with} \quad 0 \notin \sigma \succeq \sigma_*\,. \tag{15}$$

For the derivation of the optimality conditions in the following sections, we will employ all three representation depending on the context. As can be seen from the last representation of the branch problem, where only inequalities depending on x are used to describe the feasible set, the branch problems have no special structure that could be exploited.

It is natural to look at constraint qualifications for these branch problems. First, we will show that the Abadie constraint qualification does not hold in general:

Example 2.3. Based on Exam. 1.1, we define the following piecewise smooth optimization problem

$$\varphi : \mathbb{R}^2 \mapsto \mathbb{R}, \quad \varphi(x) = |x_2| + |\tfrac{1}{3}x_1^3 - x_2|$$

with the minimizer $x_* = 0 \in \mathbb{R}^2$. The abs-normal form is given by

$$z_1 = x_2, \qquad z_2 = \tfrac{1}{3}x_1^3 - x_2, \qquad \text{and} \qquad f(x,|z|) = |z_1| + |z_2|\,.$$

Then, one has

$$Z = \begin{bmatrix} 0 & 1 \\ x_1^2 & -1 \end{bmatrix}, \quad L = 0 \in \mathbb{R}^{2\times 2}, \quad a = (0\ \ 0), \quad b = (1\ \ 1)\,,$$

and the piecewise linearization of $\varphi(.)$ at $x_* = 0$ is given by

$$\Delta\varphi : \mathbb{R}^2 \mapsto \mathbb{R}, \quad \Delta\varphi(0,\Delta x) = |\Delta x_2| + |\Delta x_2|\,,$$

where we keep the two absolute value evaluations separate to illustrate that they stem from the two absolute value evaluations in the nonlinear target function. Using the formulation (15) of the branch problem, the linearized cone, i.e., the set of first order feasible directions, at x_* for the signature vector $\sigma = (1,\,1)^\top$ is given by

$$\mathcal{F}_\sigma(x_*) = \left\{d \in \mathbb{R}^2 \,|\, d_1 \in \mathbb{R},\, d_2 = 0\right\}\,.$$

For the tangent cone of this branch problem at x_*, we obtain

$$T_\sigma(x_*) = \left\{d \in \mathbb{R}^2 \,|\, d_1 \geq 0,\, d_2 = 0\right\}\,,$$

such that $\mathcal{F}_\sigma(x_*) \neq T_\sigma(x_*)$ and the Abadie constraint qualification does not hold.

Generally, we now examine when the constraint qualification LICQ and MFCQ do hold for the branch problems. For any 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 varies. According to Eq. (11), the Jacobian of all constraints is given by

$$\Sigma \nabla z^\sigma \;=\; \Sigma(I - L\Sigma)^{-1} Z \;=\; (\Sigma - L)^{-1} Z \in \mathbb{R}^{s \times n}, \tag{16}$$

where we have used again the invertibility of $\Sigma = \Sigma^{-1}$ due to the definiteness of σ. We proved in the companion paper [19], that the Jacobian of the active constraints only has a very similar structure:

LEMMA 2.4 (Jacobian of active constraints). *Consider for a definite signature vector $\sigma \in \{-1, 1\}^s$ the branch problem (15). For $\mathring{x} \in \mathbb{R}^n$, the Jacobian of the constraints that are active at $\mathring{x}$ is given by*

$$J_\sigma \equiv (\sigma_i \nabla z_i^\sigma)_{i \in \mathring{\alpha}} = \check{\Sigma}(I - \check{L}\check{\Sigma})^{-1} \check{Z} \;=\; (\check{\Sigma} - \check{L})^{-1} \check{Z} \in \mathbb{R}^{|\mathring{\alpha}| \times n} \tag{17}$$

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

In the proof of this lemma, we derived the reduces matrices constructively as follows. Since $\sigma \succeq \mathring{\sigma}$, one has

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

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

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

and $\mathring{P} \equiv |\Gamma|$, the required matrices are given by

$$\begin{array}{l} \check{Z} = (\mathring{P}\mathring{Z})_{i \in \mathring{\alpha}, 1 \leq j \leq n} \in \mathbb{R}^{|\mathring{\alpha}| \times n}, \quad \check{\Sigma} = (\mathring{P}\Gamma)_{i \in \mathring{\alpha}, j \in \mathring{\alpha}} \in \{-1, 1\}^{|\mathring{\alpha}| \times |\mathring{\alpha}|}, \quad \text{and} \\ \check{L} = (\mathring{P}\mathring{L}\mathring{P})_{i \in \mathring{\alpha}, j \in \mathring{\alpha}} \in \mathbb{R}^{|\mathring{\alpha}| \times |\mathring{\alpha}|} \;. \end{array} \tag{18}$$

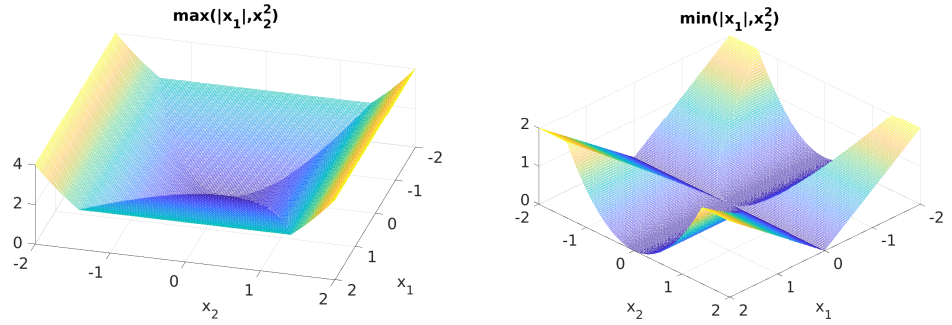
Also the following result that is based on the observation $|\det(\check{\Sigma} - \check{L})| = 1$ was proven already in [19]:

COROLLARY 2.5 (Uniformity of rank and null space). *For a given $\mathring{x}$, 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. If in the second case 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 Mangasarin Fromovitz Constraint Qualification [12] for some $\sigma \succeq \mathring{\sigma}$ requires that

$$J_\sigma v \;=\; (\check{\Sigma} - \check{L})^{-1} \check{Z} v > 0 \tag{19}$$

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 and can be



(a) $\varphi(x)$ as defined in Eq. (20) (b) $\varphi(x)$ as defined in Eq. (21) for $\varepsilon = 0$

FIGURE 1. *Example Functions from Eqs.* (20) *and* (21)

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. For a detailed discussion on the derivation of MFKQ, see the companion paper [19].

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. (19) 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. Therefore, we introduced in [19]:

DEFINITION 2.6 (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 2.4 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 in smooth optimization, it follows easily that LIKQ implies MFKQ. Furthermore, it can be shown that if $|\mathring{\alpha}| \leq n$ then LIKQ holds for almost all $[Z, L]$ [7]. 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. However, so far we needed MFKQ just for the analysis of regularity, see the companion paper [19], and it may be needed for second order necessary conditions as well. In contrast, first order necessary and second order sufficiency conditions can be derived without any kink qualification as stated below in Sec. 4.

Example Problems. To illustrate the optimality conditions and convergence rates derived in this paper, we consider six examples with varying properties:

Example 2.7. First, we consider the following convex function

$$\varphi : \mathbb{R}^2 \mapsto \mathbb{R}, \quad \varphi(x_1, x_2) = \max\left(|x_1|, x_2^2\right) \ , \tag{20}$$

see the left hand side of Fig. 1 for an illustration. Obviously, $x_* = (0, 0)$ is the only minimizer but it is not sharp. Using the reformulation

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

one obtains for $\varphi(.)$ as defined in Eq. (20) the abs-normal form

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

yielding

$$Z = \begin{bmatrix} 1 & 0 \\ 0 & -2x_2 \end{bmatrix}, \quad L = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad a = (0 \;\; x_2), \quad b = (0.5 \;\; 0.5)\,.$$

Hence, at $x_* = 0$ the matrix Z is rank deficient and thus LIKQ does not hold. For the Jacobian of the active constraints, one obtains from Eq. (17) that

$$J_\sigma = \begin{bmatrix} \sigma_1 & 0 \\ -1 & \sigma_2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \sigma_1 & 0 \\ \sigma_1\sigma_2 & \sigma_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \sigma_1 & 0 \\ \sigma_1\sigma_2 & 0 \end{bmatrix}\,.$$

Considering the signature vector $\sigma = (1 \;\; -1)^\top$, the range $R_{(1,-1)} = \{J_\sigma v \,|\, v \in \mathbb{R}^2\} = \{(v_1, -v_1)^\top \mid v_1 \in \mathbb{R}\}$ 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

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

Therefore, MFKQ does not hold. The piecewise linearization at $x_* = 0$ is given by $\Delta\varphi(x_*, \Delta x) = |\Delta x_1|$ having a first order optimal point at $\Delta x = 0$ that is not isolated.

Example 2.8. Second, we consider the following nonconvex function

$$\varphi : \mathbb{R}^2 \mapsto \mathbb{R}, \quad \varphi(x_1, x_2) = \min\left(|x_1| + \varepsilon|x_2|, x_2^2 + \varepsilon|x_1|\right) \tag{21}$$

for $\varepsilon \in \mathbb{R}$, see the right hand side of Fig. 1 for an illustration. As can be seen, for $\varepsilon = 0$, the point $x_* = 0$ is no strict minimizer. For $\varepsilon > 0$, $\varphi(.)$ the point $x_* = 0$ is an isolated minimizer but not a sharp one. If $\varepsilon < 0$ then $x_* = 0$ is only a stationary point. Using the reformulation

$$\begin{aligned} &\min(|x_1| + \varepsilon|x_2|, x_2^2 + \varepsilon|x_1|) = \\ &\qquad \tfrac{1}{2}\left(|x_1| + \varepsilon(|x_1| + |x_2|) + x_2^2 - \big||x_1| + \varepsilon(|x_2| - |x_1|) - x_2^2\big|\right)\,, \end{aligned}$$

one obtains for $\varphi(.)$ as defined in Eq. (21) the abs-normal form

$$\begin{aligned} z_1 = x_1, \quad z_2 = x_2, \quad z_3 = |z_1| + \varepsilon(|z_2| - |z_1|) - x_2^2, \\ f(x,|z|) = \tfrac{1}{2}\left(|z_1| + \varepsilon(|z_1| + |z_2|) + x_2^2 - |z_3|\right) \end{aligned}$$

yielding

$$Z = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -2x_2 \end{bmatrix}, \quad L = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1-\varepsilon & \varepsilon & 0 \end{bmatrix}, \quad a = (0 \;\; x_2), \quad b = \begin{pmatrix} \frac{1}{2}(1+\varepsilon) & \frac{1}{2}\varepsilon & -\frac{1}{2} \end{pmatrix}\,.$$

It follows that at $x_* = 0$ the matrix Z has full column but not row rank so that LIKQ does not hold. For the Jacobian of the active constraints, one obtains

$$\begin{aligned} J_\sigma &= \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ \varepsilon - 1 & -\varepsilon & \sigma_3 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -2x_2 \end{bmatrix} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ (\varepsilon-1)\sigma_1\sigma_3 & \varepsilon\sigma_2\sigma_3 & \sigma_3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -2x_2 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \\ (\varepsilon-1)\sigma_1\sigma_3 & (\varepsilon\sigma_2 - 2x_2)\sigma_3 \end{bmatrix}\,. \end{aligned}$$

Hence, for $\varepsilon = 1$ the Jacobian J_σ at $x_* = 0$ does not have full column rank such that nontrival null vectors exist, i.e., MFKQ does not hold. For $\varepsilon > 1$ and $x_* = 0$, consider $\sigma = (1\ 1\ -1)^\top$. Then, $J_\sigma v > 0$ implies

$$v_1 > 0, \qquad v_2 > 0, \qquad v_1 < \tfrac{\varepsilon}{1-\varepsilon} v_2 \ ,$$

yielding contradictory conditions for the signs of v_1 and v_2. For $\varepsilon < 1$, the requirement $J_\sigma v > 0$ yields for the signature vector $\sigma = (-1\ -1\ 1)$ contradictory conditions for the signs of v_1 and v_2 in that we should have

$$v_1 < 0, \qquad v_2 < 0, \qquad v_1 > -\tfrac{\varepsilon}{1-\varepsilon} v_2 \ .$$

The linearization of Eq. (21) at $x_* = 0$ is given by

$$\Delta\varphi(0, \Delta x) = \tfrac{1}{2}\left[|\Delta x_1| + \varepsilon(|\Delta x_1| + |\Delta x_2|) - ||\Delta x_1| + \varepsilon(|\Delta x_2| - |\Delta x_1|)|\right] \ ,$$

which has for all $\varepsilon \in \mathbb{R}$ a first order minimal point at $\Delta x = 0$ that is for $\varepsilon \geq 0$ not isolated.

Example 2.9 (Chained CB3 I [2]). The scalable chained CB3 I function is given by

$$\varphi : \mathbb{R}^n \mapsto \mathbb{R}, \varphi(x) = \sum_{i=1}^{n-1} \max\{x_i^4 + x_{i+1}^2, (2 - x_i)^2 + (2 - x_{i+1})^2, 2e^{-x_i + x_{i+1}}\} \ , \tag{22}$$

and hence a piecewise smooth convex function. Here, one has $s = 2(n-1)$ and all switching variables are active at the solution $x_* = (1 \dots 1)^\top \in \mathbb{R}^n$. Therefore, once more, LIKQ can not be fulfilled for $n > 2$. Moreover, it is easy to see that x_* is a sharp minimizer so that we can expect quadratic convergence of SPLOP.

Example 2.10 (Maxq [2]). The scalable Maxq function is defined as

$$\varphi : \mathbb{R}^n \mapsto \mathbb{R}, \varphi(x) = \max_{1 \leq i \leq n} x_i^2 \tag{23}$$

and therefore also a piecewise smooth convex function. One has $s \leq n$ depending on the specific formulation of the function evaluation. One possibility is to compare x_i^2 with the maximum of all previous square values. Alternatively, one could compare neighbors x_i^2 and x_{i+1}^2 and then compute the function value using a binary tree. Either way one would obtain at the optimal point $x_* = 0 \in \mathbb{R}^n$ that $Z = 0$, i.e., the kink qualifications LIKQ and MFKQ are not fulfilled. The minimizer x_* is isolated but not sharp.

Example 2.11 (Chained Crescent I [2]). The scalable chained Crescent I function defined as

$$\varphi : \mathbb{R}^n \mapsto \mathbb{R}, \quad \varphi(x) = \max\{\varphi_1(x), \varphi_2(x)\} \tag{24}$$
$$\varphi_1(x) = \sum_{i=1}^{n-1} (x_i^2 + (x_{i+1} - 1)^2 + x_{i+1} - 1), \ \varphi_2(x) = \sum_{i=1}^{n-1} (-x_i^2 - (x_{i+1} - 1)^2 + x_{i+1} + 1) \ ,$$

is a piecewise smooth but nonconvex function with the isolated but not sharp minimizer $x_* = (1 \dots 1)^\top \in \mathbb{R}^n$. Here, one has $s = 1$,

$$Z = (0\ 4\ \dots\ 4), \quad L = 0 \in \mathbb{R}, \quad a = (0\ 1\ \dots\ 1), \quad b = 0.5 \ ,$$

and the only switching variable is active at x_*. As can be seen, LIKQ and MFKQ hold.

Example 2.12 (LASSO problem). In statistics and machine learning, the so-called Least Absolute Shrinkage and Selection Operator (LASSO) introduced in [18] is a regression approach to perform both variable selection and regularization to enhance the prediction accuracy and interpretability of the statistical model it produces. In its Lagrangian form, for given data $w \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times n}$ the LASSO objective function is defined as

$$\varphi : \mathbb{R}^n \mapsto \mathbb{R}, \qquad \varphi(x) = \tfrac{1}{m}\|w - Ax\|_2^2 + \beta\|x\|_1 \tag{25}$$

with the penalty factor $\beta > 0$. As a possible generalization, one may consider

$$\varphi : \mathbb{R}^n \mapsto \mathbb{R}, \qquad \varphi(x) = g(x) + \beta\|x\|_1 \ ,$$

where $g : \mathbb{R}^n \mapsto \mathbb{R}$ is a sufficiently smooth function. For the abs-normal form at $\mathring{x}$, one obtains for the Lagrangian form

$$z = \mathring{x}, \qquad Z = I, \qquad b = (1 \ \ldots \ 1) \in \mathbb{R}^n, \qquad \text{and} \qquad a = \tfrac{2}{m}A^\top A\,\mathring{x} \ .$$

For the generalization, $a = g_x(x)$ is the only thing that changes. Obviously, for both formulations Z in the localized case or $\check{Z}$ in the nonlocalized case must have full rank such that LIKQ holds.

As can be seen, for these examples either LIKQ and MFKQ hold or both kink qualifications are not fulfilled. In the companion paper [19], we analyzed an example where LIKQ does not hold but MFKQ.

3. Linear Convergence Order under LIKQ. As in [9] we first consider the localized situation where all z_i vanish at the potential minimizer x_*, so that we have simply $\check{Z} = Z$ and $\check{L} = L$. Assuming throughout this section that LIKQ holds at x_*, first order optimality requires that there exists for a given localized point x_*, a Lagrange multiplier vector $\lambda_* \in \mathbb{R}^s$ such that

$$\begin{aligned} a^\top(x_*,0) + \lambda_*^\top Z(x_*,0) &= 0 && \text{Tangential Stationarity (TS)} \\ F(x_*,0) &= 0 && \text{Full kink activity} \qquad \text{and} \\ b^\top(x_*,0) + \lambda_*^\top L(x_*,0) &\geq |\lambda_*|^\top && \text{Normal Growth (NG)} \quad . \end{aligned} \tag{26}$$

The equalities represent $n + s$ equations in the unknowns (x_*, λ_*) whose Jacobian is given by the saddle point matrix

$$\begin{bmatrix} H & Z^\top \\ Z & 0 \end{bmatrix} \in \mathbb{R}^{(n+s)\times(n+s)} \tag{27}$$

with

$$H = H(x_*, \lambda_*) \equiv f(x_*,0)_{xx} + \big(\lambda^\top F(x_*,0)\big)_{xx} \in \mathbb{R}^{n\times n} \ .$$

Obviously the Hessian H is also the second derivative of the Lagrangian

$$\mathcal{L}(x,0,\lambda) = f(x,0) + \lambda^\top F(x,0) \tag{28}$$

with respect to x. The saddle point Jacobian is nonsingular provided we have SSC in that $U^\top H U \succ 0$ where $U = \check{U}$ as in Cor. 2.5. If we also have strong normal growth in that Eq. (26) holds as a strict inequality then it follows immediately that x_* is a strict minimizer.

Isolation. From the nonsingularity of the saddle point matrix (27) it follows immediately that φ can have no other tangentially stationary point that is fully localized in its neighborhood. Now the question is whether there can be other minimizers or even just tangentially stationary points $x \approx x_*$ with an active set $\alpha(x) \subsetneq \{1, \ldots, s\}$. We can reorder the components z_i such that the ones with $i \in \alpha \equiv \alpha(x)$ come first and the ones with $i \notin \alpha(x)$ come last. So we rewrite the problem in terms of $\bar{z} = (|z_i|)_{i\in\alpha}$ and $\hat{z} = (\sigma_i(x) z_i)_{i\notin\alpha(x)}$. Correspondingly we reorder the components of F into $\check{F} = (F_i)_{i\in\alpha}$ and $\hat{F} = (\sigma_i(x) F_i)_{i\notin\alpha}$. Notice that after this symmetric reordering the Jacobian of $(\check{F}^\top \ \hat{F}^\top)^\top$ w.r.t. $(\bar{z}^\top \ \hat{z}^\top)^\top$ is no longer strictly lower triangular but its two diagonal blocks $\check{F}_{\bar{z}}$ and $\hat{F}_{\hat{z}}$ still are. Since $Z = F_x$ has full row rank at x_*, LIKQ must hold in some neighborhood of x_*. In the nonlocalized case, tangential stationarity at the point x requires according to Proposition 4 of [9] for the argument $w = (x, 0, \hat{z})$ that there exists a unique multiplier vector $(\check{\lambda}^\top \ \hat{\lambda}^\top)$ such that

$$(f_x(w) \ f_{\hat{z}}(w)) \ = \ -(\check{\lambda}^\top \ \hat{\lambda}^\top) \begin{bmatrix} \check{F}_x(w) & \check{F}_{\hat{z}}(w) \\ \hat{F}_x(w) & \hat{F}_{\hat{z}}(w) - I \end{bmatrix} \ \in \ \mathbb{R}^{n+|\sigma|} \ .$$

This implies for $\sigma \equiv \sigma(x)$ the $|\sigma|$ equations

$$0 = f_{\hat{z}}(w) + \check{\lambda}^\top \check{F}_{\hat{z}}(w) + \hat{\lambda}^\top \hat{F}_{\hat{z}}(w) - \hat{\lambda}^\top \ . \tag{29}$$

We will now see that if SSC with strict complementarity, i.e., strict normal growth, hold at x_*, then the conditions (29) cannot be fulfilled with nontrivial $\hat{z}$ and $\hat{F}$, i.e., with $\alpha(x) \neq \alpha_*$. Reordering and partitioning the normal growth condition that holds strictly at x_* we get at the argument $w_* = (x_*, 0, 0)$

$$(f_x(w_*) \ f_{\hat{z}}(w_*)) + (\check{\lambda}_*^\top \ \hat{\lambda}_*^\top) \begin{bmatrix} \check{F}_{\hat{z}}(w_*) & \check{F}_{\bar{z}}(w_*) \\ \hat{F}_{\hat{z}}(w_*) & \hat{F}_{\bar{z}}(w_*) \end{bmatrix} \geq \varepsilon e_s^\top + (|\check{\lambda}_*|^\top \ |\hat{\lambda}_*|^\top)$$

with $e_s = (1, \ldots, 1)^\top \in \mathbb{R}^s$ and for some $\varepsilon > 0$. Here, no identity matrix occurs in the lower right block, since x_* is localized which implies the equality constraint $0 = \hat{F}(w_*)$. Extracting the $|\sigma|$ components that belong to the indices $i \notin \alpha(x)$, the $|\sigma|$ inequalities

$$|\hat{\lambda}_*|^\top + \varepsilon e_{|\sigma|}^\top \leq f_{\hat{z}}(w_*) + \check{\lambda}_*^\top \check{F}_{\bar{z}}(w_*) + \hat{\lambda}_*^\top \hat{F}_{\bar{z}}(w_*) \tag{30}$$

with $e_{|\sigma|} = (1 \ \ldots \ 1)^\top \in \mathbb{R}^{|\sigma|}$ must hold. Because of the assumed continuous differentiability all first derivatives of F and f evaluated at $(x, \bar{z}(x), \hat{z}(x))$ are small perturbations of their values at $(x_*, 0, 0)$. We abbreviate this discrepancies as $\mathcal{O}(\|x - x_*\|)$. Since $Z = F_x(w)$ has by LIKQ full row rank the multiplier vector $(\check{\lambda}^\top \ \hat{\lambda}^\top)$, provided it exists at all, is also a small perturbation of its value at $(\check{\lambda}_*^\top, \hat{\lambda}_*^\top)$. Hence we may subtract Eq. (29) from Eq. (30) and obtain after transposition

$$0 < \varepsilon e_{|\sigma|} \leq |\hat{\lambda}_*| + \varepsilon e_{|\sigma|} \leq \mathcal{O}(\|x - x_*\|) + \hat{\lambda} \ \leq |\hat{\lambda}_*| + \mathcal{O}(\|x - x_*\|) \ ,$$

which is obviously a contradiction for x close enough to x_*. Hence there can be no other point satisfying KKT or even just tangential stationarity in some level set $\mathcal{N} = \{x \in \mathbb{R}^n | f(x) \leq f(x_*) + \varepsilon\}$. Now we formulate this for the general case, where x_* itself is not necessarily localized.

PROPOSITION 3.1. *Suppose x_* satisfies SSC under LIKQ for the C^d_{abs} function $\varphi(x)$. Then in some neighborhood of x_* there can be no other point satisfying tangential stationarity, let alone Clarke stationarity or first order minimality (FOM).*

Proof. Regarding the last assertion, note that by Lemma 3 in [9] under LIKQ, Clarke stationarity requires tangential stationarity, and of course it is implied by FOM. In the lead-up to this proposition we have already proven the assertion for the fully active, i.e., localized, case where $\alpha(x_*) = \{1, \ldots, s\}$. For the nonlocalized situation, let us first consider the case where at a prospective tangential stationary point x the active index set $\alpha(x)$ is the same as at x_*, i.e., $\alpha(x) = \alpha(x_*) \equiv \alpha_*$. Using again the same partition as above, for $w_* = (x_*, 0, \hat{z}_*)$ tangential stationarity means that the set of equations

$$\begin{aligned} 0 \;&=\; (f_x(w_*)\; f_{\check{z}}(w_*)) + (\,\check{\lambda}_*^\top\; \hat{\lambda}_*^\top) \begin{bmatrix} \check{F}_x(w_*) & \check{F}_{\check{z}}(w_*) \\ \hat{F}_x(w_*) & \hat{F}_{\check{z}}(w_*) - I \end{bmatrix} \;\in\; \mathbb{R}^{n+|\sigma|}\,, \\ 0 \;&=\; \check{F}(w_*) \in \mathbb{R}^{|\alpha|}, \qquad\qquad \hat{z} = \hat{F}(w_*) \in \mathbb{R}^{|\sigma|} \end{aligned}$$

is satisfied. Here, the identity in the lower right block of the matrix occurs since x_* is not localized. Since x is assumed to be another stationary point, there must exist a multiplier vector $(\,\check{\lambda}^\top\; \hat{\lambda}^\top)$ such that the same set of equations is fulfilled when evaluated at the point $w = (x, 0, \hat{z})$. Thus we have in total $n+|\sigma|+|\alpha|+|\sigma| = n+s+|\sigma|$ equations in the variables $x, \hat{z}, \check{\lambda}, \hat{\lambda}$ since $\check{z} \equiv 0$ by definition. This is a square system which has isolated solutions if its Jacobian has full rank. The Jacobian is given by

$$\begin{bmatrix} \nabla^2_{xx}\mathcal{L}(w_*) & \nabla^2_{x\hat{z}}\mathcal{L}(w_*) & \check{F}_x(w_*)^\top & \hat{F}_x(w_*)^\top \\ \nabla^2_{\hat{z}x}\mathcal{L}(w_*) & \nabla^2_{\hat{z}\hat{z}}\mathcal{L}(w_*) & \check{F}_{\hat{z}}^\top(w_*) & \hat{F}_{\hat{z}}^\top(w_*) - I \\ \check{F}_x(w_*) & \check{F}_{\hat{z}}(w_*) & 0 & 0 \\ \hat{F}_x(w_*) & \hat{F}_{\hat{z}}(w_*) - I & 0 & 0 \end{bmatrix} \in \mathbb{R}^{(n+|\sigma|+|\alpha|+|\sigma|)\times(n+|\sigma|+|\alpha|+|\sigma|)}\,,$$

where $\mathcal{L}$ denotes the Lagrangian as introduced in Eq. (28). LIKQ and SSC at x_* require exactly that this symmetric matrix has full rank so that there can be no other tangentially stationary point $x \approx x_*$ with the same activity in a neighborhood of x_*.

Since in some neighborhood of x_* all $z_i(x)$ with $i \notin \alpha_*$ keep their sign, the only possibility for a change of activity is that $\alpha(x) \subsetneq \alpha_*$. As in the localized case we will exclude this by considering the strict normal growth condition assumed to hold at x_*. Without loss of generality we can assume that the last indices $i = |\alpha| + 1, \ldots, |\alpha_*|$ represent the indices $i \in \alpha_*$ that do not belong to $\alpha(x)$. Abbreviating $\tilde{z}_i = \sigma_i(x) z_i$ and $\tilde{F}_i = \sigma_i(x) F_i$ for $i = |\alpha| + 1, \ldots, |\alpha_*|$ we get at x as tangential stationarity condition the $n + |\sigma(x)|$ equations

$$0 = (f_x(w)\; f_{\check{z}}(w)\; f_{\hat{z}}(w)) + (\,\check{\lambda}^\top\; \tilde{\lambda}^\top\; \hat{\lambda}^\top) \begin{bmatrix} \check{F}_x(w) & \check{F}_{\check{z}}(w) & \check{F}_{\hat{z}}(w) \\ \tilde{F}_x(w) & \tilde{F}_{\tilde{z}}(w) - I & \tilde{F}_{\hat{z}}(w) \\ \hat{F}_x(w) & \hat{F}_{\tilde{z}}(w) & \hat{F}_{\hat{z}}(w) - I \end{bmatrix}. \tag{31}$$

In other words, we have partitioned the switching variables that are inactive at x into those, namely $\tilde{z}$ that are active only at x_* and the remainder that are inactive at x and x_*, which we still denote as $\hat{z}$. Of course we did the same for the function vector F. As before we notice that strict normal growth at x_* requires that with the same partitioning the $|\alpha_*|$ inequalities

$$(|\mathring{\lambda}_*|^\top\, |\tilde{\lambda}_*|^\top) + \varepsilon e_{|\alpha_*|}^\top \le (f_{\check{z}}(w_*)\; f_{\tilde{z}}(w_*)) + (\,\check{\lambda}_*^\top\; \tilde{\lambda}_*^\top\; \hat{\lambda}_*^\top) \begin{bmatrix} \check{F}_{\check{z}}(w_*) & \check{F}_{\tilde{z}}(w_*) \\ \tilde{F}_{\check{z}}(w_*) & \tilde{F}_{\tilde{z}}(w_*) \\ \hat{F}_{\check{z}}(w_*) & \hat{F}_{\tilde{z}}(w_*) \end{bmatrix} \tag{32}$$

hold, where $\mathring{z}_i = \sigma_i(x) z_i$ for $i = 1, \ldots, |\alpha|$ denote the switching variables that are active at x_* and x. Subtracting the middle block of Eq. (31) from the second block of Eq. (32), we get as in the localized case with $\beta = |\alpha_*| - |\alpha|$

$$|\tilde{\lambda}| + \varepsilon e_\beta \;\leq\; \tilde{\lambda} + \mathcal{O}(\|x - x_*\|) \,,$$

which yields as above a contradiction for x sufficiently close to x_*. ∎

Now we can address the question at what rate SPLOP converges to minimizers satisfying the above isolation theorem.

Linear Convergence. Under the assumption of Prop. 3.1 we know that x_* is a strict local minimizer of second order in that

$$\varphi(x) - \varphi(x_*) \;\geq\; c\|x - x_*\|^2 \quad \text{for some } c > 0 \text{ and all } \; x \approx x_* \,.$$

Hence, there is a compact level set $\mathcal{N} = \{x \in \mathbb{R}^n | f(x) \leq f(x_*) + \varepsilon\}$ containing no other first order minimal points. Moreover, we will assume that within $\mathcal{N}$ Eq. (1) holds with a constant coefficient $\gamma/2$ and that the regularization parameter q introduced in Eq. (2) is bigger than γ so that for $x_k \in \mathcal{N}$

$$\begin{aligned} \varphi(x_{k+1}) - \varphi(x_k) &\leq \Delta\varphi(x_k; x_{k+1} - x_k) + \tfrac{\gamma}{2}\|x_{k+1} - x_k\|^2 \\ &\leq \tfrac{\gamma - q}{2}\|x_{k+1} - x_k\|^2 < 0 \;\;\Longrightarrow\;\; x_{k+1} \in \mathcal{N} \,. \end{aligned}$$

That is, the iterates generated by SPLOP stay in the neighborhood of x_* and it was shown in [4] that all cluster points of the sequence must be at the very least Clarke stationary. By Prop. 3.1 there is no such point other than x_* in the neighborhood of $\mathcal{N}$, which implies convergence, i.e., $x_k \to x_*$. So now the challenge is to show that the rate of convergence is linear. We know that the piecewise linear model problems $\Delta\varphi(x_k; \Delta x)$ are similar to the one at the minimizers given by $\Delta\varphi(x_*; \Delta x)$ in that the discrepancies between the model data are of order $\mathcal{O}(\|x_k - x_*\|)$.

Let us assume once more that all kinks are active at x_* and consider the general, nonlocalized case later. Often the piecewise linear local models themselves will be unbounded and thus not have a minimizer but the regularized ones do. Their optimal points x_{k+1} if also fully active, i.e., localized, must satisfy the FOM conditions

$$\begin{aligned} a_k + q(x_{k+1} - x_k) + Z_k^\top \lambda_{k+1} &= 0 \\ F(x_k, |z(x_k)|) + Z_k(x_{k+1} - x_k) &= 0 \\ b_k^\top + \lambda_{k+1}^\top L_k \;\geq\; |\lambda_{k+1}|^\top & \end{aligned} \qquad (33)$$

Since we know already from the fact of convergence that $x_k - x_*, x_{k+1} - x_*$ and thus $x_{k+1} - x_k$ are small, the argument used to prove the isolation of x_* can also be used to show that x_{k+1} must be given in this way as it cannot have lesser activity than x_*. In other words the active kinks are properly identified by SPLOP in some vicinity of a fully active minimizer.

We can view the two block equations as a fixed point iteration and observe that by the above Prop. 3.1 the only fixed point of this iteration is the minimizer x_* with

the corresponding multiplier λ_*. Solving for (x_{k+1}, λ_{k+1}) we get

$$
\begin{aligned}
\begin{bmatrix} x_{k+1} \\ \lambda_{k+1} \end{bmatrix} &= \begin{bmatrix} qI & Z_k^\top \\ Z_k & 0 \end{bmatrix}^{-1} \begin{bmatrix} qx_k - a_k \\ Z_k x_k - F_k \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{q}(I - Z_k^+ Z_k) & Z_k^+ \\ (Z_k^+)^\top & -q(Z_k Z_k^\top)^{-1} \end{bmatrix} \begin{bmatrix} qx_k - a_k \\ Z_k x_k - F_k \end{bmatrix} \\
&= \begin{bmatrix} x_k - \frac{1}{q}(I - Z_k^+ Z_k)a_k - Z_k^+ F_k \\ -(Z_k^+)^\top a_k + q(Z_k Z_k^\top)^{-1} F_k \end{bmatrix} ,
\end{aligned} \tag{34}
$$

where $Z_k^+ = Z_k^\top (Z_k Z_k^\top)^{-1}$ is the generalized inverse of Z_k. To establish the contractivity we have to consider the eigenvalues of the Jacobian of the right hand side as follows:

LEMMA 3.2. *In the localized case where $\check{H} = H$ as defined above and with $\check{U} = U$ an orthogonal null space matrix of $\check{Z} = Z$, the nonzero eigenvalues of the Jacobian $(\partial x_{k+1}, \partial \lambda_{k+1}) / (\partial x_k, \partial \lambda_k)$ evaluated at x_* are those of the matrix difference*

$$
\left[I - U_*^\top H_* U_* / q \right] \in \mathbb{R}^{(n-s)\times(n-s)} .
$$

Proof. Since λ_k does not enter into the very right hand side of Eq. (34) at all we only have to consider the first block

$$
J_k = \tfrac{\partial x_{k+1}}{\partial x_k} = I - \tfrac{\partial}{\partial x_k} \left[\tfrac{1}{q}(I - Z_k^+ Z_k) a_k \right] - \tfrac{\partial}{\partial x_k} \left[Z_k^+ F_k \right] .
$$

Dropping the index k and evaluating the right hand side at the fixed point x_* where $F = 0$ and $a = -Z^\top \lambda_*$ we obtain

$$
J_* = (I - Z^+ Z)(I - \tfrac{1}{q} f_{xx}) - \tfrac{1}{q} \left[\tfrac{\partial}{\partial x}(Z^+ Z) \right] Z^\top \lambda_* .
$$

To get the derivative in the last term for fixed λ_* we calculate using the chain rule

$$
\begin{aligned}
\left[\tfrac{\partial}{\partial x}(Z^+ Z) \right] Z^\top \lambda_* &= \left[\tfrac{\partial}{\partial x}(Z^+ Z Z^\top) \right] \lambda_* - Z^+ Z \left[\tfrac{\partial}{\partial x} Z^\top \right] \lambda_* \\
&= \left[\tfrac{\partial}{\partial x} Z^\top \right] \lambda_* - Z^+ Z \left[\tfrac{\partial}{\partial x} Z^\top \right] \lambda_* = \left[I - Z^+ Z \right] F_{xx}^\top \lambda_* .
\end{aligned}
$$

Substitution now yields the final expression

$$
J_* = [I - Z^+ Z](I - \tfrac{1}{q}(f_{xx} + \lambda_*^\top F_{xx})) \;=\; (I - Z^+ Z)(I - \tfrac{1}{q} H)
$$

with H the Hessian of the Lagrangian as defined above. The eigenvectors of J_* with nonzero eigenvalues all belong to the range of the orthogonal projection $UU^\top = (I - Z^+ Z)$ and are thus the eigenvectors of

$$
I - \tfrac{1}{q} U^\top H U \in \mathbb{R}^{(n-s)\times(n-s)} ,
$$

which completes the proof. ∎

Hence, we see immediately that the fixed point iteration has a spectral radius less than 1 and is thus contractive provided SSC is satisfied and q sufficiently large. To formulate this result for the general, nonlocalized case let us recall the definition of $\check{Z}$ in Lemma 2.4, its null space matrix $\check{U} \in \mathbb{R}^{n\times(n-|\alpha|)}$ in Cor. 2.5 and from [9] the reduced Hessian

$$
\check{H} = \begin{bmatrix} I & \hat{z}_x^\top \end{bmatrix} \begin{bmatrix} f_{xx} + (\check{\lambda}^\top \check{F})_{xx} & f_{x\hat{z}} + (\check{\lambda}^\top \check{F})_{x\hat{z}} \\ f_{\hat{z}x} + (\check{\lambda}^\top \check{F})_{\hat{z}x} & f_{\hat{z}\hat{z}} + (\check{\lambda}^\top \check{F})_{\hat{z}\hat{z}} \end{bmatrix} \begin{bmatrix} I \\ \hat{z}_x \end{bmatrix} \in \mathbb{R}^{n\times n} , \tag{35}
$$

where

$$\hat{z}_x \equiv \frac{\partial \hat{z}}{\partial x} \;=\; (I - \hat{F}_{\hat{z}})^{-1}\hat{F}_x \;. \tag{36}$$

PROPOSITION 3.3. *Suppose x_* satisfies SSC under LIKQ for $\varphi(.) \in C^d_{abs}$. Assume that $q > \max(\gamma, \|\check{U}_*^\top \check{H}_* \check{U}_*\|)$ holds for the regularization parameter q. Then SPLOP yields local and linear convergence with the R-factor*

$$\|I - \tfrac{1}{q}\check{U}_*^\top \check{H}_* \check{U}_*\| \geq 1 - (\kappa(\check{U}_*^\top \check{H}_* \check{U}_*))^{-1} \;,$$

where κ denotes the condition number with respect to the spectral norm.

It should be noted that generally we can expect that the constant γ is much larger than $\|\check{U}^\top \check{H} \check{U}\|$ since it is a bound on all curvatures including the ones occurring in the Hessian and its projection.

Proof. To prove the general result we show that the iterates generated for the nonlocalized optimization problem are identical to the ones generated for the corresponding reduced optimization problem. Generally x_{k+1} solves the problem

$$\min \Delta\varphi(x_k; x - x_k) + \tfrac{q}{2}\|x - x_*\|^2 \;,$$

which is equivalent to the constrained problem

$$\begin{aligned} &\min a_k^\top (x - x_k) + b_k^\top |z| + \tfrac{q}{2}\|x - x_k\|^2 \\ &s.t. \quad z = F_k + Z_k(x - x_k) + L_k|z| \;. \end{aligned}$$

This problem is obtained by simply linearizing f and F. Since we know that the matrices and vectors defining these problems are $\mathcal{O}(\|x - x_*\|)$ perturbations of their values at the KKT and SSC point x_* we know that the solution of the model problem has the same kink activity. Thus after partitioning accordingly we get as optimality conditions for x and $z(x)$ the set of equations

$$\begin{aligned} 0 \;&=\; \begin{bmatrix} a_k + q(x - x_k) \\ \hat{b}_k \end{bmatrix} + \begin{bmatrix} \check{F}_x^\top & \hat{F}_x^\top \\ \check{F}_{\hat{z}}^\top & \hat{F}_{\hat{z}}^\top - I \end{bmatrix} \begin{bmatrix} \check{\lambda} \\ \hat{\lambda} \end{bmatrix} \;\in\; \mathbb{R}^{n+|\sigma|} \;, \\ 0 \;&=\; \check{F} + \check{F}_x(x - x_k) + \check{F}_{\hat{z}}(\hat{z} - \hat{z}_k) \;, \\ 0 \;&=\; \hat{F} + \hat{F}_x(x - x_k) + (\hat{F}_{\hat{z}} - I)(\hat{z} - \hat{z}_k) \;, \end{aligned}$$

where all quantities are evaluated at the current iterate (x_k, z_k). Now we can use the nonsingularity of the unitary lower triangular matrix $I - \hat{F}_{\hat{z}}$ to eliminate both $\hat{z} - z_k$ and $\hat{\lambda}$ to obtain

$$\begin{aligned} 0 \;&=\; \left[a_k + \hat{F}_x^\top (I - \hat{F}_{\hat{z}}^\top)^{-1}\hat{b}_k\right] + q(x - x_k) + \left[\; \check{F}_x^\top + \hat{F}_x^\top (I - \hat{F}_{\hat{z}}^\top)^{-1}\check{F}_{\hat{z}}^\top\right] \check{\lambda} \;, \\ 0 \;&=\; \left[\check{F} + \check{F}_{\hat{z}}(I - \hat{F}_{\hat{z}})^{-1}\hat{F}\right] + \left[\check{F}_x + \check{F}_{\hat{z}}(I - \hat{F}_{\hat{z}})^{-1}\hat{F}_x\right](x - x_k) \;. \end{aligned}$$

Thus we see that the step from x_k to x_{k+1} is exactly the same as one would obtain on the reduced problem where $\check{a} = a + \hat{F}_x^\top (I - \hat{F}_{\hat{z}}^\top)^{-1}\hat{b}$, $\check{Z}^\top = \check{F}_x^\top + \hat{F}_x^\top (I - \hat{F}_{\hat{z}}^\top)^{-1}\check{F}_{\hat{z}}^\top$ and $\hat{y} = \check{F} + \check{F}_{\hat{z}}(I - \hat{F}_{\hat{z}}^\top)^{-1}\hat{F}$. Hence we conclude that locally under the given assumptions the SPLOP iteration for the original problem and the reduced, localized problem are indeed identical. This completes the proof as it is well known [15] that continuously differentiable fixed point iterations converge locally with an R-Factor no larger than the spectral radius of its Jacobian at a given fixed point. ∎

From the examples introduced in the previous section, LIKQ only holds for the Chained Crescent I example and the LASSO problem. For the Chained Crescent I problem, see Exam. 2.11, one obtains from the tangential stationarity that $\lambda = 0.25$. Furthermore the blocks $f_{xx}, f_{xz}, f_{zx}, f_{tt}, F_{xz}, F_{zx}, F_{zz}$ are all zero at $x_* = 0 \in \mathbb{R}^n$. Only F_{xx} is a diagonal matrix with positive entries on the diagonal. Hence, SSC holds at the minimizer.

Next, we will examine the LASSO problem in more detail to derive conditions for SSC to hold. For the LASSO problem, the optimizer x_* is nonlocalized since otherwise one would obtain the trivial solution $x_* = 0$. Therefore, we have to consider for $\alpha_* \equiv \alpha(x_*)$ the reduced matrix $\check{Z}$ and a basis $\check{U}$ of its null space given by

$$\check{Z} = (e_i^\top)_{i \in \alpha_*} \in \mathbb{R}^{|\alpha_*| \times n} \qquad \text{and} \qquad \check{U} = (e_i)_{i \notin \alpha_*} \in \mathbb{R}^{n \times |\sigma_*|} \,, \tag{37}$$

where e_i denotes the ith unit vector in $\mathbb{R}^n$. Based on this observation one obtains the following result:

LEMMA 3.4. *For the LASSO problem, at the optimal point x_* the kink qualification LIKQ is always satisfied. The property SSC holds for the Lagrangian form of the LASSO problem if one has for the data $w \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$ that $m > n$ and A has full column rank. For the generalized problem SSC is fulfilled if $\check{U}^\top g_{xx}(x_*)\check{U}$ is positive definite.*

Proof. It follows directly from Eq. (37) that LIKQ is always fulfilled for the LASSO problem. Therefore, we only have to show the conditions for SSC. Combining the formulation (15) and the Lagrangian form (25) one obtains

$$\begin{aligned} &\min \tfrac{1}{m}(w - Ax)^\top (w - Ax) + \beta(\bar{z} + \hat{z}) \quad \text{s.t.} \\ &\check{\Sigma}\bar{z} = (x_i)_{i \in \alpha_*}, \quad \hat{z} = (\sigma_i x_i)_{i \in \sigma_*}, \quad \bar{z} \geq 0 \,. \end{aligned}$$

It follows that the derivative f_{xx} is the only nonzero block of the second derivative of the Lagrangian function, since this is the only nonlinear term. Hence, one has

$$\check{U}^\top \check{H} \check{U} \succ 0 \qquad \Leftrightarrow \qquad \tfrac{2}{m}\check{U}^\top A^\top A \check{U} \succ 0 \,,$$

which is the case if $m > n$ and A has full column rank. For the generalized LASSO problem only the objective function changes into

$$\min g(x) + \beta(\bar{z} + \hat{z}) \,.$$

Once more, the only nonlinear term is $g(x)$ yielding

$$\check{U}^\top \check{H} \check{U} = \check{U}^\top g_{xx}(x_*)\check{U} \succ 0 \,,$$

and therefore the assertion. ∎

It follows from the last lemma that we can expect linear convergence of SPLOP for the LASSO problem as soon as $m > n$ and A has full column rank for the Lagrangian form or $\check{U}^\top g_{xx}(x_*)\check{U}$ is positive definite in the generalized form.

4. Optimality Conditions without Kink Qualification. As demonstrated by the Examples 2.7 – 2.10, LIKQ and even MFKQ may not always be a natural condition to impose. We already observed in the introduction that a necessary condition for local optimality of φ at x_* is First Order Minimality (FOM), i.e., local minimality of $\Delta\varphi(x_*; \Delta x)$ at $\Delta x = 0$. Moreover, sharp minimality of the former (SMIN) is equivalent to strict minimality of the latter (SFOM). We now formulate equivalent conditions for FOM and thus necessary conditions for MIN as well as sufficient conditions for (SFOM) and thus sufficient conditions for (IMIN).

THEOREM 4.1 (Characterization of (Strict) First Order Minimality (FOM)). *Let $\varphi(x)$ be itself piecewise linear and given in the abs-normal form (7). Assume that the potential minimum x_* is shifted to the origin and localized in that all s kinks are active. Then, $\varphi(.)$ has a local minimum at $x_* = 0$ if and only if there exists for each definite signature vector $\sigma \in \{-1, 1\}^s$ a vector of Lagrange multipliers $0 \leq \mu_\sigma \in \mathbb{R}^s$ such that*

$$a^\top + (b - \mu_\sigma)^\top (\Sigma - L)^{-1} Z = 0 \; . \tag{38}$$

Moreover, we must have $a \in \text{range}(Z^\top)$ such that $a^\top = \lambda^\top Z$ for $\lambda \in \Lambda$ and $\Lambda \subset \mathbb{R}^s$ an affine space of multipliers. If furthermore

$$|\lambda_*|^\top \leq b^\top + \lambda_*^\top L \qquad \textit{for some} \quad \lambda_* \in \Lambda \; , \tag{39}$$

then Eq. (38) is automatically satisfied by

$$\mu_\sigma^\top = b^\top - \lambda_*^\top (\Sigma - L) \qquad \textit{for all} \quad \sigma \in \{-1, 1\}^s \; . \tag{40}$$

Finally, the minimizer $x_* = 0$ can only be sharp if Z has full column rank n and it must be sharp if in addition Eq. (39) holds as a strict inequality.

Proof. First we show the forward implication. By Farkas' Lemma [14, Lemma 12.4] we have for each σ either a feasible descent direction $d \in \mathbb{R}^n$ such that $g_\sigma^\top d < 0$ and $J_\sigma d \geq 0$ or a Lagrange multiplier vector $0 \leq \mu_\sigma \in \mathbb{R}^s$ such that $g_\sigma^\top = \mu_\sigma^\top J_\sigma$. Due to the assumed minimality of $x_* = 0$ the latter alternative must apply and we get with (8) and (17)

$$0 = g_\sigma^\top - \mu_\sigma^\top J_\sigma = a^\top + b^\top (\Sigma - L)^{-1} Z - \mu_\sigma^\top (\Sigma - L)^{-1} Z \; ,$$

which is equivalent to the assertion. The converse implication is equally simple since the nonexistence of a suitable μ_σ for at least one definite σ guarantees the existence of a descent direction on that branch problem so that $x_* = 0$ cannot be a local minimizer. Moreover we see immediately that for any σ one has $a^\top = -\lambda_\sigma^\top Z$ when setting $\lambda_\sigma^\top \equiv (b - \mu_\sigma)^\top (\Sigma - L)^{-1}$ which implies the second assertion. Notice that the case where the feasible polyhedron S_σ consists only of the origin $\{0\}$ itself is automatically taken care of by Farkas' Lemma.

The nonnegativity or even strict positivity of all μ_σ follows immediately from

$$\mu_\sigma^\top = b^\top - \lambda_*^\top (\Sigma - L) \; \geq \; b^\top - |\lambda_*|^\top + \lambda_*^\top L \tag{41}$$

and the assumptions on λ_*. Finally, if Z had a null vector $d \neq 0$ that direction d would be a feasible direction for all branch problems and we would have $g_\sigma^\top d = \mu_\sigma^\top J_\sigma d = (\Sigma - L)^{-1} Z d = 0$ so that $x_* = 0$ could not be a strict minimizer. Otherwise $v = J_\sigma d$ must have for any feasible direction d at least one positive component so that $g_\sigma^\top d = \mu_\sigma^\top v > 0$ due to the strict positivity of μ_σ. Hence x_* must be a strict minimizer under the given assumptions on Z and λ_*. ∎

Note that under LICQ we have a singleton $\Lambda = \{\lambda_*\}$ and Eq. (39) represents exactly the normal growth condition originally derived in [9]. Without the assumption of LIKQ, the set Λ will be an affine subspace of dimension one or more and Eq. (40) for μ_σ may only apply with a different $\lambda = \lambda_\sigma \in \Lambda$ for each σ. The higher the dimension the more difficult this check might become, and we conjecture it to be NP co-complete, which was shown for the test of First Order Convexity, a necessary

condition for subdifferential regularity, in [19]. It is much easier to check whether a suitable Λ_* for the sufficient condition Eq. (38) exists, namely one has to solve the LOP problem

$$\min e_s^\top v \quad \text{s.t.} \quad -v \le \lambda_* \le v, \quad 0 = Z^\top \lambda_* + a, \quad v \le b + L^\top \lambda_* - \varepsilon e_s$$

in the variables $v \in \mathbb{R}^s$ and $\lambda_* \in \mathbb{R}^s$ for given $\varepsilon \ge 0$. If for some $\varepsilon > 0$ the feasible set is nonempty we have established strict minimality. We note that without LIKQ there remains a gap between computable necessary and sufficient conditions, which could be closed by further, more detailed analysis.

In the nonlocalized scenario, one obtains as generalization of [9, Prop. 4] the following result:

COROLLARY 4.2 (Characterization of (Strict) FOM in the nonlocalized case). *Let $\varphi(x)$ be itself piecewise linear and given in the abs-normal form (7). Assume that x_* is nonlocalized with the signature $\sigma_* \equiv \sigma(x_*)$ such that with $\alpha_* \equiv \alpha(x_*)$ one has $|\alpha_*| < s$. Then, $\varphi(.)$ has a local minimizer at $x_* = 0$ if and only if there exist for each definite signature vector $\sigma \succ \sigma_*$ Lagrange multiplier vectors $\check\lambda_\sigma \in \mathbb{R}^{|\alpha_*|}$, $\hat\lambda_\sigma \in \mathbb{R}^{|\sigma_*|}$ and $\mu_\sigma \in \mathbb{R}^{|\alpha_*|}, \mu_\sigma \ge 0$, such that*

$$(f_x \; f_{\hat z}) = -(\check\lambda_\sigma^\top \; \hat\lambda_\sigma^\top) \begin{bmatrix} \check F_x & \check F_{\hat z} \\ \hat F_x & \hat F_{\hat z} - I \end{bmatrix} \quad \text{and} \quad f_{\bar z} + (\check\lambda_\sigma^\top \; \hat\lambda_\sigma^\top) \begin{bmatrix} \check F_{\bar z} - \check\Sigma \\ \hat F_{\bar z} \end{bmatrix} = \check\mu_\sigma^\top \ge 0 \,, \tag{42}$$

where all derivatives are evaluated at $(x_*, 0, \hat z(x_*, 0))$. Moreover, we must have for $\check a \equiv (f_x \; f_{\bar z})$ that

$$\check a \in \operatorname{range}\left(\begin{bmatrix} \check F_x & \check F_{\hat z} \\ \hat F_x & \hat F_{\hat z} - I \end{bmatrix}^\top \right), \quad \textit{i.e.,} \quad \check a^\top = \lambda^\top \begin{bmatrix} \check F_x & \check F_{\hat z} \\ \hat F_x & \hat F_{\hat z} - I \end{bmatrix}$$

for $\lambda^\top = (\check\lambda^\top \; \hat\lambda^\top) \in \Lambda$ and $\Lambda \subset \mathbb{R}^s$ being an affine space of multipliers. If furthermore there exists an element $\lambda_* \in \Lambda$ such that

$$|\check\lambda_*^\top| \le f_{\bar z} + \check\lambda_*^\top \check F_{\bar z} + \hat\lambda_*^\top \hat F_{\bar z} \,, \tag{43}$$

then Eq. (42) is automatically satisfied by

$$\mu_\sigma^\top \equiv f_{\bar z} + \check\lambda_*^\top (\check F_{\bar z} - \check\Sigma) + \hat\lambda_*^\top \hat F_{\bar z} \qquad \textit{for all} \quad \sigma \in \{-1, 1\}^{|\alpha|} \,.$$

Finally, the minimizer $x_* = 0$ can only be sharp if the matrix

$$\check Z \equiv \check F_x + \check F_{\hat z}(I - \hat F_{\hat z})^{-1} \hat F_x \in \mathbb{R}^{|\alpha| \times n}$$

has full column rank n and it must be sharp if in addition Eq. (43) holds as a strict inequality.

Proof. To derive these result we use the formulation (14) of the branch problems and follow the lines of the previous theorem. Hence, applying Farkas' Lemma for the branch problem defined by $\sigma \succ \sigma_*$ the point $x_* = 0$ is a minimizer if and only if there exist Lagrange multiplier vectors $\check\lambda_\sigma \in \mathbb{R}^{|\alpha_*|}$, $\hat\lambda_\sigma \in \mathbb{R}^{|\sigma_*|}$, and $\check\mu_\sigma \in \mathbb{R}^{|\alpha_*|}, \check\mu_\sigma \ge 0$, such that

$$f_x + \check\lambda_\sigma^\top \check F_x + \hat\lambda_\sigma^\top \hat F_x = 0 \,, \qquad f_{\hat z} + \check\lambda_\sigma^\top \check F_{\hat z} + \hat\lambda_\sigma^\top (\hat F_{\hat z} - I) = 0 \,, \qquad \text{and} \tag{44}$$

$$f_{\bar z} + \check\lambda_\sigma^\top (\check F_{\bar z} - \check\Sigma) + \hat\lambda_\sigma^\top \hat F_{\bar z} = \check\mu_\sigma^\top \ge 0 \,. \tag{45}$$

Combining the two equations in Eq. (44) gives

$$(f_x \; f_{\hat{z}}) = -(\check{\lambda}_\sigma^\top \; \hat{\lambda}_\sigma^\top) \begin{bmatrix} \check{F}_x & \check{F}_{\hat{z}} \\ \hat{F}_x & \hat{F}_{\hat{z}} - I \end{bmatrix}$$

and therefore the first assertion. Collecting for all definite $\sigma \succ \sigma_*$ the corresponding Lagrange vectors $\lambda_\sigma^\top \equiv (\check{\lambda}_\sigma^\top \; \hat{\lambda}_\sigma^\top)$ in one affine set Λ yields the second assertion. Here, we assume as above, that the entries of z are sorted correspondingly. Furthermore, if there exists an element $\lambda_* = (\check{\lambda}_*^\top \; \hat{\lambda}_*^\top) \in \Lambda$ such that

$$|\check{\lambda}_*^\top| \leq f_{\bar{z}} + \check{\lambda}_*^\top \check{F}_{\bar{z}} + \hat{\lambda}_*^\top \hat{F}_{\bar{z}} \; ,$$

then one can define

$$\mu_\sigma^\top \equiv f_{\bar{z}} + \check{\lambda}_*^\top (\check{F}_{\bar{z}} - \check{\Sigma}) + \hat{\lambda}_*^\top \hat{F}_{\bar{z}} \geq f_{\bar{z}} + \check{\lambda}_*^\top \check{F}_{\bar{z}} - |\check{\lambda}_*^\top| + \hat{\lambda}_*^\top \hat{F}_{\bar{z}} \geq 0$$

to obtain for each definite $\sigma \succ \sigma_*$ the required nonnegative Lagrange multiplier vector for the inequalities. Finally, if $\check{Z}$ had a null vector $\check{d} \neq 0$ then it follows by the Schur complement argument that there exists a nontrivial null vector d of the Jacobian of all equality and inequality constraints given by

$$\hat{\check{J}}_\sigma \equiv \begin{bmatrix} \check{F}_x & \check{F}_{\hat{z}} & \check{F}_{\bar{z}} - \check{\Sigma} \\ \hat{F}_x & \hat{F}_{\hat{z}} - I & \hat{F}_{\bar{z}} \\ 0 & 0 & I_{|\alpha|} \end{bmatrix} \in \mathbb{R}^{s \times (n+s)} \; ,$$

which would be also a feasible direction for all branch problems. Now Farkas' Lemma would yield $(f_x \; f_{\hat{z}} \; f_{\bar{z}})d = (\check{\lambda}_*^\top \; \hat{\lambda}_*^\top \; \mu_\sigma^\top)\hat{\check{J}}_\sigma d = 0$ so that $x_* = 0$ could not be a strict minimizer. Otherwise $v = \hat{\check{J}}_\sigma d$ must have for any feasible direction d at least one positive component so that $(f_x \; f_{\hat{z}} \; f_{\bar{z}})d = (\check{\lambda}_*^\top \; \hat{\lambda}_*^\top \; \mu_\sigma^\top)v > 0$ due to the strict positivity of μ_σ that follows from Eq. (43). Hence, x_* must be a strict minimizer under the given assumptions on $\check{Z}$ and the Lagrange multiplier vector λ_*. □

The natural progression from the first order necessary conditions above, which in the sharp case can already be sufficient, is to consider second order necessary conditions. However, so far we have not been able to develop them in a reasonably concise form, and it is not clear whether the assumption of MFKQ is really helpful in this context. Hence we restrict ourselves for the time being to second order sufficiency conditions, which like in the smooth case do not require any kink or constraint qualifications.

We consider the representation given by Eq. (13) of the branch problems. With μ_σ and λ_σ the Lagrange multipliers defined by the first order conditions for each definite signature vector σ we can formulate in the localized case the corresponding Lagrange functions

$$\mathcal{L}_\sigma(x, \bar{z}, \lambda_\sigma, \mu_\sigma) = f(x, \bar{z}) + \lambda_\sigma^\top (F(x, \bar{z}) - \Sigma \bar{z}) - \mu_\sigma^\top \bar{z} \; .$$

The Jacobian of all equality and inequality constraints is given by

$$\hat{\bar{J}}_\sigma \equiv \begin{bmatrix} \bar{J}_\sigma \\ \hat{J}_\sigma \end{bmatrix} \equiv \begin{bmatrix} Z & L - \Sigma \\ 0 & I_s \end{bmatrix} \in \mathbb{R}^{2s \times (n+s)} \; . \tag{46}$$

As can be seen, the Jacobian of the inequality constraints always has full rank due to the identity matrix I_s. Furthermore, the null space of $\hat{\bar{J}}_\sigma$ is given by $\hat{U} = [U^\top \; 0]^\top \in \mathbb{R}^{(n+s)\times(n-r)}$, where $U \in \mathbb{R}^{n \times (n-r)}$ represents a basis of the null space of Z as defined in Cor. 2.5 that has rank r. Here, it is important to note that this null space representation does not depend on σ.

THEOREM 4.3 (Sufficient second order optimality conditions).
Let $\varphi : \mathbb{R}^n \mapsto \mathbb{R}$ be a C^2_{abs} function and $\Delta\varphi(x_*; \Delta x)$ be minimal at $\Delta x = 0$ so that Theorem 4.1 applies. Assume that x_* is localized. If a λ_* satisfying Eq. (39) exists then φ attains at x_* a strict minimum provided we have

$$v^\top H_* v > 0 \qquad with \qquad H_* \equiv \begin{bmatrix} f_{xx} + \left(\lambda_*^\top F\right)_{xx} & f_{xz} + \left(\lambda_*^\top F\right)_{xz} \\ f_{zx} + \left(\lambda_*^\top F\right)_{zx} & f_{zz} + \left(\lambda_*^\top F\right)_{zz} \end{bmatrix}$$

for all $v \in C_\sigma(x_*), v \neq 0$, elements of the critical cones

$$C_\sigma(x_*) \equiv \{v \in \mathbb{R}^{n+s} \mid \bar{J}_\sigma v = 0,\ e_i^\top \hat{J}_\sigma v = 0 \text{ for } \mu_{\sigma,i} > 0,\ e_i^\top \hat{J}_\sigma v \geq 0 \text{ for } \mu_{\sigma,i} = 0\}$$

with μ_σ defined by Eq. (40) for all definite σ. Finally, if Eq. (39) holds strictly then it is sufficient for optimality that $U^\top (f + (\lambda_*^\top F))_{xx} U \succ 0$ where U spans the null space of Z as defined in Corollary 2.5

Proof. Assume that $v^\top H_* v > 0$ holds for all $0 \neq v \in C_\sigma(x_*)$ where σ varies over all definite signature vectors. Consider now a fixed definite σ. Then, it follows from the sufficient second order optimality condition for smooth constrained optimization that x_* is a strict minimizer of the corresponding branch problem [14, Theorem 12.6]. Since this true for all branch problems, x_* must be a strict minimizer for the original nonsmooth problem.

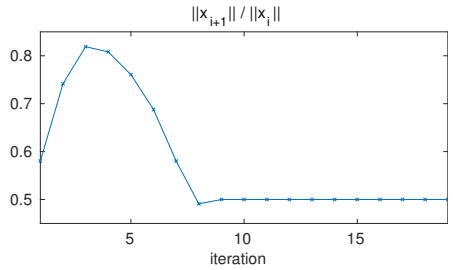
When Eq. (39) holds strictly, one has that strict complementarity holds, in that $\mu_\sigma > 0$ for all definite σ. Then the critical cones $C_\sigma(x_*)$ are all equal to the null space $\hat{U}$ of the Jacobian $\hat{\bar{J}}$ as defined in Eq. (46). Then, the final assertion follows form the representation of the Hessian H_* and of the structure of the null space basis $\hat{U} = [U^\top 0]^\top$. ∎

Notice that the requirement that $\hat{U}^\top H_* \hat{U} \succ 0$ is the second order sufficiency condition we already obtained in [9] under LIKQ, where λ_* is uniquely defined by tangential stationarity. Of course this second order condition can also be formulated for the nonlocalized case by performing the corresponding matrix and vector transformations.

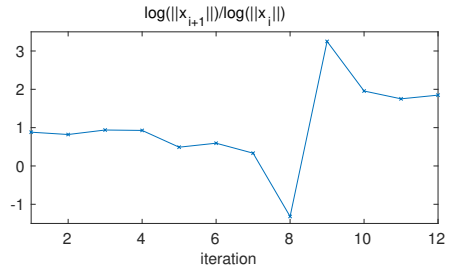
5. Numerical examples. One version called LiPsMin of the SPLOP algorithm discussed in Sec. 1 is analyzed in [4, 10]. Its C++ based optimization is used to generate the numerical results presented in this section.

For Exam. 2.7, we used $x_0 = (2, 3)$ as starting point. Figure 2a shows the ratio of the norm of two successive iterates. Here, we observe linear convergence with a contraction factor of 0.5 despite the fact that LIKQ does not hold. To minimize the nonconvex objective function given in Exam. 2.8, we chose $x_0 = (2, -2)$ as initial value. Figure 2b shows the ratio of the logarithm of the norm of two successive iterates. Hence, since the ratio reaches the value 2 for the last iterates this indicates local quadratic convergence. Here, it is interesting to note that the minimizer is not strict. These two example show that the convergence behavior of SPLOP is even better than expected from the theoretical results available so far.

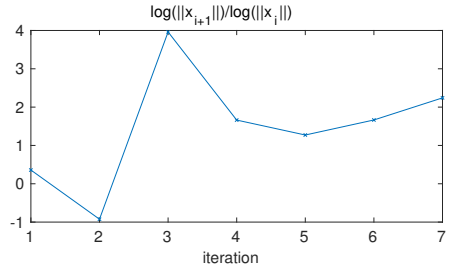
The objective function of the Exam. 2.9 does indeed have a sharp minimizer so we expect from the theory derived in this paper local quadratic convergence. For $n = 10$, we chose the initial point $x_0 = (2 \ldots 2) \in \mathbb{R}^{10}$ as stated in the literature, see [2]. Figure 2c shows the ratio of the logarithm of the norm of two successive iterates. Once more, we observe a value around 2 for this ratio at the final iterates indicating local quadratic convergence such that this example fits nicely to the theory developed in this paper.



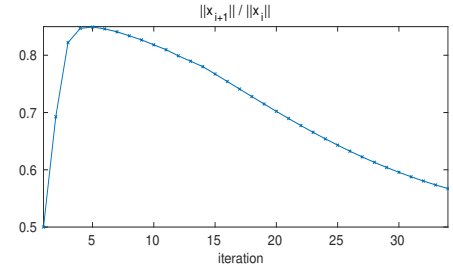
(a) Convergence history for $\varphi(x)$ as defined in Eq. (20)



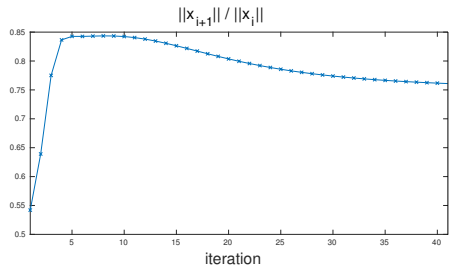
(b) Convergence history for $\varphi(x)$ as defined in Eq. (21) for $\varepsilon = 0.5$



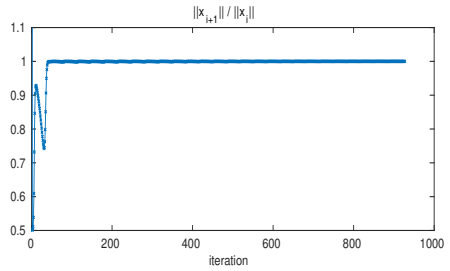
(c) Convergence history for $\varphi(x)$ as defined in Eq. (22) for $\varepsilon = 0.5$



(d) Convergence history for $\varphi(x)$ as defined in Eq. (23)



(e) Convergence history for $\varphi(x)$ as defined in Eq. (24)



(f) Convergence history for $\varphi(x)$ as defined in Eq. (25)

FIGURE 2. *Convergence history for example functions*

For the Exam. 2.10 and $n = 20$, we took the initial point $x_{0,i} = i$ for $i = 1, \ldots, 10$ and $x_{0,i} = -i$ for $i = 11, \ldots, 20$ as stated in the literature, see [2]. Figure 2d shows the ratio of the norm of two successive iterates. Here, we observe once more a linear rate of convergence despite the fact that that LIKQ does not hold.

The initial point to optimize the nonconvex objective function of Exam. 2.11 was chosen as $x_{0,i} = 2.0$ if i is even and $x_{0,i} = -1.5$ if i is odd according to the problem statement in the literature, see [2]. For this problem, LIKQ and SSC hold and therefore we can expect linear convergence. As illustrated by Fig. 2e this linear convergence can be observed with a contraction factor of about 0.75.

Finally we consider as one example of the LASSO problem the prostate cancer data set already used in [18] for the introduction of this method. Note that the application of SPLOP to this class of problems corresponds exactly to the proximal point method due to the problem structure. As shown in Lem. 3.4, we can expect

linear convergence and Fig. 2f confirms this convergence rate. However, the matrix $A^\top A$ which determines this convergence rate has a large condition number, namely $\kappa(A^\top A) = 63645.49$. Hence, the reduction factor is less than but very close to one.

6. Summary and Outlook. In this paper we analyzed for the optimization of C^d_{abs} objectives, which form a large class of piecewise smooth functions, the algorithm of successive piecewise linear optimization with a proximal term (SPLOP). We showed that SPLOP converges quadratically to sharp minimizers, where the function exhibits linear growth. Furthermore, under the Linear Independence Kink Qualification (LIKQ) and SSC with strict complementary, i.e. strict normal growth, we proved that SPLOP achieves a linear rate of convergence. Numerical examples based on a version of SPLOP called LiPsMin [10, 4] confirm these theoretical results.

Furthermore, without any kink qualifications we derived for the considered function class necessary and sufficient first order optimality conditions, specifically that local optimality of the nonlinear objective always requires local optimality of its piecewise linearization, and strict minimality of the latter is in fact equivalent to sharp minimality of the former. In this setting, it was also possible to state sufficient second order optimality conditions without any kink qualifications. These results are independent of the particular function representation, and allow in particular duplications of switching variables and other intermediates. So far it was not possible to derive reasonable necessary second order optimality conditions without LIKQ. Therefore, this will be the subject of further research.

With respect to the algorithmic development the approach to solve the local model problems must be improved to guarantee that all cluster points satisfy the first order minimality conditions given in Theo. 4.1. Also our code LiPsMin can solve the sequence of similar problems more efficiently by using warm starts as well as exploiting sparsity and other structure. A very promising prospect is the generalization of the identity matrix defining the proximal term to an approximation of the Hessian of the Lagrangian in order to achieve superlinear convergence.

Acknowledgement. The authors would like to thank Florian Jarre for discussions on second order optimality conditions.

REFERENCES

[1] J. Abadie. On the Kuhn-Tucker theorem. Nonlinear Programm., NATO Summer School Menton 1964, 19-36, 1967.

[2] A. Bagirov, N. Karmitsa, and M. Mäkelä. *Introduction to nonsmooth optimization. Theory, practice and software*. Springer, 2014.

[3] W. de Oliveira and C. Sagastizábal. Bundle methods in the XXIst century: A birds'-eye view. *Pesquisa Operacional*, 34(3):647–670, 2014.

[4] S. Fiege, A. Walther, and A. Griewank. An algorithm for nonsmooth optimization by successive piecewise linearization. Technical Report SPP1962-007, Universität Paderborn, 2016. available at optimization-online.

[5] A. Griewank. Automatic directional differentiation of nonsmooth composite functions. In *Recent developments in optimization. 7th French-German conference on optimization, Dijon, France, June 27-July 2, 1994*, pages 155–169. Springer, 1995.

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

[7] 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.

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

[9] 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.

[10] 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.
[11] F. Jarre and J. Stoer. *Optimierung.* Springer, 2004.
[12] 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.
[13] Y. Nesterov. Lexicographic differentiation of nonsmooth functions. *Mathematical Programming Series A*, 104(2-3):669–700, 2005.
[14] J. Nocedal and S. Wright. *Numerical Optimization.* Springer Series in Operations Research. Springer, 2006.
[15] J.M. Ortega and W.C. Rheinboldt. *Iterative solution of nonlinear equations in several variables.* Academic Press, 1970.
[16] S. Scholtes. *Introduction to Piecewise Differentiable Functions.* Springer, 2012.
[17] N.Z. Shor. *Nondifferentiable optimization and polynomial problems.* Kluwer, 1998.
[18] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B*, 58(1):267–288, 1996.
[19] A. Walther and A. Griewank. Characterizing and testing subdifferential regularity for piecewise smooth objective functions. Technical Report SPP1962-038, Universität Paderborn, 2017. available at optimization-online.