RELAXING KINK QUALIFICATIONS AND PROVING CONVERGENCE RATES IN PIECEWISE SMOOTH OPTIMIZATION[*]

ANDREAS GRIEWANK[†] AND ANDREA WALTHER[‡]

Abstract. In the paper [*Optim. Methods Softw.*, 31 (2016), pp. 904–930] we derived first order (KKT) and second order (second order sufficiency condition (SOSC)) 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. Firstly, we show here that under LIKQ and SOSC with strict complementarity, the natural algorithm of successive abs-linear minimization with a proximal term (SALMIN) achieves a linear rate of convergence. A version of SALMIN called LiPsMin has already been implemented and tested in [*Math. Program.*, 158 (2016), pp. 383–415; *Math. Program.* (2018), https://doi.org/10.1007/s10107-018-1273-5]. 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 SALMIN 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.

Key words. abs-normal form, sharp minimizer, linear independence kink qualification, LIKQ, Karush–Kuhn–Tucker, KKT, second order sufficiency condition, SOSC, strict complementary, linear convergence, quadratic convergence

AMS subject classifications. 49J52, 90C56

DOI. 10.1137/17M1157623

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 [12], 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 at least $d > 0$ times continuously differentiable, we now also allow piecewise linear elemental functions like abs, min, and max as part of the mix. In this way, we arrive at a class of piecewise smooth functions [23] that may be denoted by C_{abs}^d and contains most nonsmooth test functions from the literature. Its members 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 represents 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 [9], where we also show that

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

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 defined in [23].

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 way of linear algebra tests on the so-called abs-linear 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 [13] 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 the linear independence kink qualification (LIKQ), which is generic [10], i.e., always satisfiable by arbitrary small perturbations of a given abs-linear 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 Mangasarian–Fromovitz constraint qualification (MFCQ), which naturally generalizes to the Mangasarian–Fromovitz kink qualification (MFKQ) in our nonsmooth setting. The corresponding, more general, optimality conditions are developed to obtain the implications

$$(2) \qquad \text{MIN} \underset{(\text{POSC})}{\overset{\Longrightarrow}{\Longleftarrow}} \text{FOM}.$$

Here, FOM stands for first order minimal, which is equivalent to *criticality* in the sense of 0 being a Fréchet subdifferential of the Lipschitz function φ (see, e.g., [2]). Of course, FOM implies Clarke stationarity (CS) in the sense of 0 being contained in the larger Clarke generalized gradient $\partial_C \varphi$. It should be noted that for many nonsmooth optimization algorithms, the cluster points of the iterates generated can only be guaranteed to be CS. Naturally, in the convex case FOM and CS are both equivalent to MIN, but throughout this paper we will focus on the nonconvex case.

The upper implication in (2) is fairly obvious and has been proven formally in [7, 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 [16, Chapter 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 in some sense weakest constraint qualification defined by Guignard [15] holds. This implies also that the Abadie constraint qualification [1] is not satisfied so that the tangent cone of the nonlinear problem at its minimizer does not coincide with the tangent cone of the 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

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

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 optimizing 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 $x_2 = \tfrac{1}{3} x_1^3$. 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 $x_2 = \tfrac{1}{3} x_1^3$. 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 [9]. $\square$

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 (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 piecewise smooth 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 [23]. For larger Δx the piecewise linearization $\Delta(\mathring{x}; \Delta x)$ is in general nonhomogeneous. Indicating sharpness by a leading S and isolation by a leading I, we obtain the equivalences

$$\text{SMIN} \quad \Longleftrightarrow \quad \text{SFOM} \quad \Longleftrightarrow \quad \text{IFOM}.$$

In Theorem 4.1 we 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 FOM we have a gap between necessary and sufficient conditions, whose verification is simple under LIKQ but potentially combinatorial otherwise.

Successive abs-linear optimization. In view of relation (1) a natural algorithmic strategy, to be called successive abs-linear minimization with a proximal term (SALMIN), is to successively update

$$(3) \qquad x_{k+1} = x_k + \arg\min_{\Delta x} \left\{ \Delta\varphi(x_k; \Delta x) + \tfrac{q}{2}\|\Delta x\|^2 \right\}.$$

Here the quadratic coefficient q should be large enough to bound the discrepancy between φ and its piecewise linearization. In practical algorithms, the proximal parameter q must be adjusted successively as, for example, proposed and analyzed in [7]. For the theoretical analysis in this paper we will just assume that q has been chosen sufficiently large, which is customary for penalty parameters.

The iteration (3) was originally proposed in [9] and shown to generate a sequence of iterates $\{x_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ whose cluster points are FOM. As we will see, the rates of convergence established in this paper are closely related to those achieved by the proximal point iteration [2]

$$(4) \qquad x_{k+1} = x_k + \arg\min_{\Delta x} \left\{ \varphi(x_k + \Delta x) + \tfrac{\Delta q}{2}\|\Delta x\|^2 \right\}.$$

Here, Δq can be any positive constant or vary within some interval. The practical challenge for the proximal point concept is that the inner problem of minimizing the right-hand side of (4) seems almost as hard as the direct minimization of φ. Hence, one may also view SALMIN as an approximate proximal point method. Even closer

is the relation to quadratic or cubic overestimation, as proposed for example in [11], [20], and [5].

If the inner problem of minimizing the regularized piecewise linear model is not solved exactly, but one accepts increments Δx_k that reduce the objective and are simply Clarke stationary for $\Delta\varphi$, then the cluster points are guaranteed to be also Clarke stationary as shown in [7]. 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 $x_* \in \mathbb{R}^n$,

$$(5) \quad \Delta\varphi(x_k; x_{k+1} - x_k) - \Delta\varphi(x_k; x_* - x_k) \leq \tfrac{q}{2}(\|x_k - x_*\|^2 - \|x_{k+1} - x_k\|^2).$$

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 order term on the right-hand side of* (1) *is bounded by $\frac{\gamma}{2}\|\Delta x\|^2$ then the SALMIN method as described above with a proximal parameter $q > \gamma$ converges quadratically to x_* from all x_0 in some ball $B_\rho(x_*)$.*

Proof. Starting with the sharpness, we derive from (5) that

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

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 SALMIN 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 and bundle methods generally only obtain a harmonic convergence rate close to $1/\sqrt{k}$ or $1/k$ [24]. However, for structured problems and in the convex case, a superlinear rate has been established for one particular bundle approach in [18]. Here, we will prove a linear convergence rate under the assumption of SOSC with strict complementarity under LIKQ. These combined assumptions are quite strong. Essentially, they ensure that eventually all active kinks causing the nonsmoothness at the minimizers are identified so that SALMIN reduces locally to sequential quadratic optimization (SQP) 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 seven 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 SALMIN under LIKQ. In section 4, we generalize the optimality conditions of [13] to 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 [19] but no longer piecewise smooth.

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

$$(6) \qquad\qquad\qquad z = F(x, |z|),$$

$$(7) \qquad\qquad\qquad y = f(x, |z|),$$

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 [8] 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* (6), (7) *with $f, F \in C^d(\overline{\mathcal{D}} \times \overline{\mathbb{R}_+^s})$ is denoted by $C_{\mathrm{abs}}^d(\overline{\mathcal{D}})$.*

Recall that $C^d(\overline{\Omega})$ is the set of functions that possess continuous dth 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 $\varphi \in C_{\mathrm{abs}}^d(\overline{\mathcal{D}})$ 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

$$(8) \qquad Z \equiv \frac{\partial}{\partial x} F(x, |z|) \in \mathbb{R}^{s \times n}, \quad a = \frac{\partial}{\partial x} f(x, |z|) \in \mathbb{R}^n, \quad b = \frac{\partial}{\partial|z|} f(x, |z|) \in \mathbb{R}^s.$$

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

$$(9) \quad y = y_0 + a^\top(x - x_0) + b^\top(|z| - |z_0|) \quad \text{with} \quad z = z_0 + Z(x - x_0) + L(|z| - |z_0|),$$

where $z_0 = z(x_0)$ and $y_0 = \varphi(x_0)$. 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 $\{-1, 0, 1\}^s$.

Exploiting $|z| \equiv \Sigma z$, one obtains for z the identity

$$(I - L\Sigma)z = z_0 - Zx_0 - L|z_0| + Zx \quad \text{and}$$
$$z = (I - L\Sigma)^{-1}(z_0 - Zx_0 - L|z_0| + Zx).$$

Notice that due to the strict triangularity of $L\Sigma$ the inverse of $(I - L\Sigma)$ is well defined and polynomial in the entries of L. Substituting the expression obtained for z into the equation for $y = \varphi(x)$ and differentiating with respect to x yields

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

where the last equality only holds if $\sigma \in \{-1, 1\}^s$ so that Σ is nonsingular and thus its own inverse. As observed already in [14] for the nonlinear case, g_σ is one element of the limiting gradient set

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

whose convex hull is the Clarke generalized gradient.

The signature vectors define the domains

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

as a decomposition of the argument space, where one has

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

and φ_σ is one of finitely many differentiable selection functions in the sense of Scholtes [23]. They can easily be extended to an open neighborhood of S_σ. 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

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

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_{abs}^d function. If all switching variables vanish for a given point x, i.e.,*

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

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

$$(13) \qquad z = F(x, \Sigma z)$$

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

$$(14) \qquad \nabla z^\sigma \equiv \frac{\partial}{\partial x} z^\sigma = (I - L\Sigma)^{-1} Z \in \mathbb{R}^{s \times n},$$

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 or even the Guignard constraint qualification. Therefore, we will now examine under what conditions the sets S_σ as defined in (11) 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

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

This partial ordering of the signature vectors has already been used in [14]. 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*

$$(15) \qquad \hat{S}_\sigma \equiv \{x \in \mathbb{R}^n \,|\, \sigma \succ \sigma(x)\} \supset \bar{S}_\sigma.$$

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

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

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

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

As observed in [13, section 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*

$$(16) \qquad \min f(x, \bar{z}) \quad \text{s.t.} \quad \Sigma \bar{z} = F(x, \bar{z}) \quad \text{and} \quad \bar{z} \geq 0 \quad \text{for} \quad 0 \notin \sigma \succ \sigma(x_*) \equiv \sigma_*,$$

where $\bar{z} \equiv \Sigma z$ irrespective of 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

$$(17) \qquad \min f(x, \bar{z}, \hat{z}) \quad \text{s.t.} \quad \check{\Sigma}\bar{z} = \check{F}(x, \bar{z}, \hat{z}), \quad \hat{z} = \hat{F}(x, \bar{z}, \hat{z}), \quad \bar{z} \geq 0,$$

where $\check{\Sigma} \in \{-1, 1\}^{|\alpha(x_*)|}$. Given the smooth vector function z^σ as defined for a fixed definite σ by (13), we can describe the feasible set $\hat{S}_\sigma$ based on inequality constraints, i.e.,

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

This yields the third formulation of the branch problems given by

$$(18) \qquad \min f_\sigma(x) \equiv f(x, \Sigma z^\sigma(x)) \quad \text{s.t.} \quad x \in \hat{S}_\sigma \quad \text{with} \quad 0 \notin \sigma \succ \sigma_*.$$

For the derivation of the optimality conditions in the following sections, we will employ all three representations 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 and also the Guignard constraint qualification do not hold in general.

Example 2.3. Based on Example 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 $\varphi(\cdot)$ 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 (18) 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

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

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

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

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

As dual cones for $\mathcal{F}_\sigma(x_*)$ and $T_\sigma(x_*)$ one obtains

$$\mathcal{F}_\sigma^*(x_*) = \left\{v \in \mathbb{R}^2 \mid v_1 = 0,\, v_2 \in \mathbb{R}\right\} \quad \text{and} \quad T_\sigma^*(x_*) = \left\{v \in \mathbb{R}^2 \mid v_1 \geq 0,\, v_2 \in \mathbb{R}\right\}.$$

It follows that $\mathcal{F}_\sigma^*(x_*) \neq T_\sigma^*(x_*)$ such that the Guignard constraint qualification does not hold either.

Now, we examine when the constraint qualifications 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 (14), the Jacobian of all constraints is given by

$$(19) \qquad \Sigma \nabla z^\sigma = \Sigma(I - L\Sigma)^{-1}Z = (\Sigma - L)^{-1}Z \in \mathbb{R}^{s \times n},$$

where we have again used the invertibility of $\Sigma = \Sigma^{-1}$ due to the definiteness of σ. We proved in the companion paper [26] that the Jacobian of the active constraints alone 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 (18). For $\mathring{x} \in \mathbb{R}^n$, the Jacobian of the constraints that are active at $\mathring{x}$ is given by*

$$(20) \qquad J_\sigma \equiv (\sigma_i \nabla z_i^\sigma)_{i \in \mathring{\alpha}} = \check{\Sigma}(I - \check{L}\check{\Sigma})^{-1}\check{Z} = (\check{\Sigma} - \check{L})^{-1}\check{Z} \in \mathbb{R}^{|\mathring{\alpha}| \times n}$$

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 reduced 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

$$(21) \qquad \begin{aligned} &\check{Z} = (\mathring{P}\mathring{Z})_{i \in \mathring{\alpha}, 1 \leq j \leq n} \in \mathbb{R}^{|\mathring{\alpha}| \times n}, \quad \check{\Sigma} = (\mathring{P}\Gamma)_{i \in \mathring{\alpha}, j \in \mathring{\alpha}} \in \{-1, 1\}^{|\mathring{\alpha}| \times |\mathring{\alpha}|}, \quad \text{and} \\ &\check{L} = (\mathring{P}\mathring{L}\mathring{P})_{i \in \mathring{\alpha}, j \in \mathring{\alpha}} \in \mathbb{R}^{|\mathring{\alpha}| \times |\mathring{\alpha}|}. \end{aligned}$$

Also the following result, which is based on the observation $|\det(\check{\Sigma} - \check{L})| = 1$, was proven in [26].

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 Mangasarian–Fromovitz constraint qualification [17] for some $\sigma \succeq \mathring{\sigma}$ requires that

$$(22) \qquad J_\sigma v = (\check{\Sigma} - \check{L})^{-1}\check{Z}v > 0$$

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
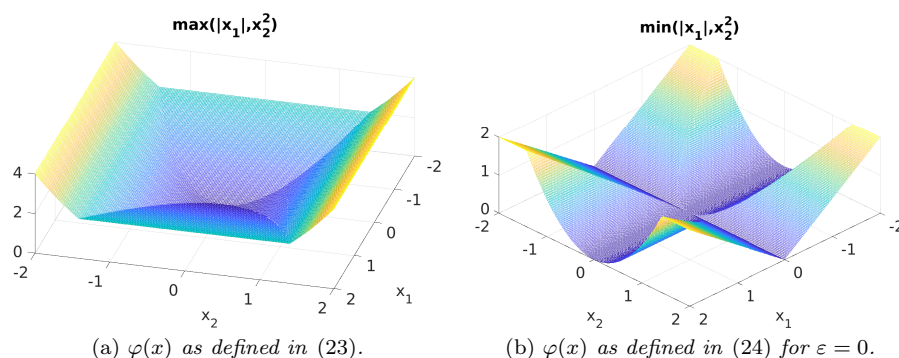
(a) $\varphi(x)$ *as defined in* (23).   (b) $\varphi(x)$ *as defined in* (24) *for* $\varepsilon = 0$.

Fig. 1. *Example functions from* (23) *and* (24).

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 [26].

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 (22) 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 [26] the following definition.

DEFINITION 2.6 (LIKQ and MFKQ). *For* $\varphi \in C^1_{\mathrm{abs}}(\overline{\mathcal{D}})$ *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 $[Z, L]$ [10]. While LIKQ just requires a rank determination for $\check{Z}$, we have so far not found a way to avoid the combinatorial effort of testing the weaker condition MFKQ for each branch problem defined by $\hat{\sigma} \succeq \mathring{\sigma}$. Indeed, we conjecture that MFKQ cannot be tested in polynomial time. However, so far we needed MFKQ just for the analysis of regularity (see the companion paper [26]) 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 section 4.

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

Example 2.7. First, we consider the convex function

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

see the left-hand side of Figure 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 $\varphi(\cdot)$ as defined in (23) 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 (20) 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 $R_{(1,-1)} = \{J_\sigma v \,|\, v \in \mathbb{R}^2\} = \{(v_1, -v_1)^\top \,|\, v_1 \in \mathbb{R}\}$ 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

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

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 nonconvex function

$$(24) \qquad \varphi : \mathbb{R}^2 \mapsto \mathbb{R}, \quad \varphi(x_1, x_2) = \min\left(|x_1| + \varepsilon|x_2|, x_2^2 + \varepsilon|x_1|\right)$$

for $\varepsilon \in \mathbb{R}$; see the right-hand side of Figure 1 for an illustration. As can be seen, for $\varepsilon = 0$, the point $x_* = 0$ is not a strict minimizer. For $\varepsilon > 0$ 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

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

one obtains for $\varphi(\cdot)$ as defined in (24) the abs-normal form

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

yielding

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

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

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

Hence, for $\varepsilon = 1$ the Jacobian J_σ at $x_* = 0$ does not have full column rank such that nontrivial 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 (24) at $x_* = 0$ is given by

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

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. Next, we consider the function

$$z = F(x) \equiv x \quad \text{and} \quad \varphi(x) = f(x, |z|) \equiv \tfrac{1}{2}(x^2 + |z|^2) = x^2,$$

which is already given in abs-normal form. Here, $x_* = 0$ is clearly the only minimizer and the two branch problems are given by

$$\min f(x, \bar z) \ \text{s.t.} \ \bar z = x,\ \bar z \geq 0, \qquad \text{and} \qquad \min f(x, \bar z) \ \text{s.t.} \ \bar z = -x,\ \bar z \geq 0,$$

which are equivalent due to the special nature of the objective function. The Jacobians of the equality constraint and the active inequality constraint are 2×2 triangular matrices with determinant ± 1. Hence, we have LICQ, but since the gradient of the objective vanishes at the minimizer, the two Langrange multipliers must vanish. Consequently, strict complementarity is violated and the Hessian of the Lagrangian is just that of f, namely the identity. It is positive definite in all directions and, in particular, for the critical directions of the branch problems, namely $(\Delta x, \Delta \bar z) = (1, 1)$ and $(\Delta x, \Delta \bar z) = (-1, 1)$, respectively. Hence, second order sufficiency in the sense of Nocedal and Wright is satisfied.

The piecewise linearization at the point x is given by

$$\Delta\varphi(x; \Delta x) = x\Delta x + |x|(|x + \Delta x| - |x|)$$

with the approximation error

$$\big|(x + \Delta x)^2 - x^2 - x\Delta x - |x||x + \Delta x| + x^2\big|$$
$$= \big|x\Delta x + (\Delta x)^2 - |x||x + \Delta x| + x^2\big| \ \leq \ 4(\Delta x)^2.$$

Here the last inequality can be easily checked when $x + \Delta x$ has the same sign as x and also when the signs differ and hence $|x| \leq |\Delta x|$.

Example 2.10 (chained CB3 I [3]). The scalable chained CB3 I function is given by

$$(25) \quad \varphi : \mathbb{R}^n \mapsto \mathbb{R}, \varphi(x) = \sum_{i=1}^{n-1} \max\{x_i^4 + x_{i+1}^2, (2 - x_i)^2 + (2 - x_{i+1})^2, 2e^{-x_i + x_{i+1}}\},$$

and hence is a piecewise smooth convex function. Here, one has $s = 2(n-1)$ and all switching variables are active at the solution $x_* = (1 \cdots 1)^\top \in \mathbb{R}^n$. Therefore, once more, LIKQ cannot 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 SALMIN.

Example 2.11 (Maxq [3]). The scalable Maxq function is defined as

$$(26) \qquad \varphi : \mathbb{R}^n \mapsto \mathbb{R}, \quad \varphi(x) = \max_{1 \leq i \leq n} x_i^2$$

and therefore also a piecewise smooth convex function. One has $s < n$ 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 2.12 (chained Crescent I [3]). The scalable chained Crescent I function defined as

$$(27) \quad \varphi : \mathbb{R}^n \mapsto \mathbb{R}, \quad \varphi(x) = \max \{\varphi_1(x), \varphi_2(x)\},$$

$$\varphi_1(x) = \sum_{i=1}^{n-1} (x_i^2 + (x_{i+1}-1)^2 + x_{i+1} - 1), \ \varphi_2(x) = \sum_{i=1}^{n-1} (-x_i^2 - (x_{i+1}-1)^2 + x_{i+1} + 1),$$

is a piecewise smooth but nonconvex function with the isolated but not sharp minimizer $x_* = (0 \cdots 0)^\top \in \mathbb{R}^n$. Here, one has $s = 1$,

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

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

Example 2.13 (LASSO problem). In statistics and machine learning, the so-called least absolute shrinkage and selection operator (LASSO) introduced in [25] 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

$$(28) \qquad \varphi : \mathbb{R}^n \mapsto \mathbb{R}, \qquad \varphi(x) = \tfrac{1}{m} \|w - Ax\|_2^2 + \beta \|x\|_1$$

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-linear form at $\mathring{x}$, one obtains for the Lagrangian form

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

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 [26], we analyzed an example where LIKQ does not hold but MFKQ does.

3. Linear convergence order under LIKQ. As in [13] 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

$$a^\top(x_*, 0) + \lambda_*^\top Z(x_*, 0) = 0 \qquad \text{(tangential stationarity)},$$
$$F(x_*, 0) = 0 \qquad \text{(full kink activity), and}$$
$$(29) \qquad b^\top(x_*, 0) + \lambda_*^\top L(x_*, 0) \geq |\lambda_*|^\top \qquad \text{(normal growth)}.$$

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

$$(30) \qquad \begin{bmatrix} H & Z^\top \\ Z & 0 \end{bmatrix} \in \mathbb{R}^{(n+s)\times(n+s)}$$

with

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

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

$$(31) \qquad \mathcal{L}(x, 0, \lambda) = f(x, 0) + \lambda^\top F(x, 0)$$

with respect to x. The saddle point Jacobian is nonsingular provided we have SOSC in that $U^\top H U \succ 0$, where $U = \check{U}$ as in Corollary 2.5. If we also have strong normal growth in that (29) holds as a strict inequality then it follows immediately that x_* is a strict minimizer.

Isolation. From the nonsingularity of the saddle point matrix (30) 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 [13], 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

$$(32) \qquad 0 = f_{\hat{z}}(w) + \check{\lambda}^\top \check{F}_{\hat{z}}(w) + \hat{\lambda}^\top \hat{F}_{\hat{z}}(w) - \hat{\lambda}^\top.$$

We will now see that if SOSC with strict complementarity, i.e., strict normal growth, holds at x_*, then the conditions (32) 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 $w_* = (x_*, 0, 0)$,

$$(f_x(w_*)\, f_{\hat{z}}(w_*)) + (\,\check{\lambda}_*^\top\; \hat{\lambda}_*^\top\,) \left[\begin{array}{cc} \check{F}_{\hat{z}}(w_*) & \check{F}_{\bar{z}}(w_*) \\ \hat{F}_{\hat{z}}(w_*) & \hat{F}_{\bar{z}}(w_*) \end{array} \right] \geq \varepsilon e_s^\top + (|\check{\lambda}_*|^\top\; |\hat{\lambda}_*|^\top)$$

with $e_s = (1 \;\cdots\; 1)^\top \in \mathbb{R}^s$ 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

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

with $e_{|\sigma|} = (1 \;\cdots\; 1)^\top \in \mathbb{R}^{|\sigma|}$ 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 these 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 (32) from (33) 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 SOSC with strict complementarity under LIKQ for the C_{abs}^d 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 [13] 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

$$0 = (f_x(w_*)\, f_{\hat{z}}(w_*)) + (\,\check{\lambda}_*^\top\; \hat{\lambda}_*^\top\,) \left[\begin{array}{cc} \check{F}_x(w_*) & \check{F}_{\hat{z}}(w_*) \\ \hat{F}_x(w_*) & \hat{F}_{\hat{z}}(w_*) - I \end{array} \right] \in \mathbb{R}^{n+|\sigma|},$$

$$0 = \check{F}(w_*) \in \mathbb{R}^{|\alpha|}, \qquad \hat{z} = \hat{F}(w_*) \in \mathbb{R}^{|\sigma|}$$

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

that 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 (31). LIKQ and SOSC 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

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

In other words, we have partitioned the switching variables that are inactive at x into those that are active only at x_*, namely $\tilde{z}$, 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

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

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 (34) from the second block of (35), 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_*. $\qquad\square$

Now we can address the question concerning the rate at which SALMIN converges to minimizers satisfying the above isolation theorem.

Linear convergence. Under the assumption of Proposition 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 $\mathcal{N} = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_*) + \varepsilon\}$ containing no other first order minimal points. Moreover, we will assume that within $\mathcal{N}$ (1) holds

with a constant coefficient $\gamma/2$ and that the proximal parameter q introduced in (3) is bigger than γ so that for $x_k \in \mathcal{N}$,

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

That is, the iterates generated by SALMIN stay in the neighborhood of x_* and it was shown in [7] that all cluster points of the sequence must be at the very least Clarke stationary. By Proposition 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 for the model problem, i.e., localized, must satisfy the FOM conditions

(36)
$$\begin{aligned} a_k + q(x_{k+1} - x_k) + Z_k^\top \lambda_{k+1} &= 0, \\ F(x_k, |z(x_k)|) + Z_k(x_{k+1} - x_k) &= 0, \\ b_k^\top + \lambda_{k+1}^\top L_k &\geq |\lambda_{k+1}|^\top. \end{aligned}$$

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 on the model problem it cannot have lower activity than x_* on the underlying abs-normal problem. In other words, the active kinks are properly identified by SALMIN in some vicinity of a fully active minimizer.

We can view the two block equations as a fixed point iteration and observe that by Proposition 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

(37)
$$\begin{aligned} \begin{bmatrix} x_{k+1} \\ \lambda_{k+1} \end{bmatrix} &= \begin{bmatrix} qI & Z_k^\top \\ Z_k & 0 \end{bmatrix}^{-1} \begin{bmatrix} qx_k - a_k \\ Z_k x_k - F_k \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{q}(I - Z_k^+ Z_k) & Z_k^+ \\ (Z_k^+)^\top & -q(Z_k Z_k^\top)^{-1} \end{bmatrix} \begin{bmatrix} qx_k - a_k \\ Z_k x_k - F_k \end{bmatrix} \\ &= \begin{bmatrix} x_k - \frac{1}{q}(I - Z_k^+ Z_k)a_k - Z_k^+ F_k \\ -(Z_k^+)^\top a_k + q(Z_k Z_k^\top)^{-1} F_k \end{bmatrix}, \end{aligned}$$

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 $\check{H} = H$ is defined as 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 (37) at all, we only have to consider the first block

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

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 SOSC is satisfied and q is 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 Corollary 2.5, and from [13] the reduced Hessian

$$(38)\qquad \check{H} = \begin{bmatrix} I & \hat{z}_x^\top \end{bmatrix}\begin{bmatrix} f_{xx} + (\check{\lambda}^\top \check{F})_{xx} & f_{x\hat{z}} + (\check{\lambda}^\top \check{F})_{x\hat{z}} \\ f_{\hat{z}x} + (\check{\lambda}^\top \check{F})_{\hat{z}x} & f_{\hat{z}\hat{z}} + (\check{\lambda}^\top \check{F})_{\hat{z}\hat{z}} \end{bmatrix}\begin{bmatrix} I \\ \hat{z}_x \end{bmatrix} \in \mathbb{R}^{n\times n},$$

where

$$(39)\qquad \hat{z}_x \equiv \frac{\partial \hat{z}}{\partial x} = (I - \hat{F}_{\hat{z}})^{-1}\hat{F}_x.$$

PROPOSITION 3.3. *Suppose x_* satisfies SOSC with strict complementarity under LIKQ for $\varphi(\cdot) \in C_{abs}^d$. Assume that $q > \max(\gamma, \|\check{U}_*^\top \check{H}_* \check{U}_*\|)$ holds for the proximal parameter q. Then SALMIN 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

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

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 SOSC 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

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

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

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

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 SALMIN iteration for the original problem and the reduced, localized problem are indeed identical. This completes the proof as it is well known [22] 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 Example 2.12) 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, SOSC holds at the minimizer.

Next, we will examine the LASSO problem in more detail to derive conditions for SOSC 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

$$(40) \qquad \check{Z} = (e_i^\top)_{i \in \alpha_*} \in \mathbb{R}^{|\alpha_*| \times n} \qquad \text{and} \qquad \check{U} = (e_i)_{i \notin \alpha_*} \in \mathbb{R}^{n \times |\sigma_*|},$$

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 SOSC 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 SOSC is fulfilled if $\check{U}^\top g_{xx}(x_*)\check{U}$ is positive definite.*

Proof. It follows directly from (40) that LIKQ is always fulfilled for the LASSO problem. Therefore, we only have to show the conditions for SOSC. Combining the formulation (18) and the Lagrangian form (28) one obtains

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

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 holds.   □

It follows from the last lemma that we can expect linear convergence of SALMIN 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.

As suggested by one of the referees, we will briefly examine the question of whether the R-linear convergence result can be obtained without the assumption of strict complementarity. There are indeed such results even for superlinear convergence of SQP methods, like, for example, Theorem 18.5 in [21]. However, it appears that these results apply mostly to the final equality constrained phase, where the active constraints have already been identified. On the other hand, the original papers on the identification for active constraints assumed strict complementarity as well; see, e.g., [4]. In more recent publications [6], more ingenious identification functions were proposed that do not rely on strict complementarity. However, SALMIN does not have this desirable feature, as can be seen from Example 2.9. For this example, we may choose $\gamma = 8$ in the second order term of (1). With a proximal parameter $q > 8$ the local model problem for this example is given by

$$\min_{\Delta x} \left\{ x\Delta x + |x|(|x + \Delta x| - |x|) + \tfrac{q}{2}(\Delta x)^2 \right\}.$$

For $\Delta x = 0$ the value of this function is zero and for $\Delta x = -x$ the value is $(\tfrac{q}{2} - 2)x^2 > 2x^2$. Hence, the minimizer cannot be attained at the kink but must be located in the smooth part. Differentiating with respect to Δx we get the stationarity condition $0 = x \pm x + q\Delta x$, so that $\Delta x = -(x \pm x)/q$. While $\Delta x = 0$ yields no reduction, we get for $\Delta x = -2 x/q$ the value $-2x^2/q - 2x^2/q + 2x^2/q = -2x^2/q < 0$. Thus we see

that $x + \Delta x = (1 - 2/q)x$, which means linear convergence at the rate $(1 - 2/q)$. In particular, the constraint $z = x = 0$ is never identified.

Hence, proving linear convergence without assuming strict complementarity requires a much deeper analysis as kink activity may change from step to step throughout the iteration. We will try to meet this challenge in a follow-up paper.

4. Optimality conditions without kink qualification. As demonstrated by Examples 2.7–2.11, 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) FOM). *Let $\varphi(x)$ be itself piecewise linear and given in the abs-linear form* (9). *Assume that the potential minimum x_* is shifted to the origin and localized in that all s kinks are active. Then, $\varphi(\cdot)$ 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*

$$(41) \qquad a^\top + (b - \mu_\sigma)^\top (\Sigma - L)^{-1} Z = 0.$$

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

$$(42) \qquad |\lambda_*|^\top \leq b^\top + \lambda_*^\top L \qquad for\ some \quad \lambda_* \in \Lambda,$$

then (41) *is automatically satisfied by*

$$(43) \qquad \mu_\sigma^\top = b^\top - \lambda_*^\top (\Sigma - L) \qquad for\ all \quad \sigma \in \{-1, 1\}^s.$$

Finally, the minimizer $x_* = 0$ can only be sharp if Z has full column rank n and it must be sharp if in addition (42) *holds as a strict inequality.*

Proof. First we show the forward implication. By Farkas's lemma [21, 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 (10) and (20) that

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

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's lemma.

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

$$(44) \qquad \mu_\sigma^\top = b^\top - \lambda_*^\top (\Sigma - L) \ \geq \ b^\top - |\lambda_*|^\top + \lambda_*^\top L$$

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_*. $\qquad\qquad\square$

Note that under LICQ we have a singleton $\Lambda = \{\lambda_*\}$ and (42) represents exactly the normal growth condition originally derived in [13]. Without the assumption of LIKQ, the set Λ will be an affine subspace of dimension one or more and (43) 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 [26]. It is much easier to check whether a suitable Λ_* for the sufficient condition (41) exists, namely one has to solve the LOP problem

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

in the variables $v \in \mathbb{R}^s$ and $\lambda_* \in \mathbb{R}^s$ for given $\varepsilon \geq 0$. 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 [13, Proposition 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-linear form (9). Assume that x_* is nonlocalized with the signature $\sigma_* \equiv \sigma(x_*)$ such that with $\alpha_* \equiv \alpha(x_*)$ one has $|\alpha_*| < s$. Then, $\varphi(\cdot)$ 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 $\mu_\sigma \in \mathbb{R}^{|\alpha_*|}$, $\mu_\sigma \geq 0$, such that*

$$(45) \quad (f_x \ f_{\hat{z}}) = -(\check{\lambda}_\sigma^\top \ \hat{\lambda}_\sigma^\top) \begin{bmatrix} \check{F}_x & \check{F}_{\hat{z}} \\ \hat{F}_x & \hat{F}_{\hat{z}} - I \end{bmatrix} \quad and \quad f_{\bar{z}} + (\check{\lambda}_\sigma^\top \ \hat{\lambda}_\sigma^\top) \begin{bmatrix} \check{F}_{\bar{z}} - \check{\Sigma} \\ \hat{F}_{\bar{z}} \end{bmatrix} = \check{\mu}_\sigma^\top \geq 0,$$

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

$$\check{a} \in \mathrm{range}\left(\begin{bmatrix} \check{F}_x & \check{F}_{\hat{z}} \\ \hat{F}_x & \hat{F}_{\hat{z}} - I \end{bmatrix}^\top \right), \quad i.e., \quad \check{a}^\top = \lambda^\top \begin{bmatrix} \check{F}_x & \check{F}_{\hat{z}} \\ \hat{F}_x & \hat{F}_{\hat{z}} - I \end{bmatrix}$$

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

$$(46) \qquad\qquad\qquad |\check{\lambda}_*^\top| \leq f_{\bar{z}} + \check{\lambda}_*^\top \check{F}_{\bar{z}} + \hat{\lambda}_*^\top \hat{F}_{\bar{z}},$$

then (45) is automatically satisfied by

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

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 (46) holds as a strict inequality.

Proof. To derive these results we use the formulation (17) of the branch problems and follow the lines of the previous theorem. Hence, applying Farkas's 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 $\check{\mu}_\sigma \in \mathbb{R}^{|\alpha_*|}$, $\check{\mu}_\sigma \geq 0$, such that

$$(47) \qquad f_x + \check{\lambda}_\sigma^\top \check{F}_x + \hat{\lambda}_\sigma^\top \hat{F}_x = 0, \qquad f_{\hat{z}} + \check{\lambda}_\sigma^\top \check{F}_{\hat{z}} + \hat{\lambda}_\sigma^\top (\hat{F}_{\hat{z}} - I) = 0, \qquad \text{and}$$

$$(48) \qquad f_{\bar{z}} + \check{\lambda}_\sigma^\top (\check{F}_{\bar{z}} - \check{\Sigma}) + \hat{\lambda}_\sigma^\top \hat{F}_{\bar{z}} = \check{\mu}_\sigma^\top \geq 0.$$

Combining the two equations in (47) 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 also be a feasible direction for all branch problems. Now Farkas's 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 (46). Hence, x_* must be a strict minimizer under the given assumptions on $\check{Z}$ and the Lagrange multiplier vector λ_*. $\qquad \square$

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 (16) 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

$$(49) \qquad \hat{\bar{J}}_\sigma \equiv \begin{bmatrix} \bar{J}_\sigma \\ \hat{J}_\sigma \end{bmatrix} \equiv \begin{bmatrix} Z & L - \Sigma \\ 0 & I_s \end{bmatrix} \in \mathbb{R}^{2s \times (n+s)}.$$

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 Corollary 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 (42) exists then φ attains at x_* a strict minimum provided we have*

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

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

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

with μ_σ defined by (43) for all definite σ. Finally, if (42) 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 [21, Theorem 12.6]. Since this is true for all branch problems, x_* must be a strict minimizer for the original nonsmooth problem.

When (42) 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 (49). Hence, the final assertion follows from the representation of the Hessian H_* and of the structure of the null space basis $\hat{U} = [U^\top 0]^\top$. $\qquad\square$

Notice that the requirement that $\hat{U}^\top H_* \hat{U} \succ 0$ is the second order sufficiency condition we already obtained in [13] 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 SALMIN algorithm discussed in section 1 is analyzed in [7, 14]. Its C++-based optimization is used to generate the numerical results presented in this section.

For Example 2.7, we used $x_0 = (2, 3)$ as the starting point. Figure 2 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 Example 2.8, we chose $x_0 = (2, -2)$ as the initial value. Figure 2 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
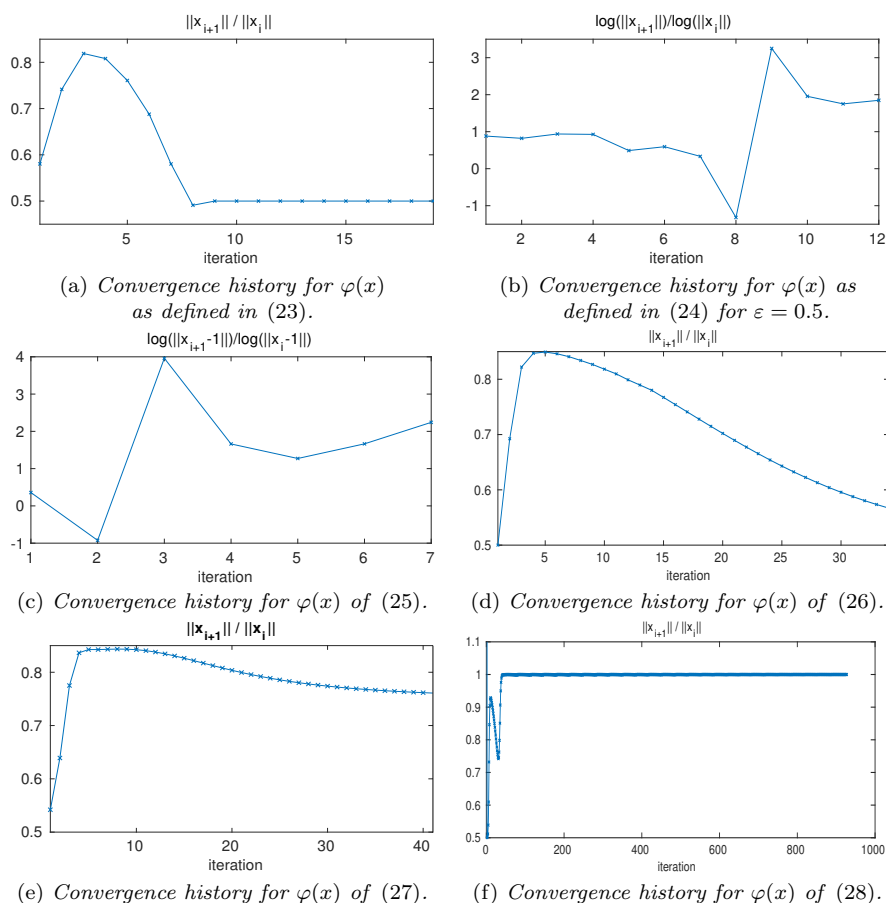
(a) *Convergence history for $\varphi(x)$ as defined in* (23).

(b) *Convergence history for $\varphi(x)$ as defined in* (24) *for $\varepsilon = 0.5$.*

(c) *Convergence history for $\varphi(x)$ of* (25).

(d) *Convergence history for $\varphi(x)$ of* (26).

(e) *Convergence history for $\varphi(x)$ of* (27).

(f) *Convergence history for $\varphi(x)$ of* (28).

FIG. 2. *Convergence history for example functions.*

local quadratic convergence. Here, it is interesting to note that the minimizer is not strict. These two examples show that the convergence behavior of SALMIN is even better than expected from the theoretical results available so far.

The objective function of Example 2.10 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 $x_0 = (2 \cdots 2) \in \mathbb{R}^{10}$, as stated in the literature; see [3]. Figure 2 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. For Example 2.11 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 [3]. Figure 2 shows the ratio of the norm of two successive iterates. Here, we observe once more a linear rate of convergence despite the fact that LIKQ does not hold. The initial point to optimize the nonconvex objective function of Example 2.12 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 [3]. For this problem, LIKQ and SOSC hold and therefore we can expect linear convergence. As illustrated by Figure 2, 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 [25] for the introduction of this method. Note that the application of SALMIN to this class of problems corresponds exactly to the proximal point method due to the problem structure. As shown in Lemma 3.4, we can expect linear convergence and Figure 2 confirms this convergence rate since four out of eight components at the solution are zero and the remaining four components attain small values. However, the matrix $A^\top A$ that 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_{abs}^d objectives, which form a large class of piecewise smooth functions, the algorithm of successive abs-linear minimization with a proximal term (SALMIN). We showed that SALMIN converges quadratically to sharp minimizers, where the function exhibits linear growth near its minimizer. Furthermore, under the linear independence kink qualification (LIKQ) and SOSC with strict complementary, i.e., strict normal growth, we proved that SALMIN achieves a linear rate of convergence. Numerical examples based on a version of SALMIN called LiPsMin [14, 7] 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 Theorem 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.

REFERENCES

[1] J. Abadie, *On the Kuhn–Tucker theorem*, in Nonlinear Programming, North-Holland, New York, 1967, pp. 19–36.

[2] H. Attouch and J. Bolte, *On the convergence of the proximal algorithm for nonsmooth functions involving analytic features*, Math. Program., 116 (2009), pp. 5–16.

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

[4] J. V. Burke and J. J. Moré, *On the identification of active constraints*, SIAM J. Numer. Anal., 25 (1988), pp. 1197–1211.

[5] C. Cartis, N. Gould, and P. Toint, *Adaptive cubic regularisation methods for unconstrained optimization. Part* I*: Motivation, convergence and numerical results*, Math. Program., 127 (2011), pp. 245–295.

[6] F. FACCHINEI, A. FISCHER, AND C. KANZOW, *On the accurate identification of active constraints*, SIAM J. Optim., 9 (1998), pp. 14–32.

[7] S. FIEGE, A. WALTHER, AND A. GRIEWANK, *An algorithm for nonsmooth optimization by successive piecewise linearization*, Math. Program. (2018), https://doi.org/10.1007/s10107-018-1273-5.

[8] A. GRIEWANK, *Automatic directional differentiation of nonsmooth composite functions*, in Recent Developments in Optimization, Lecture Notes in Econom. Math. Systems 429, R. Durier and C. Michelot, eds., Springer, Berlin, 1995, pp. 155–169.

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

[10] A. GRIEWANK, J.-U. BERNT, M. RANDONS, AND T. STREUBEL, *Solving piecewise linear equations in abs-normal form*, Linear Algebra Appl., 471 (2015), pp. 500–530.

[11] A. GRIEWANK, J. FISCHER, AND T. BOSSE, *Cubic overestimation and secant updating for unconstrained optimization of $C^{2,1}$ functions*, Optim. Methods Softw., 29 (2014), pp. 1075–1089.

[12] A. GRIEWANK AND A. WALTHER, *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, Other Titles Appl. Math., SIAM, Philadelphia, PA, 2008.

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

[14] A. GRIEWANK, A. WALTHER, S. FIEGE, AND T. BOSSE, *On Lipschitz optimization based on gray-box piecewise linearization*, Math. Program., 158 (2016), pp. 383–415.

[15] M. GUIGNARD, *Generalized Kuhn–Tucker conditions for mathematical programming problems in a Banach space*, SIAM J. Control, 7 (1969), pp. 232–241.

[16] F. JARRE AND J. STOER, *Optimierung*, Springer, Berlin, 2004.

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

[18] R. MIFFLIN AND C. SAGASTIZÁBAL, *A $\mathcal{VU}$-algorithm for convex minimization*, Math. Program., 104 (2005), pp. 583–608.

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

[20] Y. NESTEROV AND B. POLYAK, *Cubic regularization of Newton method and its global performance*, Math. Program., 108 (2006), pp. 177–205.

[21] J. NOCEDAL AND S. WRIGHT, *Numerical Optimization*, 2nd ed., Springer Ser. Oper. Res. Financ. Eng., Springer, New York, 2006.

[22] J. ORTEGA AND W. RHEINBOLDT, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

[23] S. SCHOLTES, *Introduction to Piecewise Differentiable Equations*, SpringerBriefs Optim., Springer, New York, 2012.

[24] N. Z. SHOR, *Nondifferentiable Optimization and Polynomial Problems*, Kluwer Academic Publishers, Norwell, MA, 1998.

[25] R. TIBSHIRANI, *Regression shrinkage and selection via the lasso*, J. R. Stat. Soc. Ser. B Stat. Methodol., 58 (1996), pp. 267–288.

[26] A. WALTHER AND A. GRIEWANK, *Characterizing and testing subdifferential regularity for piecewise smooth objective functions*, Optim. Online, 2017; available at http://www.optimization-online.org/DB_HTML/2017/11/6314.html.