An active signature method for constrained abs-linear minimization

Timo Kreimeier* Andrea Walther* Andreas Griewank*

April 28, 2023

In this paper we consider the solution of optimization tasks with a piecewise linear objective function and piecewise linear constraints. First, we state optimality conditions for that class of problems given in the so-called abs-linear form and prove that they can be verified in polynomial time. Subsequently, we propose an algorithm called Constrained Active Signature Method that explicitly exploits the piecewise linear structure to solve such optimization problems as the main contribution of this work. Convergence of the algorithm within a finite number of iterations is proven. Numerical results for various test cases including linear complementarity constraints and a bi-level problem illustrate the performance of the new algorithm.

Keywords: piecewise linear; constrained optimization; abs-linear form; linear independence kink qualification (LIKQ); optimality conditions; active signature method (ASM)

1 Introduction

Motivated by numerous applications, e.g., from machine learning, there has been a growing interest in optimization problems that lack differentiability. That is, the objective function and/or the constraints are not differentiable everywhere. One important class of such problems is given by piecewise linear functions, where corresponding optimization tasks arise, e.g., in train time tabling [5], as local models [22] or in the training of deep neural networks with the Rectified Linear Unit (ReLU) as activation function [6, 27].

So far, there is only a limited number of algorithms to solve constrained nonsmooth optimization problems available. Possible approaches comprise quasi-Newton methods [3], bundle methods [23] or extensions of Franke-Wolfe approaches [25]. They all have in common that they do not exploit the structure that is available in the nonsmooth

*Institute for Mathematics, Humboldt-Universität zu Berlin, Berlin, 10099, Germany.

setting. For unconstrained optimization problems with piecewise linear objective functions, the so-called Active Signature Method (ASM) for determining local minima has been proposed in [9]. This approach explicitly builds on the nonsmooth structure to verify corresponding optimality conditions that can be verified in polynomial time even if the target function is nonsmooth. In [14, 16], the more general setting of nonlinear, so-called abs-smooth constrained optimization problems was considered. To derive optimality conditions that can be verified in polynomial time, Hegerhorst-Schultchen and Steinbach reformulated the inequality constraints into equality constraints using slack variables. Furthermore, it has been shown that each abs-smooth nonlinear optimization problem has an equivalent formulation as mathematical program with equilibrium constraints (MPEC) [15]. In the same paper, an equivalence of the corresponding regularity conditions was shown. These are, on the one hand, the MPEC-LICQ and, on the other hand, the regularity conditions for abs-smooth problems, the so-called Linear Independence Kink Qualification (LIKQ) that will also be considered here.

In this paper, an extension of ASM will be presented, which, in addition to the piecewise linear objective function, also takes piecewise linear functions as equality and inequality constraints into account and exploits explicitly the nonsmooth structure of the optimization problem. To verfiy that an optimal point is reached by the algorithm, i.e., to define a suitable termination criteria, we will derive optimality conditions directly for this problem class providing an alternative proof for the piecewise linear case in comparison to [14, 16].

A piecewise linear function can always be given in its abs-linear form as introduced for the first time in [8]:

Definition 1.1 (Abs-linear form, switching vector)**.** A continuous piecewise linear function $\varphi : \mathbb{R}^n \to \mathbb{R}$ is in *abs-linear form* if $y \equiv \varphi(x)$ is given by

$$y = d + a^\top x + b^\top z \ , \tag{1a}$$

$$z = c + Zx + Mz + L|z| \ , \tag{1b}$$

with $x \in \mathbb{R}^n$ the argument vector, $z \in \mathbb{R}^s$ the vector of switching variables, called *switching vector*, and constants $d \in \mathbb{R}$, $a \in \mathbb{R}^n$, $b, c \in \mathbb{R}^s$, $Z \in \mathbb{R}^{s\times n}$, $L, M \in \mathbb{R}^{s\times s}$, where the last two matrices are strictly lower triangular. Eq. (1b) is called *switching system.*

Here and throughout, $|z|$ denotes the component-wise modulus of a vector z. Without loss of generality, we can always assume that $d = 0$. Using the reformulation

$$\max(x_1, x_2) = \frac{1}{2}(x_1 + x_2 + |x_1 - x_2|) \ , \quad \min(x_1, x_2) = \frac{1}{2}(x_1 + x_2 - |x_1 - x_2|) \tag{2}$$

and Prop. 2.2.2 of [26] it follows that every continuous piecewise linear function can be represented in an abs-linear form. In contrast to previous publications, e.g., [9, 11, 10], here the matrix M appears on the right side of Eq. (1b). One easily obtains the new matrix M from the previous notation by subtracting the unit matrix there. The reason for the adjustment is on one hand the coverage of more general formulations of the target function and on the other hand a unification with the constraints considered in this paper for the first time.

Using the signatures of the switching vector, it is possible to decompose $\mathbb{R}^n$ into polyhedra [7]. This decomposition plays an essential role in the algorithm we will present in this paper, because one of the main ideas of the algorithm will be the solution of suitable adapted smooth optimization problems on these polyhedra.

Definition 1.2 (Signature vector and signature matrix)**.** Let a piecewise linear function be given in an abs-linear form (1). For each $x \in \mathbb{R}^n$, we define the *signature vector*

$$\sigma(x) \equiv (\mathbf{sgn}(z_1(x)), \ldots, \mathbf{sgn}(z_s(x))) \in \{-1, 0, 1\}^s \ ,$$

where the set on the right hand side comprises all possible signature vectors. The corresponding *signature matrix* is given by $\Sigma(x) = \mathbf{diag}(\sigma(x))$. A signature vector $\sigma(x)$ is called *definite*, if no component vanishes, i.e., $\sigma(x) \in \{-1, 1\}^s$. This situation is denoted by $0 \notin \sigma(x)$. Otherwise it is called *indefinite*.

Since for Eq. (1b) the matrix L is assumed to be strictly lower triangular, $|z_s(x)|$ does not contribute to the value of the abs-linear objective function and hence does not impose any nonsmoothness. To simplify notation, we assume that the first $s-1$ components $z_1, \ldots, z_{s-1}$ are arguments of the absolute value such that they impose nonsmoothness. This fact has to be taken into account correspondingly for calculating a step size as described later, see Section 4. Since we will also consider frequently fixed signature vectors, we will state the dependence on x if there is any explicitly. Based on the signature vectors, it is possible to decompose the $\mathbb{R}^n$ into polyhedra as follows.

Definition 1.3 ((Extended) Signature domain)**.** For a fixed $\sigma \in \{-1, 0, 1\}^s$, we define

$$\mathcal{P}_\sigma \equiv \{x \in \mathbb{R}^n \mid \mathbf{sgn}(z(x)) = \sigma\} \subset \overline{\mathcal{P}}_\sigma \equiv \{x \in \mathbb{R}^n \mid \Sigma z(x) = |z(x)|\} \ .$$

The set $\mathcal{P}_\sigma$ is called *signature domain* and the set $\overline{\mathcal{P}}_\sigma$ *extended signature domain.*

Note that here and throughout the symbol $\subset$ denotes a subset relation that also allows equality of sets.

The domains $\mathcal{P}_\sigma$ are given as inverse images of the corresponding σ and represent a disjoint decomposition of $\mathbb{R}^n$ into relatively open polyhedra. The boundaries of the polyhedra $\mathcal{P}_\sigma$ are usually the sets where φ is nonsmooth. Motivated by the graphical representation in low dimensions as illustrated also in Ex. 3.1, the sets of points at which φ is nonsmooth are called kinks. As shown in [12] it is possible to define a partial ordering as follows

$$\sigma \prec \tilde{\sigma} \quad \iff \quad \sigma_l^2 \leq \tilde{\sigma}_l \sigma_l \quad \text{for } 1 \leq l \leq s \quad \iff \quad \overline{\mathcal{P}}_\sigma \subset \overline{\mathcal{P}}_{\tilde{\sigma}}.$$

For more details about the decomposition see for example [7] or [9].

Next, we state the optimization problem for which we will present and analyze a solution algorithm in this paper, i.e., the *constrained abs-linear optimization problem*

(CALOP). It has the following structure:

$$\begin{aligned}
&\min_{x\in\mathbb{R}^n, z\in\mathbb{R}^s} \quad a^\top x + b^\top z \\
\text{s.t.}\quad & 0 = g + Ax + Bz + C|z| \; , \\
& 0 \geq h + Dx + Ez + F|z| \; , \\
& z = c + Zx + Mz + L|z| \; ,
\end{aligned} \tag{CALOP}$$

where $g \in \mathbb{R}^m, h \in \mathbb{R}^p, A \in \mathbb{R}^{m\times n}, B, C \in \mathbb{R}^{m\times s}, D \in \mathbb{R}^{p\times n}$ and $E, F \in \mathbb{R}^{p\times s}$. Note that the component z_s may now introduce nonsmoothness. This is an important difference to the unconstrained case considered in earlier papers, e.g. [9, 10]. However, since we want to allow constraints that are as general as possible we do not impose further restrictions on the matrices C and F in the remainder of this paper. Alternatively, i.e., if the nonsmoothness of the constraints is also pushed into the switching equation, the value of s may increase significantly, depending on the number of constraints and their nonsmoothness leading to a significant increase of the run time of the Algo. 1 presented later. As can be seen, we assume that the objective function combined with the switching system in the last constraint is in abs-linear form, cf. Eq. (1). The first constraint in (CALOP) represents the equality constraint and the second one the inequality constraint. For later use, we define

$$\begin{aligned}
& f : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}, && (x, z) \mapsto a^\top x + b^\top z \; , \\
& G : \mathbb{R}^n \times \mathbb{R}^s \times \mathbb{R}^s \to \mathbb{R}^m, && (x, z, |z|) \mapsto g + Ax + Bz + C|z| \; , \\
\text{and}\quad & H : \mathbb{R}^n \times \mathbb{R}^s \times \mathbb{R}^s \to \mathbb{R}^p, && (x, z, |z|) \mapsto h + Dx + Ez + F|z| \; .
\end{aligned} \tag{3}$$

The paper is organized as follows. Section 2 briefly describes the ASM published in [9] to prepare the ground for the extensions derived in the present paper. Optimality conditions for constrained nonsmooth optimization problems of the form (CALOP) are derived in Section 3. For this purpose, the optimality conditions of the unconstrained case as introduced in [7] and discussed further in [10] are modified to take the feasibility with respect to the constraints into account. In Section 4, the ASM is extended to cover also piecewise linear constrained optimization problems of the form (CALOP) as one main contribution of the paper. The convergence analysis of the resulting new algorithm is also given in Section 4.2. This includes also a statement on finite convergence representing another important contribution of this paper. Numerical results for several test problems are presented in Section 5. Finally, the paper concludes with a summary and an outlook in Section 6.

2 The active signature method for unconstrained problems

In this section, the Active Signature Method (ASM) published in [9] is explained shortly, such that it can be extended subsequently to constrained problems of the form (CALOP). It should be noted that the notation has been adjusted in comparison to [9] because the abs-linear form is used in a slightly different representation, as motivated above.

The aim of the ASM is to determine a local minimum for a piecewise linear objective function $\varphi : \mathbb{R}^n \to \mathbb{R}$ given in abs-linear form (1). At this point we would like to mention that whenever it is not explicitly stated otherwise, the term minimum refers to a local minimum. To find such a minimum, the basic idea is to decompose the $\mathbb{R}^n$ into polyhedra as sketched above using the signature vectors and optimize a penalized version of the objective function on these domains, switching from one polyhedron to the next in an appropriate way. To achieve this behavior, for each $x \in \mathbb{R}^n$ and the corresponding $z = z(x)$, information about the structure of nonsmoothness is exploited. Using the abs-linear form of the objective function φ, we obtain the following equivalent *abs-linear optimization problem*

$$\begin{aligned} \min_{x\in\mathbb{R}^n, z\in\mathbb{R}^s} \quad & a^\top x + b^\top z \\ \text{s.t.} \quad & z = c + Zx + Mz + L|z| \;. \end{aligned} \tag{ALOP}$$

Since piecewise linear functions may be unbounded below, we add the term $\frac{1}{2}x^\top Qx$ with a positive definite matrix $Q = Q^\top \in \mathbb{R}^{n\times n}$ to the objective. Furthermore, we fix one signature vector $\sigma \in \{-1, 0, 1\}^s$ to obtain from (ALOP) the smooth quadratic optimization problem

$$\min_{x\in\mathbb{R}^n, z\in\mathbb{R}^s} \quad a^\top x + b^\top z + \frac{1}{2}x^\top Qx \tag{4a}$$

$$\text{s.t.} \quad z = c + Zx + Mz + L\Sigma z \;, \tag{4b}$$

$$0 = (I_s - |\Sigma|)z \;, \tag{4c}$$

$$0 \le \Sigma z \;, \tag{4d}$$

on $\overline{\mathcal{P}}_\sigma$, i.e., all $x \in \overline{\mathcal{P}}_\sigma$ are feasible. Here, I_s denotes the identity matrix in $\mathbb{R}^{s\times s}$. Due to the penalty term, it is ensured that a global minimum exists on $\overline{\mathcal{P}}_\sigma$. As shown in [9] applying standard KKT theory for the smooth constrained quadratic optimization problem Eq. (4) yields the following system of necessary optimality conditions

$$\begin{bmatrix} Q & 0 & Z^\top \\ 0 & I_s - |\Sigma| & \Sigma(M^\top + \Sigma L^\top - I_s) \\ Z & M + L\Sigma - I_s & 0 \end{bmatrix} \begin{bmatrix} x \\ z \\ \lambda \end{bmatrix} = - \begin{bmatrix} a \\ \Sigma b \\ c \end{bmatrix} \;. \tag{5}$$

Due to its consistency, the system of equations has a solution, which in general does not necessarily have to be unique. A solution of Eq. (5) is denoted by $(\hat{x}, \hat{z}, \lambda)$, where λ represents the Lagrange multiplier associated with the equality constraint (4b). As described also in [9], the system (5) can be solved efficiently using the special structure of the triangular matrices L and M. However, Eq. (5) ignores the inequality constraints (4d) of the optimization problem (4). Hence, the resulting $\Sigma\hat{z}$ may have negative components. In this case, there must be a so called blocking constraint or more specifically a blocking kink on the line segment from the current iterate (x, z) to the infeasible point $(\hat{x}, \hat{z})$. With $\hat{z}$ part of the solution of the system given in Eq. (5) and z the current iterate, this situation can be easily detected by calculating the maximal step size β^z as

$$\beta^z = \inf_{1\le l\le s} \left\{ \beta^z_l \equiv \frac{-z_l}{\hat{z}_l - z_l} \,\middle|\, (\hat{z}_l - z_l)\sigma_l < 0 \right\} \in \; (0, \infty] \tag{6}$$

with $\inf \emptyset \equiv \infty$. Note that in comparison to [9] we replace the factor z_l by σ_l in the test of the sign. Since only the sign is important the formulation in Eq. (6) is equivalent to the original formulation but numerically more stable.

If $\beta^z < \infty$ the first index for which the minimum is attained is denoted by j^z. For $\beta^z \leq 1$, there exists at least one blocking constraint and the next iterate is given by

$$x^+ = (1 - \beta^z)x + \beta^z \hat{x} \quad \text{and} \quad z^+ = (1 - \beta^z)z + \beta^z \hat{z} \ .$$

The part x^+ lies on the boundary of $\mathcal{P}_\sigma$, i.e., at least one component of $\sigma(x^+)$ drops to zero in comparison to σ. Therefore, one updates $\sigma^+ = \sigma - \sigma_{j^z} e_{j^z}$ setting $\sigma^+_{j^z} = 0$, which amounts to the activation of a kink. In other words the feasible domain of the optimization problem (4) is effectively reduced to the face polyhedron $\overline{\mathcal{P}}_{\sigma^+} \subset \overline{\mathcal{P}}_\sigma$. This can be seen as restricting one component of the signature vector. After finitely many such kink activations one must have $\beta^z = 1$ so that the full step reaches the unique minimizer x_σ within the current $\mathcal{P}_\sigma$. For more details on this, see [9] and for the handling of constraints Section 4 in this article.

Definition 2.1 (Signature optimal point)**.** Let an optimization problem of the form (ALOP) be given. Consider a fixed signature vector $\sigma \in \{-1, 0, 1\}^s$. A minimizer $x_\sigma \in \mathcal{P}_\sigma$ of the optimization problem

$$\begin{aligned} \min_{x \in \mathbb{R}^n, z \in \mathbb{R}^s} \quad & a^\top x + b^\top z && \text{(7a)} \\ \text{s.t.} \quad & z = c + Zx + Mz + L\Sigma z \ , && \text{(7b)} \\ & 0 = (I_s - |\Sigma|)z \ , && \text{(7c)} \\ & 0 \leq \Sigma z \ , && \text{(7d)} \end{aligned}$$

is called *signature optimal point* of the original, unconstrained optimization problem (ALOP).

Note that for many or even most σ the polyhedra $\mathcal{P}_\sigma$ do not contain minimizers, in which case the solutions of (7) lie on their relative boundary. The piecewise linear and convex function $\varphi : \mathbb{R}^2 \to \mathbb{R}$,

$$\varphi(x) = \max\{\max\{-100, 2x_1 + 5|x_2|\}, 3x_1 + 2|x_2|\} \tag{8}$$

considered already by Hiriart-Urruty and Lemaréchal in [17] will be used to illustrate this observation. In contrast to the original formulation given in [17] the representation given in Eq. (8) results in only four switching variables:

$$\begin{aligned} z_1 &= x_2 \ , & z_2 &= -100 - 2x_1 + 5|z_1| \ , \\ z_3 &= -50 - 4x_1 + 0.5|z_1| + 0.5|z_2| \ , & z_4 &= 2.25|z_1| + 0.25|z_2| + 0.5|z_3| \ , \\ y &= -25 + x_1 + z_4 \ . \end{aligned}$$

Fig. 1 shows the decomposition of the $\mathbb{R}^2$ in the different polyhedra $\mathcal{P}_\sigma$ with definite signature vectors. The lines correspond to polyhedra with indefinite signature vectors,

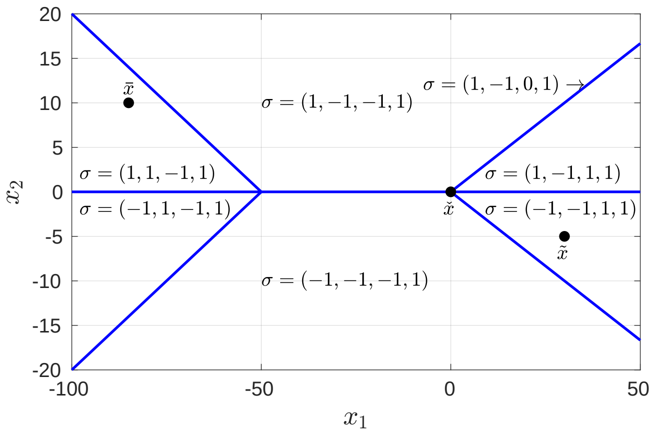


Figure 1: Signature domains and (non) signature optimal points

where we mark one of them to give an example. As can be seen, the point $\tilde{x} = (30, -5)$ has the signature vector $\sigma(\tilde{x}) = (-1,-1,1,1) \equiv \sigma$. When optimizing over $\overline{\mathcal{P}}_\sigma$ one obtains the minimizer $\check{x} = (0,0)$ with the signature vector $\sigma(\check{x}) = (0,-1,0,1) \neq \sigma$. Hence, one has $\check{x} \notin \mathcal{P}_{(-1,-1,1,1)}$ but $\check{x}$ is signature optimal on the polyhedron $\mathcal{P}_{(0,-1,0,1)}$ that contains only $\check{x}$. The function $\varphi(.)$ is constant on the polyhedra $\mathcal{P}_{(1,1,-1,1)}$ and $\mathcal{P}_{(-1,1,-1,1)}$. Hence, $\bar{x} = (-85, 10)$ with the signature $\sigma(\bar{x}) = (1,1,-1,1)$ is a minimizer on the polyhedron $\mathcal{P}_{(1,1,-1,1)}$. Therefore, $\bar{x}$ is signature optimal.

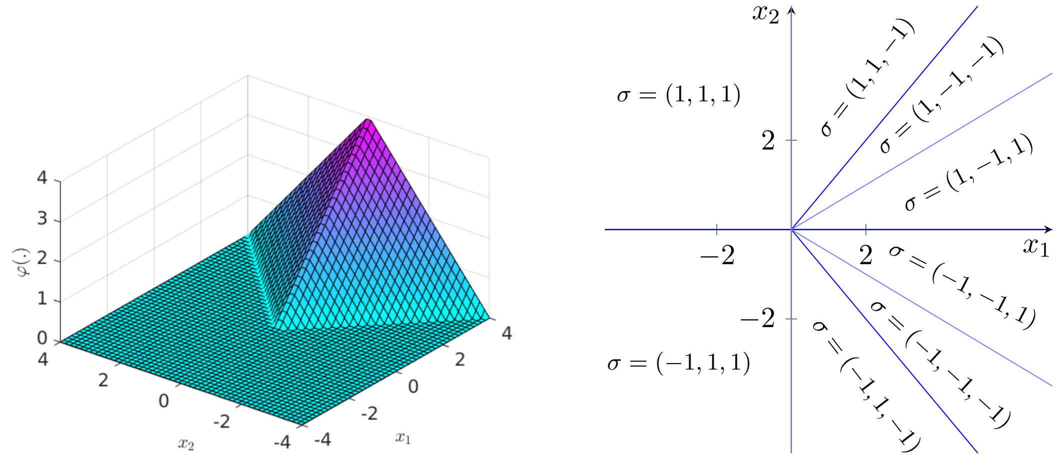
There exist optimality conditions that can be used to verify in polynomial time whether a signature optimal point x_σ is a minimizer of the full optimization problem (ALOP) or not, see, e.g., [10, 14, 16]. For this purpose, let $\tilde{\sigma} \succ \sigma$ so that $\overline{\mathcal{P}}_{\tilde{\sigma}} \supset \overline{\mathcal{P}}_\sigma$. Any such $\tilde{\sigma}$ can be decomposed into $\sigma + \gamma$, where $|\sigma|^\top |\gamma| = 0$ holds. It was shown in [10] that minimality of x_σ on $\mathcal{P}_{\tilde{\sigma}}$ then requires

$$0 \leq b^\top \gamma + \lambda^\top L|\gamma| - \lambda^\top (I_s - M)\gamma = \Big(b^\top - \lambda^\top (I_s - M)\Big)\gamma + \lambda^\top L|\gamma| \,.$$

This optimality condition is violated if and only if there is at least one index $k < s$ such that $\gamma = -\mathbf{sgn}(b_k - \lambda^\top (I_s - M)e_k)e_k$ satisfies

$$0 > \Big(b^\top - \lambda^\top (I_s - M)\Big)\gamma + \lambda^\top L|\gamma| = -\Big|b^\top - \lambda^\top (I_s - M)\Big| e_k + \lambda^\top L e_k \tag{9}$$

with $\sigma_k = 0$. Here, the fact that $|z_s(x)|$ does not contribute to the value of $f(x,z)$ and therefore does not lead to a nonsmoothness, has to be taken into account. Hence, $k = s$ must not be considered in the unconstrained case. If the optimality condition does not hold one possible strategy is to choose the index k for which the right-hand side of Eq. (9) is minimal, which represents a heuristic as known for example from active set methods [24]. By updating $\sigma^+ = \sigma + \gamma$, the resulting signature vector σ^+ has one component less that equals zero. This can be interpreted as releasing a kink in that one does not insist anymore that the corresponding absolute value is evaluated at zero.



The function $\varphi()$

Polyhedra that belong to the corresponding definite signature vectors

Figure 2: Illustration of the unconstrained case from Ex. 3.1

These considerations result in the Active Signature Method that is described in much more detail in [9].

3 Optimality conditions of the constrained optimization problem

In this section, we derive optimality conditions for the constrained optimization problem (CALOP) as defined in Sec. 1 and show that they can be verified in polynomial time at a given point. In [14, 16], for the more general class abs-smooth problems optimality conditions that can be verified in polynomial time were already derived. However, since we restrict ourselves to the special class of abs-linear functions and admit both z and $|z|$ as arguments for all functions, we prove the optimality conditions directly for our notation instead of showing the equivalence of the problem formulations in [14, 16, 15]. This allows also to make a direct connection to the termination criteria of the CASM algorithm presented and analysed in the next section.

For the constrained optimization problem (CALOP), the functions G and H may or may not depend on the value $|z_s|$ as illustrated in the next example. Therefore, we denote the total number of all switching variables z_i that occur as arguments in an evaluation of the absolute value function in the target function or in the constraints by $\tilde{s} \leq s$ and assume that they are located in the first $\tilde{s}$ switching variables. If this is not the case, the abs-linear representation of (CALOP) can be adapted correspondingly such that $\tilde{s} \in \{s-1, s\}$.

Example 3.1. *Let the function* $\varphi(x_1, x_2) = \max\{0, x_1 - |x_2|\}$ *be given. This nonsmooth nonconvex function is illustrated on the left hand side of Fig. 2. Using the reformulation* *of the* max-*function given in Eq.* (2)*, we obtain*

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

which can be converted into the following abs-linear form, see Eq. (1)*:*

$$\begin{aligned}
z = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} &= \begin{pmatrix} x_2 \\ -x_1 + |z_1| \\ -\frac{1}{2}|z_1| + \frac{1}{2}|z_2| \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \begin{bmatrix} 0 & 1 \\ -1 & 0 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \begin{pmatrix} |z_1| \\ |z_2| \\ |z_3| \end{pmatrix} \\
y &= 0 + \begin{pmatrix} \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = f(x,z)\,.
\end{aligned}$$

If we consider this function as objective of an unconstrained optimization problem as in Sec. 2, the polyhedra resulting from definite signature vectors are shown in Fig. 2, where we used the corresponding signature vectors as labels for the polyhedra. The blue lines mark the arguments that cause the nonsmoothness and the light blue lines correspond to the last component of the switching vector that does not leads to any nonsmoothness in the target function. Now, we add the constraint

$$|z_3| = \left| -\frac{1}{2}|x_2| + \frac{1}{2}\big| - x_1 + |x_2|\big| \right| \leq 2$$

that can be formulated as

$$H(x,z,|z|) = -2 + \begin{bmatrix} 0 & 0 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} + \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} |z_1| \\ |z_2| \\ |z_3| \end{pmatrix} \leq 0$$

to obtain a constrained optimization problem of the form (CALOP)*. Then,* $|z_3|$ *contributes explicitly to the evaluation of the abs-linear constraint.*

Fig. 3 shows for the constrained situation the polyhedra resulting from the definite signature vectors using the corresponding σ *as label. In comparison to the unconstrained case, further kinks are added resulting in more polyhedra. The red area represents the feasible set. All points that lie inside or on the edges of the red area are feasible.*

Next, we define polyhedra that take the additional constraints into account:

Definition 3.2 (Feasible (extended) signature domain)**.** For a fixed signature vector $\sigma \in \{-1,0,1\}^s$, we define

$$\mathcal{F}_\sigma \equiv \left\{ x \in \mathbb{R}^n \;\middle|\; \begin{array}{l} G(x,z(x),\Sigma z(x)) = 0, \\ H(x,z(x),\Sigma z(x)) \leq 0, \\ \mathbf{sgn}(z(x)) = \sigma, \end{array} \right\} \subset \overline{\mathcal{F}}_\sigma \equiv \left\{ x \in \mathbb{R}^n \;\middle|\; \begin{array}{l} G(x,z(x),|z(x)|) = 0, \\ H(x,z(x),|z(x)|) \leq 0, \\ \Sigma z(x) = |z(x)| \end{array} \right\}.$$

The set $\mathcal{F}_\sigma$ is called *feasible signature domain* and $\overline{\mathcal{F}}_\sigma$ the *feasible extended signature domain.*

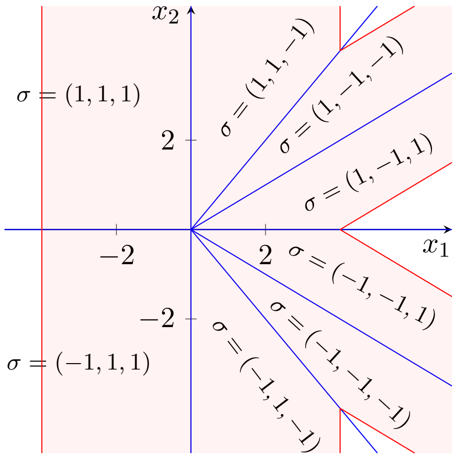


Figure 3: Polyhedra for the definite signature vectors in the constrained case of Ex. 3.1.

With these definitions, the inclusions $\mathcal{F}_\sigma \subset \mathcal{P}_\sigma$ and $\overline{\mathcal{F}}_\sigma \subset \overline{\mathcal{P}}_\sigma$ hold, where $\mathcal{F}_\sigma$ may be empty. In a similar way as we introduced the signature vector for kinks, we define a vector containing the signs of the inequality constraints.

Definition 3.3 (Signature vector and signature matrix of inequality constraints)**.** Let a point $x \in \mathbb{R}^n$ be given that fulfills the equality and inequality constraints of (CALOP). We define the *signature vector of the inequality constraints* as

$$\omega(x) \equiv \mathbf{sgn}(H(x, z, |z|)) \in \{-1, 0\}^p \ .$$

The jth inequality constraint is called *active* if $\omega_j(x) = 0$ and *inactive* otherwise. The *signature matrix of the inequality constraints* is denoted by $\Omega(x) = \mathbf{diag}(\omega(x))$. Furthermore, $\mathcal{I} \equiv \mathcal{I}(x)$ collects the indices of the active inequality constraints at x. The projection onto the active components of $H(x)$ is defined as $P_\mathcal{I} \equiv (e_i^\top)_{i \in \mathcal{I}} \in \mathbb{R}^{|\mathcal{I}| \times p}$ with e_i denoting the ith unit vector of appropriate size.

Next, we prepare the formulation of optimality conditions that can be verified in polynomial time. For this purpose, we introduce the following notations:

Definition 3.4 (Active switching variables)**.** A switching variable z_i is called *active* at x if $z_i(x) = 0$. The *active switching set* $\alpha(x)$ collects all indices of active switching variables that directly depend on x and occur as arguments of the absolute value, i.e.,

$$\alpha(x) \equiv \{i \in \{1, \dots, \tilde{s}\} \mid z_i(x) = 0\} \ .$$

The projection onto the active components of $z(x)$ is defined as $P_\alpha \equiv (e_i^\top)_{i \in \alpha} \in \mathbb{R}^{|\alpha| \times s}$ with e_i denoting the ith unit vector of appropriate size.

For each fixed signature vector $\sigma \in \{-1, 0, 1\}^s$, we obtain from (CALOP) similar to Eq. (7) the smooth optimization problem

$$\min_{x \in \mathbb{R}^n, z \in \mathbb{R}^s} \quad a^\top x + b^\top z \tag{10a}$$

$$\text{s.t.} \quad 0 = g + Ax + Bz + C\Sigma z \ , \tag{10b}$$

$$0 \geq h + Dx + Ez + F\Sigma z \ , \tag{10c}$$

$$z = c + Zx + Mz + L\Sigma z \ , \tag{10d}$$

$$0 = (I_s - |\Sigma|)z \ , \tag{10e}$$

$$0 \leq \Sigma z \ . \tag{10f}$$

Now, we can extend the concept of signature optimality to the situation considered in this section:

Definition 3.5 (Feasible signature optimal point)**.** Let an optimization problem of the form (CALOP) be given. Consider a fixed signature vector $\sigma \in \{-1, 0, 1\}^s$. A minimizer $x_\sigma \in \mathcal{F}_\sigma$ of the optimization problem (10) is called *feasible signature optimal point* of the original, constrained optimization problem (CALOP).

To reformulate the optimization problem (10), we define

$$\tilde{Z} = (I_s - M - L\Sigma)^{-1} Z \qquad \text{and} \qquad \tilde{c} = (I_s - M - L\Sigma)^{-1} c \ . \tag{11}$$

Then one can combine Eqs. (10d) and (10e) to one equality constraint and obtains the following optimization problem that is equivalent to the one stated in Eq. (10)

$$\min_{x \in \mathbb{R}^n, z \in \mathbb{R}^s} \quad a^\top x + b^\top |\Sigma| z \tag{12a}$$

$$\text{s.t.} \quad 0 = g + Ax + B|\Sigma|z + C\Sigma z \ , \tag{12b}$$

$$0 \geq h + Dx + E|\Sigma|z + F\Sigma z \ , \tag{12c}$$

$$0 = |\Sigma|z - \tilde{c} - \tilde{Z}x \ , \tag{12d}$$

$$0 \leq \Sigma z \ . \tag{12e}$$

Since we consider only linear constraints, one has for the optimization problem (12) that the set of feasible directions at x coincides with the tangent cone at x, see [24, Lem. 12.7]. In this case, no further constraint qualification is needed to ensure the existence of Lagrange multipliers but then their uniqueness is not guaranteed. Our goal is to derive optimality conditions that can be verified in polynomial time. Hence, any dependence on the signature vectors that would lead to a combinatorial complexity in 2^s in the worst case must be avoided. Therefore, we have to ensure that the Lagrange multipliers are unique, see also [11]. For this reason, we adapt the kink qualification LIKQ that was introduced in [7] for the unconstrained case appropriately. In [16], LIKQ has already been extended for constrained nonsmooth nonlinear optimization problems. However, since we focus in this paper on the piecewise linear case, LIKQ can be specified in its matrix representation.

In the unconstrained case, LIKQ requires the full rank of the matrix $P_\alpha \tilde{Z}$, i.e., the active Jacobian of the reformulated switching system. To derive a similar result for the constrained case, we analyze the optimization problem (12) for a feasible point x_σ in more detail. Due to the continuity of all involved functions and the relation $\Sigma z = |z|$, the components z_i, $i \notin \alpha$, of the vector z determined by Eq. (12d) will not drop to zero in an open neighborhood $U(x_\sigma)$ of x_σ. In combination with the identity $\Sigma z = \Sigma|\Sigma|z$, in this neighborhood $U(x_\sigma)$ the optimization problem (12) is then equivalent to

$$\min_{x \in U(x_\sigma)} \quad a^\top x + b^\top |\Sigma|(\tilde{c} + \tilde{Z}x) \tag{13a}$$

$$\text{s.t.} \quad 0 = g + Ax + B|\Sigma|(\tilde{c} + \tilde{Z}x) + C\Sigma(\tilde{c} + \tilde{Z}x) \ , \tag{13b}$$

$$0 \geq h + Dx + E|\Sigma|(\tilde{c} + \tilde{Z}x) + F\Sigma(\tilde{c} + \tilde{Z}x) \ , \tag{13c}$$

$$0 = P_\alpha(\tilde{c} + \tilde{Z}x) \ . \tag{13d}$$

Definition 3.6 (Active Jacobian)**.** Consider for the constrained optimization problem (CALOP) and a given signature vector $\sigma \in \{-1, 0, 1\}^s$ a point x_σ that is feasible for the problem given by Eq. (13). The *active Jacobian* at x_σ is given by

$$\mathcal{J}_\sigma \equiv \begin{bmatrix} A + B|\Sigma|\tilde{Z} + C\Sigma\tilde{Z} \\ P_\mathcal{I}(D + E|\Sigma|\tilde{Z} + F\Sigma\tilde{Z}) \\ P_\alpha \tilde{Z} \end{bmatrix} \in \mathbb{R}^{(m+|\mathcal{I}|+|\alpha|)\times n} \ .$$

Now, the required kink qualification can be stated for the setting considered in this paper:

Definition 3.7 (LIKQ (constrained case))**.** Let a constrained optimization problem of the form (CALOP) and a signature vector $\sigma \in \{-1, 0, 1\}^s$ be given. We say that the *Linear Independence Kink Qualification (LIKQ)* holds at a feasible point x_σ if the active Jacobian $\mathcal{J}_\sigma$ at x_σ has full row rank $m + |\mathcal{I}| + |\alpha|$.

After these preparations, we are able to show that the optimality of a feasible signature optimal point can be verified in polynomial time extending the results given in [10] to the constrained case.

Theorem 3.8 (Necessary and sufficient optimality conditions)**.** *Let a constrained optimization problem of the form* (CALOP) *and a signature vector* $\sigma \in \{-1, 0, 1\}^s$ *be given. Assume that* x_σ *is feasible signature optimal for* (CALOP) *and that LIKQ holds at* x_σ*. Then* x_σ *is a local minimizer of* (CALOP) *if and only if there exist unique Lagrange multipliers* $\delta \in \mathbb{R}^m$*,* $0 \leq \nu \in \mathbb{R}^p$ *and* $\lambda \in \mathbb{R}^s$*, such that*

$$0 = a^\top + b^\top |\Sigma|\tilde{Z} + \delta^\top(A + B|\Sigma|\tilde{Z} + C\Sigma\tilde{Z}) + \nu^\top(D + E|\Sigma|\tilde{Z} + F\Sigma\tilde{Z}) - \lambda^\top P_\alpha^\top P_\alpha \tilde{Z} \ , \tag{14}$$

$$0 = b^\top |\Sigma| + \delta^\top \left(B|\Sigma| + C\Sigma\right) + \nu^\top \left(E|\Sigma| + F\Sigma\right) + \lambda^\top |\Sigma| \tag{15}$$

and

$$|P_\alpha(b + B^\top \delta + E^\top \nu + \lambda)| \leq P_\alpha(C^\top \delta + F^\top \nu - \tilde{L}^\top \lambda) \tag{16}$$

with $\tilde{L}$ given by

$$\tilde{L} = (I_s - M - L\Sigma)^{-1}L \; .$$

Proof. First, let x_σ be a local minimizer of (CALOP). Since x_σ is feasible signature optimal for the given signature vector σ, x_σ is also a minimizer of the optimization problem (13). Then, we obtain from standard KKT theory that there exist unique Lagrange multipliers $\delta \in \mathbb{R}^m$, $0 \le \nu \in \mathbb{R}^p$ and $\check{\lambda} \in \mathbb{R}^{|\alpha|}$ associated with the equality constraint (13b), the inequality constraint (13c) and the reformulated switching system (13d) such that

$$0 = a^\top + b^\top|\Sigma|\tilde{Z} + \delta^\top(A + B|\Sigma|\tilde{Z} + C\Sigma\tilde{Z}) + \nu^\top(D + E|\Sigma|\tilde{Z} + F\Sigma\tilde{Z}) + \check{\lambda}^\top P_\alpha \tilde{Z} \; .$$

Hence, together with $\delta \in \mathbb{R}^m$ and $0 \le \nu \in \mathbb{R}^p$, each vector $\lambda \in \mathbb{R}^s$ such that $\check{\lambda} = -P_\alpha\lambda$ fulfills Eq. (14).

As introduced before, $\omega = \omega(x_\sigma)$ denotes the signature vector of the inequality constraints. Then, it is necessary and sufficient for local minimality that $(x_\sigma, z(x_\sigma))$ is a minimizer of $f(.,.)$ as defined in Eq. (3) on all feasible extended signature domains $\overline{\mathcal{F}}_{\tilde{\sigma}}$ with definite $\tilde{\sigma} \succ \sigma$. Any such $\tilde{\sigma} \succ \sigma$ can be written as $\tilde{\sigma} = \sigma + \gamma$ with $\gamma \in \{-1, 0, 1\}^s$ structurally orthogonal to σ such that for $\Gamma \equiv \mathbf{diag}(\gamma)$ we have the matrix equations

$$\tilde{\Sigma} = \Sigma + \Gamma \quad \text{and} \quad \Sigma\Gamma = 0 = |\Sigma|\Gamma \; . \tag{17}$$

Then we can express $z(x) = z_{\tilde{\sigma}}(x)$ for $x \in \mathcal{P}_{\tilde{\sigma}}$ as

$$z_{\tilde{\sigma}}(x) = z_{\sigma+\gamma}(x) = (I_s - M - L\Sigma - L\Gamma)^{-1}(c + Zx) = (I_s - \tilde{L}\Gamma)^{-1}(\tilde{c} + \tilde{Z}x) \; . \tag{18}$$

Since x_σ must be a minimizer of the objective function also on $\overline{\mathcal{F}}_{\tilde{\sigma}}$, it solves the smooth optimization problem

$$\begin{aligned}
&\min_{x \in \mathbb{R}^n, z \in \mathbb{R}^s} \quad a^\top x + b^\top (|\Sigma| + |\Gamma|)\, z && \text{(19a)}\\
&\text{s.t.} \quad 0 = g + Ax + B\,(|\Sigma| + |\Gamma|)\, z + C(\Sigma + \Gamma)z \; , && \text{(19b)}\\
&\qquad\;\; 0 \ge h + Dx + E\,(|\Sigma| + |\Gamma|)\, z + F(\Sigma + \Gamma)z \; , && \text{(19c)}\\
&\qquad\;\; 0 = (I_s - \tilde{L}\Gamma)z - \tilde{c} - \tilde{Z}x \; , && \text{(19d)}\\
&\qquad\;\; 0 \le P_\alpha \Gamma z \; . && \text{(19e)}
\end{aligned}$$

Once more, we obtain from KKT theory that there exist Lagrange multipliers $\delta \in \mathbb{R}^m$, $0 \le \nu \in \mathbb{R}^p$, $\lambda \in \mathbb{R}^s$ and $0 \le \mu \in \mathbb{R}^{|\alpha|}$ associated with the equality constraint, the inequality constraint, the reformulated switching system and the sign conditions such that

$$0 = a^\top + \delta^\top A + \nu^\top D - \lambda^\top \tilde{Z} \quad \text{and} \tag{20}$$

$$\begin{aligned}
0 = {}& b^\top (|\Sigma| + |\Gamma|) + \delta^\top \left(B\,(|\Sigma| + |\Gamma|) + C(\Sigma + \Gamma)\right)\\
&+ \nu^\top \left(E\,(|\Sigma| + |\Gamma|) + F(\Sigma + \Gamma)\right) + \lambda^\top \left(I_s - \tilde{L}\Gamma\right) - \mu^\top P_\alpha \Gamma \; .
\end{aligned} \tag{21}$$

Since the optimization problem (19) is linear, these, together with the feasibility of the variables and the complementarity condition, are necessary and sufficient for x_σ to be a minimizer. Multiplying the last equation from the right by $|\Sigma|\tilde{Z}$, we obtain with the identity $\Sigma = \Sigma|\Sigma|$ and Eq. (17)

$$0 = b^\top|\Sigma|\tilde{Z} + \delta^\top\left(B|\Sigma| + C\Sigma\right)\tilde{Z} + \nu^\top\left(E|\Sigma| + F\Sigma\right)\tilde{Z} + \lambda^\top|\Sigma|\tilde{Z}\ . \tag{22}$$

Adding this equation to (20) and exploiting

$$I_s = |\Sigma| + P_\alpha^\top P_\alpha \tag{23}$$

yields

$$0 = a^\top + b^\top|\Sigma|\tilde{Z} + \delta^\top(A + B|\Sigma|\tilde{Z} + C\Sigma\tilde{Z}) + \nu^\top(D + E|\Sigma|\tilde{Z} + F\Sigma\tilde{Z}) - \lambda^\top P_\alpha^\top P_\alpha\tilde{Z}\ . \tag{24}$$

Hence, it follows that the Lagrange multipliers $\delta \in \mathbb{R}^m$, $\nu \in \mathbb{R}^p$, $\lambda \in \mathbb{R}^s$ fulfill Eq. (14) with $\check{\lambda} = -P_\alpha\lambda$. Due to the kink qualification LIKQ, one also has that the vectors $\delta \in \mathbb{R}^m$ and $\nu \in \mathbb{R}^p$ as well as the components $P_\alpha\lambda \in \mathbb{R}^{|\alpha|}$ are determined uniquely. The remaining components of $\lambda \in \mathbb{R}^s$ can be obtained by multiplying Eq. (21) this time only with $|\Sigma|$ from the right yielding

$$0 = b^\top|\Sigma| + \delta^\top\left(B|\Sigma| + C\Sigma\right) + \nu^\top\left(E|\Sigma| + F\Sigma\right) + \lambda^\top|\Sigma|$$

and thus Eq. (15). To derive the third condition (16), we multiply Eq. (21) from the right by $\Gamma P_\alpha^\top$. Using

$$P_\alpha^\top P_\alpha = \Gamma\Gamma = |\Gamma| \text{ and } P_\alpha P_\alpha^\top = I_{|\alpha|} \tag{25}$$

and $\mu \geq 0$, it follows that

$$\begin{aligned} -(b^\top + \delta^\top B + \nu^\top E + \lambda^\top)\Gamma P_\alpha^\top &= (\delta^\top C + \nu^\top F - \lambda^\top\tilde{L})\Gamma\Gamma P_\alpha^\top - \mu^\top \\ &\leq (\delta^\top C + \nu^\top F - \lambda^\top\tilde{L})P_\alpha^\top\ . \end{aligned}$$

Now the key observation is that this condition is linear in Γ and is strongest for the choice $\gamma_i = -\mathbf{sgn}(\lambda^\top + b^\top + \delta^\top B + \nu^\top E)_i$ for $i \in \alpha$ yielding the inequalities

$$|(b + B^\top\delta + E^\top\nu + \lambda)_i| \leq e_i(C^\top\delta + F^\top\nu - \tilde{L}^\top\lambda) \quad \text{for} \quad i \in \alpha$$

showing Eq. (16) and therefore the necessary optimality conditions.

Second, we show that these conditions are also sufficient. For this purpose, we consider again all adjacent extended signature domains $\overline{\mathcal{F}}_{\tilde{\sigma}}$. Therefore, we multiply Eq. (21) again from the right by $\Gamma P_\alpha^\top$ and use Eqs. (25) and (16) to obtain

$$\mu^\top = \left(b^\top + \delta^\top B + \nu^\top E + \lambda^\top\right)\Gamma P_\alpha^\top + \left(\delta^\top C + \nu^\top F - \lambda^\top\tilde{L}\right)P_\alpha^\top \geq 0$$

and thus the feasibility. Exploiting Eq. (23), Eq. (15) multiplied from the right by $\tilde{Z}$ and Eq. (14) yields

$$\begin{aligned}
\lambda^\top \tilde{Z} &= \lambda^\top \left(|\Sigma| + P_\alpha^\top P_\alpha\right) \tilde{Z} = \lambda^\top |\Sigma| \tilde{Z} + \lambda^\top P_\alpha^\top P_\alpha \tilde{Z} \\
&= -b^\top |\Sigma| \tilde{Z} - \delta^\top \left(B|\Sigma| + C\Sigma\right) \tilde{Z} - \nu^\top \left(E|\Sigma| + F\Sigma\right) \tilde{Z} \\
&\quad + a^\top + b^\top |\Sigma| \tilde{Z} + \delta^\top \left(A + B|\Sigma|\tilde{Z} + C\Sigma\tilde{Z}\right) + \nu^\top \left(D + E|\Sigma|\tilde{Z} + F\Sigma\tilde{Z}\right) \\
&= a^\top + \delta^\top A + \nu^\top D
\end{aligned}$$

and hence Eq. (20). Using Eq. (23), Eq. (21) holds if and only if

$$\begin{aligned}
0 &= b^\top \left(|\Sigma| + |\Gamma|\right) + \delta^\top \left(B\left(|\Sigma| + |\Gamma|\right) + C(\Sigma + \Gamma)\right) \\
&\quad + \nu^\top \left(E\left(|\Sigma| + |\Gamma|\right) + F(\Sigma + \Gamma)\right) + \lambda^\top \left(|\Sigma| + P_\alpha^\top P_\alpha - \tilde{L}\Gamma\right) - \mu^\top P_\alpha \Gamma
\end{aligned}$$

holds true. Using Eq. (15) the last equation is equivalent to

$$0 = b^\top |\Gamma| + \delta^\top \left(B|\Gamma| + C\Gamma\right) + \nu^\top \left(E|\Gamma| + F\Gamma\right) + \lambda^\top \left(P_\alpha^\top P_\alpha - \tilde{L}\Gamma\right) - \mu^\top P_\alpha \Gamma$$

Multiplying the last equation from the right by $\Gamma P_\alpha^\top$ and exploiting Eq. (25), we obtain

$$\mu^\top = -\lambda^\top \tilde{L} P_\alpha^\top + \left(b^\top + \delta^\top B + \nu^\top E + \lambda^\top\right) \Gamma P_\alpha^\top + \left(\delta^\top C + \nu^\top F\right) P_\alpha^\top \,.$$

Thus, defining the Lagrange multiplier μ as given above, it satisfies Eq. (21) by using Eq. (15) and (23)

$$\begin{aligned}
& b^\top \left(|\Sigma| + |\Gamma|\right) + \delta^\top \left(B\left(|\Sigma| + |\Gamma|\right) + C\left(\Sigma + \Gamma\right)\right) \\
& + \nu^\top \left(E\left(|\Sigma| + |\Gamma|\right) + F\left(\Sigma + \Gamma\right)\right) + \lambda^\top \left(I_s - \tilde{L}\Gamma\right) \\
& - \left(-\lambda^\top \tilde{L} P_\alpha^\top + \left(b^\top + \delta^\top B + \nu^\top E + \lambda^\top\right) \Gamma P_\alpha^\top + \left(\delta^\top C + \nu^\top F\right) P_\alpha^\top\right) P_\alpha \Gamma \\
= {} & b^\top \left(|\Sigma| + |\Gamma|\right) + \delta^\top \left(B\left(|\Sigma| + |\Gamma|\right) + C\left(\Sigma + \Gamma\right)\right) \\
& + \nu^\top \left(E\left(|\Sigma| + |\Gamma|\right) + F\left(\Sigma + \Gamma\right)\right) + \lambda^\top \left(I_s - \tilde{L}\Gamma\right) \\
& + \lambda^\top \tilde{L}\Gamma - \left(b^\top |\Gamma| + \delta^\top B|\Gamma| + \nu^\top E|\Gamma| + \lambda^\top |\Gamma| + \delta^\top C\Gamma + \nu^\top F\Gamma\right) \\
= {} & - \lambda^\top |\Sigma| + \lambda^\top \tilde{L}\Gamma - \lambda^\top |\Gamma| + \lambda^\top I_s - \lambda^\top \tilde{L}\Gamma \\
= {} & - \lambda^\top |\Sigma| - \lambda^\top |\Gamma| + \lambda^\top |\Sigma| + \lambda^\top P_\alpha^\top P_\alpha = -\lambda^\top |\Gamma| + \lambda^\top \Gamma\Gamma = -\lambda^\top |\Gamma| + \lambda^\top |\Gamma| = 0 \,.
\end{aligned}$$

Since the optimization problem (19) is a linear optimization problem, the KKT conditions are necessary and sufficient for the minimality of x_σ. Thus, we have shown that x_σ satisfies the KKT conditions for all adjacent extended signature domains. Therefore, x_σ is also a minimizer of (19) and hence of (CALOP).

For the uniqueness of the Lagrange multipliers, see again the first paragraph of this proof. There it was stated that the Lagrange multipliers δ and ν as well as the components λ_i belonging to the index set $\alpha(x_\sigma)$ are unique. Finally, for the remaining $i \in \alpha^C$, the complement of α, the components λ_i can be uniquely determined by Eq. (15). □

It is important to note that for given Lagrange multipliers δ, ν, and λ, it can be verified in polynomial time whether the conditions (14)–(16) hold. Hence, this optimality test at a feasible signature optimal point is independent from the combinatorial complexity caused by all the possible values of Γ.

Furthermore, for the unconstrained case, i.e., $A = 0$, $B = 0$, $C = 0$, $D = 0$, $E = 0$, $F = 0$ in the appropriate dimensions, one rediscovers the conditions

$$0 = a^\top + b^\top |\Sigma| \tilde{Z} + \lambda^\top P_\alpha \tilde{Z} \qquad \text{and} \qquad |P_\alpha(b + \lambda)| \leq P_\alpha(-\tilde{L}^\top \lambda) \ ,$$

i.e., tangential stationarity and normal growth as introduced in [7].

4 The active signature method for constrained problems

In this section, we extend the ASM described in Sec. 2 to problems with abs-linear constraints of the form (CALOP). It should be mentioned that due to the additional equality and inequality constraints, the set of feasible points could be empty. Throughout, we assume that this is not the case such that the iteration can start with a feasible point. Subsequently, feasibility is maintained, i.e., the derived algorithm is a feasible point method. If no feasible starting point is given from the application context, one can calculate such a starting point with a Phase-I-like method known from linear optimization (cf. [24, Chapter 16]). If a feasible starting point exists, one can show that the resulting algorithm terminates within a finite number of iterations.

In the following we divide the section into two subsections. In the first one, the algorithm itself is described and in the second one the convergence is analyzed.

4.1 The Algorithm

To explain the algorithm and its individual parts, we divide this subsection into different paragraphs. For the given iterate, first, a search direction is calculated as described in the following first paragraph. Then, for the search direction, a step size is calculated as explained in the second paragraph. The third paragraph deals with the optimality condition and the control of σ and ω in case of nonoptimality. Based on these three main components the whole algorithm is described in the final paragraph of this subsection.

Computing a direction for given σ and ω Similar to the unconstrained case, we add a quadratic penalty term with a positive definite matrix $Q = Q^\top \in \mathbb{R}^{n\times n}$ to the target function ensuring that the problem is bounded from below. Hence, we want to solve

$$\min_{x\in\mathbb{R}^n, z\in\mathbb{R}^s} \quad a^\top x + b^\top |\Sigma| z + \frac{1}{2} x^\top Q x \tag{26a}$$

$$\text{s.t.} \quad 0 = g + Ax + B|\Sigma|z + C\Sigma z \ , \tag{26b}$$

$$0 \geq h + Dx + E|\Sigma|z + F\Sigma z \ , \tag{26c}$$

$$0 = |\Sigma|z - \tilde{c} - \tilde{Z}x \ , \tag{26d}$$

$$0 \leq \Sigma z \ , \tag{26e}$$

with $\tilde{Z}$ and $\tilde{c}$ as defined in Eq. (11). Due to the fixed signature vector, this optimization problem is smooth with a quadratic target function and linear constraints. Hence, it could be solved with a standard QP method. However, we want to exploit the structure provided by the signature vector as additional feature.

Once more, standard KKT theory can be applied. With Lagrange multipliers $\delta \in \mathbb{R}^m, \nu \in \mathbb{R}^p, \lambda \in \mathbb{R}^s$ and $\mu \in \mathbb{R}^s$, we obtain the following necessary optimality conditions

$$
\begin{aligned}
&0 = a^\top + x^\top Q + \delta^\top A + \nu^\top D - \lambda^\top \tilde{Z} \;, && \text{(27a)}\\
&0 = b^\top |\Sigma| + \delta^\top (B|\Sigma| + C\Sigma) + \nu^\top (E|\Sigma| + F\Sigma) + \lambda^\top |\Sigma| - \mu^\top \Sigma \;, && \text{(27b)}\\
&0 = g + Ax + B|\Sigma|z + C\Sigma z \;, && \text{(27c)}\\
&0 \geq h + Dx + E|\Sigma|z + F\Sigma z \;, && \text{(27d)}\\
&0 = |\Sigma|z - \tilde{c} - \tilde{Z}x \;, && \text{(27e)}\\
&0 \leq \Sigma z \;, \quad 0 \leq \mu \;, \quad 0 = \mu^\top \Sigma z \;, && \text{(27f)}\\
&0 \leq \nu \;, \qquad 0 = \nu^\top (h + Dx + E|\Sigma|z + F\Sigma z) \;. && \text{(27g)}
\end{aligned}
$$

Multiplying Eq. (27b) by Σ from the right and using Eq. (27f) yields

$$
0 \leq \mu^\top |\Sigma| = b^\top \Sigma + \delta^\top (B\Sigma + C|\Sigma|) + \nu^\top (E\Sigma + F|\Sigma|) + \lambda^\top \Sigma \;. \tag{28}
$$

Due to the complementarity condition $\mu^\top \Sigma z = 0$, this inequality must hold as an equality. Hence, it follows that

$$
-b^\top \Sigma = \delta^\top (B\Sigma + C|\Sigma|) + \nu^\top (E\Sigma + F|\Sigma|) + \lambda^\top \Sigma \;.
$$

Thus with $\omega = \mathbf{sgn}(H(x, |z|))$ and $\Omega = \mathbf{diag}(\omega)$ denoting as before the projection onto the active inequality constraints, we get the linear system

$$
\begin{bmatrix}
Q & 0 & -\tilde{Z}^\top & A^\top & D^\top \\
0 & 0 & \Sigma & \Sigma B^\top + |\Sigma| C^\top & \Sigma E^\top + |\Sigma| F^\top \\
\tilde{Z} & -|\Sigma| & 0 & 0 & 0 \\
A & B|\Sigma| + C\Sigma & 0 & 0 & 0 \\
\bar{\Omega} D & \bar{\Omega}(E|\Sigma| + F\Sigma) & 0 & 0 & \Omega
\end{bmatrix}
\begin{bmatrix} \hat{x} \\ \hat{z} \\ \lambda \\ \delta \\ \nu \end{bmatrix}
= -
\begin{bmatrix} a \\ \Sigma b \\ \tilde{c} \\ g \\ \bar{\Omega} h \end{bmatrix} , \tag{29}
$$

where $\bar{\Omega} = I_p - |\Omega|$ forces the inactive inequalities to vanish. The matrix Ω in the right lower corner ensures that ν is zero for the inactive inequality constraints. As in the unconstrained case, see Eq. (5), due to its consistency and the assumption that a feasible starting point exists, the system of equations (29) always has a solution, which in general does not necessarily have to be unique. We denote a solution by $(\hat{x}, \hat{z}, \lambda, \delta, \nu)$ and define for the current iterate x and z

$$
\Delta x := \hat{x} - x \quad \text{and} \quad \Delta z := \hat{z} - z \tag{30}
$$

as directions towards the next iterate.

If $\Delta x = 0$ one also has $\Delta z = 0$ due to the assumed structure when computing z. Furthermore, for $\Delta x = 0$ and $\Delta z = 0$ it follows from Eqs. (29) and (26e) that already x and z satisfy the KKT conditions for the optimization problem (26), i.e., x is feasible signature optimal. Hence, in this case one can directly check the optimality conditions.

Computing a step size β Given $(\hat{x}, \hat{z})$ as solution of (29), we must now check whether $\sigma(\hat{x}) = \sigma$ is still valid and that the inequality constraints of Eq. (27) still hold to ensure feasibility. For this purpose, we calculate two step sizes. As in the unconstrained case, the first step size is the step length from the current iterate x in the direction Δx to a possible kink, i.e., a sign change in one component of z. Therefore, this step size is also denoted by β^z and defined as

$$\beta^z = \inf_{1 \le l \le s} \left\{ \beta_l^z \equiv \frac{-z_l}{\hat{z}_l - z_l} \middle| (\hat{z}_l - z_l)\sigma_l < 0 \right\} \in (0, \infty] \ . \tag{31}$$

Once more, if $\beta^z < \infty$ the first index for which the minimum is attained is denoted by j^z. For $\beta^z \le 1$, there exists a blocking kink with the same consequences as in the unconstrained case. Note that due to the fact that for a given signature vector σ and $x \in \mathcal{P}_\sigma$ one has that for $\sigma_i \ne 0$ also $z_i(x) \ne 0$ holds. Thus, for σ with $\sigma_i \ne 0$ one has $z_i(x) \ne 0$ and $\beta^z > 0$ must hold.

The second step size is the step length from the current iterate x in the direction Δx to a possible inequality constraint $H_l(x, z, \Sigma z), 1 \le l \le p$, that becomes active. In a similar way to the computation of β^z this step size β^H is given by

$$\beta^H = \inf_{1 \le l \le p} \left\{ \beta_l^H \equiv \frac{H_l}{H_l - \hat{H}_l} \middle| (\hat{H}_l - H_l)\omega_l < 0 \right\} \in (0, \infty] \ , \tag{32}$$

where $H \equiv H(x, z, \Sigma z)$, $\hat{H} \equiv H(\hat{x}, \hat{z}, \Sigma \hat{z})$ and l denotes the lth component of H and $\hat{H}$, respectively. Similar to the first step size, we denote by j^H the smallest index for which the minimum is attained. For $\beta^H < 1$, there exists a blocking inequality constraint, i.e., the solution $\hat{x}$ is not feasible. Therefore, the new iterate x^+ should be chosen such that the j^Hth components of $H(x^+, z^+, \Sigma z^+)$ and $\omega(x^+)$ drop to zero in comparison to $H(x, z, \Sigma z)$ and ω, respectively. Setting $\omega^+_{j^H} = 0$ changes the optimality system (27) and a new solution of system (29) has to be computed. If $\beta^H \le \beta^z$ then we have $z_{j^H} \hat{z}_{j^H} \ge 0$ such that the iterate $\hat{x}$ is still contained in $\mathcal{P}_\sigma$, i.e., $\sigma(\hat{x}) = \sigma$ is still valid. Since all active constraints are encoded in ω, one must have $\beta^H > 0$.

The step sizes β^z and β^H are illustrated in Fig. 4, where the blue line represents a blocking kink and the red one a blocking inequality constraint. The yellow arrows indicate the corresponding step sizes, i.e., on the left hand side β^z and on the right hand side β^H.

Finally, we determine the actual step size

$$\beta = \min\{\beta^z, \beta^H, 1\} \in (0, 1] \ , \tag{33}$$

where the upper bound 1 on β ensures with the update

$$x^+ = (1 - \beta)x + \beta\hat{x} = x + \beta\Delta x$$

that the next iterate is still contained in $\overline{\mathcal{F}}_\sigma$. As can be seen, the case $\beta < 1$ corresponds to the activation of a kink or inequality constraint, respectively. Therefore, we will refer to this situation as a restriction of σ of ω, respectively. If $\beta = 1$, one has for the new iterate $x^+ = \hat{x}$ that $\sigma(x^+) = \sigma$ and $\omega(x^+) = \omega$. In this case, x^+ is called *signature stationary* since the two signature vectors are kept.

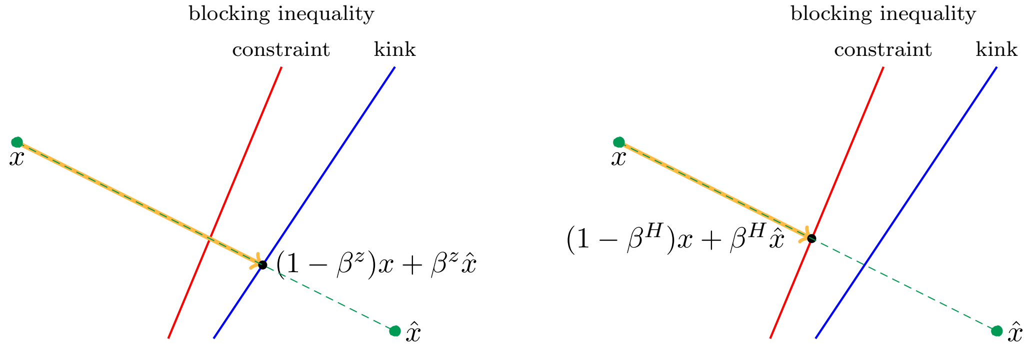


Figure 4: The two different step sizes β^z and β^H

Checking the optimality If x^+ is signature stationary on the current polyhedron $\mathcal{P}_\sigma$, one has to check whether x^+ is a minimizer of (CALOP). If this is the case the iteration stops. Otherwise, the optimization continues in one of the neighboring polyhedra $P_{\tilde{\sigma}}$ with $\tilde{\sigma} \succ \sigma$. Such a $\tilde{\sigma}$ can be again decomposed into $\sigma+\gamma$ where $|\sigma|^\top|\gamma| = 0$. Replacing Σ in the optimality conditions (27) by the corresponding $\Sigma+\Gamma$ and using Eq. (18) we see that most of the relations are still fulfilled by the current values $\hat{x}, \hat{z}$ and λ. The only thing that changes is that Eq. (28) has as many new nontrivial component as γ which can be written as

$$\begin{aligned} 0 \le \mu^\top|\Gamma| &= b^\top\Gamma + \delta^\top(B\Gamma + C|\Gamma|) + \nu^\top(E\Gamma + F|\Gamma|) + \lambda^\top(I_s - \tilde{L}\Gamma)\Gamma \\ &= (b^\top + \delta^\top B + \nu^\top E + \lambda^\top)\Gamma + (\delta^\top C + \nu^\top F - \lambda^\top\tilde{L})|\Gamma| \, . \end{aligned}$$

This condition is violated if and only if there exists at least one index k such that $\gamma \equiv -\mathbf{sgn}(b_k + \delta^\top Be_k + \nu^\top Ee_k + \lambda_k)e_k$ satisfies using $\Gamma = \mathbf{diag}(\gamma)$

$$\begin{aligned} 0 &> (\delta^\top C + \nu^\top F - \lambda^\top\tilde{L})e_k - \left|b^\top + \delta^\top B + \nu^\top E + \lambda^\top\right| e_k \quad \text{and} \\ \sigma_k &= 0 \, . \end{aligned} \tag{34}$$

As mentioned in the proof of Theo. 3.8 this is the strongest condition. This is due to the fact that this condition relates directly to the optimality condition (16) derived in Theo. 3.8. In addition we must check whether any one of the components ν_l for $1 \le l \le p$ of the Lagrange multiplier ν associated with the inequality constraints is negative. If such a violation occurs, one possible strategy is to take the most negative component of ν and Eq. (34) as discussed in the next paragraph in more detail.

If $\nu \ge 0$ does not holds, we choose the component for which $\nu \ge 0$ is most violated and drop the corresponding constraint. Hence, the associated entry of ω is set to -1 relaxing ω. If $\nu \ge 0$ holds, the current iterate is feasible signature optimal. Therefore, we check the optimality condition given by Eq. (34). If Eq. (34) is fulfilled for at least one index k the current point is not a minimizer of (CALOP) and we leave a kink by relaxing σ, i.e., setting the corresponding entry σ_k to a nonzero value via $\sigma^+ = \sigma + \gamma$.

Algorithm 1 Constrained active signature method (CASM)

Require: Feasible start point $x \in \mathbb{R}^n, n \in \mathbb{N}$, $s, m, p \in \mathbb{N} \cup \{0\}$, $a \in \mathbb{R}^n$, $b, c \in \mathbb{R}^s$, $Z \in \mathbb{R}^{s \times n}$, $L, M \in \mathbb{R}^{s \times s}$ strictly lower triangular, $Q = Q^\top \in \mathbb{R}^{n \times n}$ positive definite, $g \in \mathbb{R}^m$, $h \in \mathbb{R}^p$, $A \in \mathbb{R}^{m \times n}$, $B, C \in \mathbb{R}^{m \times s}$, $D \in \mathbb{R}^{p \times n}$, $E, F \in \mathbb{R}^{p \times s}$, $\beta = 0$

Set: $z := z(x)$ via Eq. (1a), $\sigma := \sigma(x)$ and $\omega := \omega(x)$

1: **loop**
2:     Compute $(\hat{x}, \hat{z}, \lambda, \delta, \nu)$ by solving Eq. (29)
3:     Compute β^z via Eq. (31), β^H via Eq. (32) and β via Eq. (33)
4:     Set $(x^+, z^+) = (1 - \beta)(x, z) + \beta(\hat{x}, \hat{z})$
5:     **if** $\beta^H = \beta$ **then** Restrict ω ▷ Add constraint
6:     **if** $\beta^z = \beta$ **then** Restrict σ ▷ Add kink
7:     **if** $\beta = 1$ **then**   ▷ x^+ is feasible signature stationary
8:         **if** $\nu \not\geq 0$ **then**
9:             Relax ω, set $\beta = 0$ ▷ Drop constraint
10:         **else**   ▷ x^+ is feasible signature optimal
11:             **if** Eq. (34) holds true **then**
12:                 Relax σ, set $\beta = 0$ ▷ Drop kink
13:             **else**   ▷ x^+ is local optimal
14:                 **return** (x^+, z^+) ▷ Problem solved
15:     Set $(x, z) = (x^+, z^+)$

The overall algorithm Combining all the considerations described above, one obtains Algo. 1 that consists of three main parts: First, the computation of the search direction (cf. line 2 of Algo. 1). Second, computing the step size and in case of blocking kinks and/or inequality constraints restrict σ and/or ω, respectively (cf. line 3-6). Third, checking the optimality and relaxing kinks or constraints in case of nonoptimality (cf. line 7-14).

4.2 Convergence analysis of CASM

Next, we analyze the convergence behavior of Algo. 1. For this purpose, we first examine the question, whether CASM yields a monotone decreasing sequence of function values. In each iteration, the optimality system (29) is solved which corresponds to the computation of a Newton step for the smooth optimization problem (26) when one ignores the inequality constraints.

First, we will show that under mild assumptions solving the saddle point system (29) yields a descent direction. The existence of such a descent direction when the tangential stationarity, the positivity of the Lagrange multiplier ν or the normal growth condition is violated is shown in the dissertation [18]. Since the proof is rather technical without offering any major mathematical added benefit, we refer here to the corresponding result in the thesis, i.e., [18, Lemma 4.10].

Next, we examine the optimization problem which belongs directly to our saddle point system (29), i.e., only the active inequality constraints, again denoted by $\mathcal{I}$ (cf. section 3), are taken into account in addition to the equality constraints:

$$\begin{aligned} \min_{x\in\mathbb{R}^n, z\in\mathbb{R}^s} \quad & a^\top x + b^\top |\Sigma| z + \frac{1}{2} x^\top Q x \\ \text{s.t.} \quad & 0 = g + Ax + B|\Sigma|z + C\Sigma z \\ & 0 = P_{\mathcal{I}}(h + Dx + E|\Sigma|z + F\Sigma z) \\ & 0 = |\Sigma|z - \tilde{c} - \tilde{Z}x \ . \end{aligned} \tag{35}$$

Based on the search directions defined in Eq. (30), it is possible to reformulate problem (35). For this purpose, we denote by $f_Q(x,z) = a^\top x + b^\top |\Sigma| z + \frac{1}{2} x^\top Q x$ the target function and by $(\bar{x}, \bar{z})$ the current point to obtain

$$\begin{aligned} & f_Q(\bar{x}+\Delta x, \bar{z}+\Delta z) = a^\top(\bar{x}+\Delta x) + b^\top|\Sigma|(\bar{z}+\Delta z) + \frac{1}{2}(\bar{x}+\Delta x)^\top Q(\bar{x}+\Delta x) \\ = \ & \underbrace{a^\top \bar{x} + b^\top|\Sigma|\bar{z} + \frac{1}{2}\bar{x}^\top Q\bar{x}}_{=f_Q(\bar{x},\bar{z})=\text{const.}} + a^\top\Delta x + b^\top|\Sigma|\Delta z + \frac{1}{2}\Delta x^\top Q\Delta x + \underbrace{\frac{1}{2}\bar{x}^\top Q\Delta x + \frac{1}{2}\Delta x^\top Q\bar{x}}_{=\bar{x}^\top Q\Delta z} \\ = \ & (a^\top + \bar{x}^\top Q)\Delta x + b^\top|\Sigma|\Delta z + \frac{1}{2}\Delta x^\top Q\Delta x + f_Q(\bar{x},\bar{z}). \end{aligned} \tag{36}$$

Moreover, we consider only the equality constraints and active inequality constraints, yielding with $\varphi(\Delta x, \Delta z) := (a + Q\bar{x})^\top \Delta x + b^\top|\Sigma|\Delta z + \frac{1}{2}\Delta x^\top Q \Delta x$ the following problem

$$\begin{aligned} \min_{(\Delta x,\Delta z)\in\mathbb{R}^{n+s}} \quad & \varphi(\Delta x, \Delta z) \\ \text{s.t.} \quad & 0 = A\Delta x + B|\Sigma|\Delta z + C\Sigma\Delta z \ , \\ & 0 = P_{\mathcal{I}}^\top(D\Delta x + E|\Sigma|\Delta z + F\Sigma\Delta z) \ , \\ & 0 = |\Sigma|\Delta z - \tilde{Z}\Delta x \ . \end{aligned} \tag{37}$$

This optimization problem considers exactly the same constraints, i.e. the same active inequality constraints, as the saddle point system (29). The only difference is that here we minimize (37) along the search direction for fixed $\bar{x}$ and $\bar{z}$, whereas in the original problem (35) we search for the point where the minimum is attained. After this reformulation, we can show that this gives a descent direction.

Lemma 4.1. *Suppose that $(\Delta x^*, \Delta z^*)$ is the solution of (37) with $\Delta x^* \neq 0$ and let the zero vector be no solution of (37). Then the objective function $f_Q(\cdot,\cdot)$ is strictly decreasing along the direction $(\Delta x^*, \Delta z^*)$. If LIKQ holds, then this direction is unique.*

Proof. Since the zero vector is a feasible point but no solution of (37) and $(\Delta x^*, \Delta z^*)$ is a solution, one has that

$$\varphi(\Delta x^*, \Delta z^*) < \varphi(0) \quad \Rightarrow \quad (a + Q\bar{x})^\top \Delta x^* + b^\top|\Sigma|\Delta z^* + \frac{1}{2}\Delta x^* Q \Delta x^* < 0 \tag{38}$$

Since Q is positive definite, we have $\frac{1}{2}(\Delta x^*)^\top Q \Delta x^* > 0$ and it follows with Eq. (38):

$$(a + Q\bar{x})^\top \Delta x^* + b^\top |\Sigma| \Delta z^* < 0.$$

Therefore, Eq. (36) yields

$$\begin{aligned} f_Q(\bar{x} + \alpha \Delta x^*, \bar{z} + \alpha \Delta z^*) &= f_Q(\bar{x}, \bar{z}) + \underbrace{\alpha (a + Q\bar{x})^\top \Delta x^* + \alpha b^\top |\Sigma| \Delta z^* + \frac{1}{2}\alpha^2 \Delta x^* Q \Delta x^*}_{<0} \\ &< f_Q(x, z) \end{aligned}$$

for all $\alpha > 0$ sufficiently small. The uniqueness follows from the assumption that LIKQ holds true (cf. Section 3). $\square$

Now, we analyze the convergence of the Algo. 1.

Theorem 4.2. *Suppose that an optimization problem of the form* (CALOP) *is given and that $x \in \mathbb{R}^n$ is a feasible starting point for* (CALOP). *Let $Q = Q^\top \in \mathbb{R}^{n \times n}$ be a positive definite matrix. Then, Algo. 1 terminates after finitely many iterations at a minimizer of the quadratically penalized optimization problem*

$$\begin{aligned} \min_{x \in \mathbb{R}^n, z \in \mathbb{R}^s} \quad & a^\top x + b^\top z + \frac{1}{2} x^\top Q x \\ s.t. \quad & 0 = g + Ax + Bz + C|z| \\ & 0 \geq h + Dx + Ez + F|z| \\ & z = c + Zx + Mz + L|z| \; . \end{aligned} \tag{39}$$

Proof. Algo. 1 prioritizes the multiplier ω of the inequality constraints, see line 8 versus line 11 of Algo. 1. Therefore, as long as the signature vector σ does not change, the proposed approach resembles an active set method to solve QPs. Furthermore, we always ensure a decrease in the function value, see Lemma 4.1, and the fact that $\beta > 0$. Such an approach determines a minimizer of problems with the structure (26) in finitely many steps, see, e.g., [24, Chap. 16].

If the current iterate is a feasible signature stationary point of (26) on $\overline{\mathcal{F}}_\sigma$, Algo. 1 may change also the signature vector σ, see line 12, resulting in a change to a different polyhedron $\overline{\mathcal{F}}_{\tilde{\sigma}}$. However, since there are only finitely many polyhedra and the value of the function value is consistently reduced, Algo. 1 can modify the signature vector only finitely many times leading to a finite convergence of the overall algorithm. $\square$

The theorem considers the penalized version (39) of the original optimization task (CALOP). This ensures also that the optimization problem is bounded below such that a minimizer must exist. Hence, when Algo. 1 stops at a local minimizer of (39) in line 14, one has to check the first optimality condition (14) given in Theo. 3.8 to verify that the current point is also a minimizer of (CALOP). If this is not the case, one has to reduce the influence of the quadratic penalty term and start Algo. 1 again. If (CALOP) has a minimizer and the influence of the penalty term is driven to zero in finitely many steps this yields convergence to a minimizer of (CALOP) as proven next. In our numerical tests preformed so far, such a reduction was not necessary.

Iteration	x^i	σ^i	ω^i
0	(8.00, 3.00)	(1, -1, 1)	-1
1	(7.33, 3.66)	(1, -1, 0)	-1
2	(0.00, 0.00)	(1, 0, 0)	-1
3	(0.00, 0.00)	(1, 0, 0)	-1

Table 1: Optimization history of Algo. 1 for Ex. 5.1.

Theorem 4.3. *Let f be bounded from below on the feasible set given by* (CALOP)*, denoted by $\mathcal{F}$. Then, for $Q \to 0$ the solutions generated by Algo. 1 converge to the solution of the optimization problem* (CALOP)*.*

Proof. Since f is bounded from below it attains a minimum on $\mathcal{F}$. The optimality conditions (27b) to (27g) are independent from Q and therefore coincide with the corresponding optimality conditions of (CALOP) (cf. Theo. 3.8). Thus, the only optimality condition that depends on Q is stated in Eq. (27a). For reasons of continuity, if Q tends to zero, Eq. (27a) converges to the same optimality condition as given in Eq. (20). Thus, the solution generated by Algo. 1 coincides with the solution of (CALOP). □

5 Numerical results

To illustrate the algorithm proposed in this paper, we implemented Algo. 1 in Matlab and applied it to some constrained piecewise linear test problems.

Example 5.1. *Consider again the constrained optimization problem given in Ex. 3.1. For the starting point $x^0 = (8, 3)$, Fig. 5.1 shows the iterates generated by Algo. 1. Once more, the resulting kinks are given by the blue lines and the feasible set is marked by the red area. Four iterations are performed. The corresponding iterates are stated in Table 1 together with the signature vector σ and the signature vector of the constraints ω.*

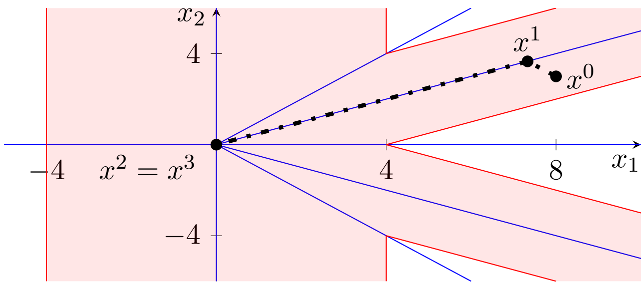


Figure 5: Iterates generated by Algo. 1 for Ex. 5.1.

Example 5.2. *(Constrained HUL) As mentioned already before, Hiriart-Urruty and Lemaréchal considered the piecewise linear and convex function $\varphi : \mathbb{R}^n \to \mathbb{R}$,*

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

To test our algorithm, we add the two constraints

$$\begin{aligned} H_1(x) &= -0.25x_1 - x_2 - 10 \le 0 \ , \\ H_2(x) &= 2 - 0.2|x_1 + 9| - |x_2 + 1| \le 0 \ , \end{aligned}$$

and choose the feasible starting point $x^0 = (9, -2.5)$. This optimization problem requires six switching variables, e.g., one has $n = 2$, $s = 6$, $m = 0$ and $p = 2$. Using Algo. 1, 15 iterations are needed.

Fig. 6 shows a plot of the resulting kinks originating from the objective function (blue lines) and from the constraints (cyan blue lines). The inequality constraints are marked by the red lines and therefore the red area represents the feasible set. Finally, the iterates generated by Algo. 1 are denoted by the black dots. In the plot only eight of the 15 iterations are marked. This is due to the fact that some of the iterations duplicate the point x when σ and ω are restricted or relaxed, i.e., kinks or constraints are activated or deactivated.

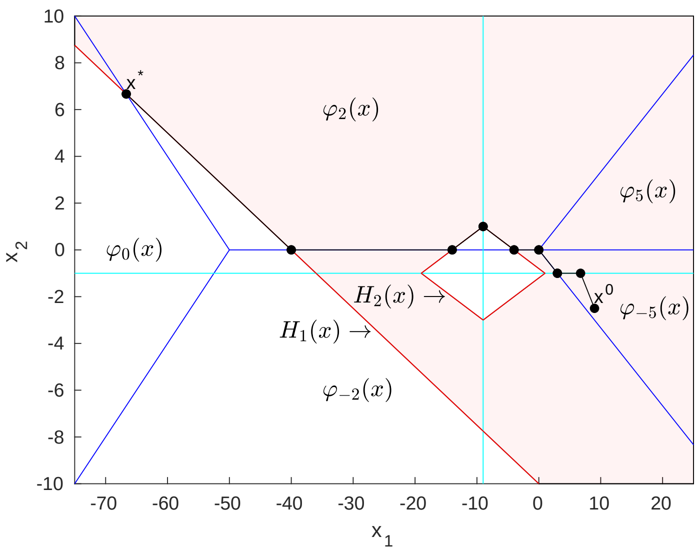


Figure 6: Iterates generated by Algo. 1 for Ex. 5.2.

Example 5.3. *(Constrained Rosenbrock-Nesterov II) According to [13], Nesterov suggested the Rosenbrock-like test function*

$$\varphi : \mathbb{R}^n \to \mathbb{R}, \quad \varphi(x) = \frac{1}{4}|x_1 - 1| + \sum_{i=1}^{n-1} |x_{i+1} - 2|x_i| + 1|$$

that is piecewise linear and nonconvex. It has the unique global minimizer $x^* = (1, 1, \ldots, 1) \in \mathbb{R}^n$ *and* $2^{n-1}-1$ *other Clarke stationary points non of which is a local minimizer. For the starting point*

$$x_1^0 = -1, \quad x_i^0 = 1 \quad \textit{for} \quad 2 \le i \le n \;,$$

the paper [9] contains numerical results and comparisons to other solvers showing that nonsmooth optimization algorithms may get stuck at one of these stationary points that are no minimizers. From the literature [13, 9] it is known that the selected starting point is particularly well suited, since from this initial point most algorithms first run through all stationary points. Since we consider constrained problems in this paper, we add the piecewise linear constraint

$$\sum_{i=1}^{n} |x_i - 1| \ge \frac{1}{2n} \;.$$

Hence, there is an n*-dimensional rhombus around the global optimum which is cut out of the* $\mathbb{R}^n$*. The remaining* $2^{n-1} - 1$ *stationary points are still feasible. To derive an abs-linear representation of this constrained optimization problem, we define* $s = 3\,n - 1$ *switching variables, namely*

$$z_i = x_i \quad \textit{for} \quad 1 \le i < n, \qquad z_{n+i} = x_{i+1} - 1 \quad \textit{for} \quad 0 \le i < n,$$
$$z_{2n+i} = x_{i+2} - 2\,|z_{i+1}| + 1 \quad \textit{for} \quad 0 \le i < n-1, \qquad z_{3n-1} = \frac{1}{4}|z_n| + \sum_{i=0}^{n-2} |z_{2n+i}| \;.$$

Hence, we obtain the matrices and vectors

$$Z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & I_{n-2} & 0 \\ 1 & 0 & 0 \\ 0 & I_{n-2} & 0 \\ 0 & 0 & 1 \\ 0 & I_{n-2} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \in \mathbb{R}^{s \times n}, \quad L = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ -2\,I_{n-1} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & \mathbf{1}^\top & 0 \end{bmatrix} \in \mathbb{R}^{s \times s}, \quad M = 0\,,$$

$$h = \tfrac{1}{2n}, \quad D = 0, \quad E = 0, \quad F = (\underbrace{0, \ldots, 0}_{n-1}, \underbrace{-1, \ldots, -1}_{n}, \underbrace{0, \ldots, 0}_{n})\,,$$

$$a = 0 \in \mathbb{R}^n, \quad b = e_{3n-1} \in \mathbb{R}^s, \quad c = (\underbrace{0, \ldots, 0}_{n-1}, \underbrace{-1, \ldots, -1}_{n}, \underbrace{1, \ldots, 1}_{n-1}, 0)$$

with $\mathbf{1} \in \mathbb{R}^{n-1}$ *as the vector with* 1 *in every component. Consider the point*

$$x_i^* = 1 - \frac{2^{i-1}}{2^n - 1} \cdot \frac{1}{2n} \in (0, 1) \quad \textit{for } 1 \le i \le n \;.$$

n	1	2	3	4	5	6	7	8	9	10	11	12
#	2	5	14	27	64	117	238	439	856	1685	3382	6807

n	13	14	15	16	17	18	19	20
#	13592	26285	42994	82995	131096	262173	605342	1119907

Table 2: Ex. 5.3: Number of iterations for different values of n.

Then one has

$$\sigma(x^*) = (\underbrace{1,\ldots,1}_{n-1},\underbrace{-1,\ldots,-1}_{n},\underbrace{0,\ldots,0}_{n})^\top \Rightarrow \alpha = \{2n,\ldots,3n-2\} \quad \textit{and} \quad \omega(x^*) = 0\ .$$

Note that the index $3n-1$ *is not contained in* α *according to Def. 3.4, since* z_{3n-1} *does not occur as an argument of an absolute value. Then, one obtains*

$$\begin{bmatrix} P_{\mathcal{I}}(D + E|\Sigma|\tilde{Z} + F\Sigma\tilde{Z}) \\ P_\alpha \tilde{Z} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & \cdots & \cdots & 1 \\ -2 & 1 & 0 & & & 0 \\ 0 & -2 & 1 & 0 & & 0 \\ & \ddots & \ddots & \ddots & \ddots & 0 \\ & & \ddots & \ddots & \ddots & 0 \\ & & & 0 & -2 & 1 \end{bmatrix} \in \mathbb{R}^{(n-1)\times n},$$

such that LIKQ holds. The optimality conditions (14), (15) *and* (16) *require*

$$\nu^\top = \lambda^\top P_\alpha^\top P_\alpha \tilde{Z}\ , \qquad \nu^\top = -\lambda^\top|\Sigma| \qquad \textit{and} \qquad |P_\alpha \lambda| \le -P_\alpha \tilde{L}^\top \lambda\ .$$

These three conditions hold for $\nu = 0 \in \mathbb{R}$ *and* $\lambda = 0 \in \mathbb{R}^s$*. Hence,* x_* *is a minimizer.*

For varying values of n*, the number of iterations required by Algo. 1 is shown in Table 2. The number of switches is given by* $s = 3n-1$*. As can be seen from the iteration counts, the number of visited polyedra is much less than the total number of polyhedra with definite signatures given by* 2^s*. We also applied MPBNGC solver [23] that is a multiobjective proximal bundle method for nonconvex, nonsmooth and generally constrained minimization. For* $n = 1$*, seven iterations are needed. Already for* $n = 2$*, the solver gets stuck after seven iterations at a stationary point. The same can be observed for larger values of* n*.*

Example 5.4. *As a fourth example, we consider a linear complementarity problem (LCP) given by*

$$Mx + q \ge 0 \quad \textit{and} \quad x^\top(Mx+q) = 0 \tag{40}$$

for $0 \le x \in \mathbb{R}^n$*,* $M \in \mathbb{R}^{n\times n}$ *and* $q \in \mathbb{R}^n$*. In [2], the LCP is formulated as a system of piecewise linear equations*

$$\min(x, Mx+q) = 0\ , \tag{41}$$

where the minimum operator acts componentwise. In the same paper, the authors present an algorithm that can be viewed as a semismooth Newton method and show nonconvergence for a special choice of the matrix M. They pointed out that the problem has a unique solution for any $q \in \mathbb{R}^n$ if and only if M is a $\boldsymbol{P}$-matrix, i.e., M has positive principal minors $\det M_{II} > 0$ for all nonempty $I \subset \{1, \ldots, n\}$.

To solve Eq. (41) with Algo. 1 we reformulate the problem (40) *as*

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{n} |\min(x_i, (Mx+q)_i)| \,.$$

For the matrix M, we set

$$M_3 \equiv \begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} \quad \textit{and} \quad M_4 \equiv \begin{bmatrix} 1 & 0 & \frac{1}{2} & \frac{4}{3} \\ \frac{4}{3} & 1 & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{4}{3} & 1 & 0 \\ 0 & \frac{1}{2} & \frac{4}{3} & 1 \end{bmatrix}$$

and $q = \mathbf{1}$ as the vector with 1 *in every component of appropriate dimension as considered also in [2]. As starting point we use the first unit vector in $\mathbb{R}^n$ as proposed in [2]. Then, Algo. 1 needs five iterations in both cases, i.e., for M_3 and M_4, respectively, to reach the solution* 0 *as zero vector of the appropriate dimension. In [2, Proposition 3.7] it is shown that the algorithm proposed in that paper does not converge but generates a circle of three resp. four reoccurring iterates.*

Bi-level problems, i.e., problems where a lower level optimization problem has to be solved and its solution impacts the upper level optimization problem, play an important role in many real-world applications and are closely related to linear complementarity problems as in Ex. 5.4. Here, we consider a bi-level problem, where all functions appearing as objective functions of the upper and lower level as well as all constraints are linear. For the lower level problem we use standard KKT theory to convert it into a set of equations and inequalities representing the necessary and sufficient optimality conditions for the lower level. Subsequently, these constraints substitute the lower level problem. However, the resulting complementarity condition is no longer a linear function. For the application of CASM, we can reformulate this constraint analogous to Eq. (41) as a piecewise linear function. Thus, the Lagrange multipliers from the lower problem also become optimization variables.

Example 5.5. *Consider the following linear bi-level problem taken from [28, Chap. 7]:*

$$\begin{aligned} \min_{x,y} \quad & 3x_1 + 2x_2 + y_1 + y_2 \\ \text{s.t.} \quad & x_1 + x_2 + y_1 + y_2 \leq 4 \,, \\ & y \in \operatorname*{argmin}_{\tilde{y}} \quad 4\tilde{y}_1 + \tilde{y}_2 \\ & \qquad \text{s.t.} \qquad 3x_1 + 5x_2 + 6\tilde{y}_1 + 2\tilde{y}_2 \geq 15 \,, \\ & x \in \mathbb{R}^2_{\geq 0}, \; y \in \mathbb{R}^2_{\geq 0} \,. \end{aligned}$$

We use the starting point

$$x = (2.5, 1.5) \ , \quad y = (0, 0) \ , \quad \textit{and} \quad \mu = (0, 4, 1) \ ,$$

where μ represents the Lagrange multiplier resulting from the lower level problem as described above. Table 3 shows the iterates when solving this problem with CASM. In [28], a structurally quite different algorithm is used to solve the problem, making it difficult to compare the effort. Both algorithms perform some preparatory work in that a pre-solve is performed before applying the algorithm proposed in [28] and a feasible starting point has to be determined for CASM. Subsequently, the algorithm presented in [28] requires three iterations, each of which requires the solution of two linear programs. CASM needed six iterations, where one system of equations with a 27×27 system matrix must be solved in each iteration. Both algorithms attain the same solution.

i	x^i	y^i	μ^i	σ^i	ω^i
0	(2.5, 1.5)	(0, 0)	(0.0, 4.0, 1.0)	(0, 1, 1)	(0, -1, -1, 0, 0, 0, 0, -1, -1)
1	(2.5, 1.5)	(0, 0)	(0.0, 4.0, 1.0)	(0, 1, 1)	(-1, -1, -1, 0, 0, 0, 0, -1, -1)
2	(0.0, 3.0)	(0, 0)	(0.0, 4.0, 1.0)	(0, 1, 1)	(-1, 0, -1, 0, 0, 0, 0, -1, -1)
3	(0.0, 3.0)	(0, 0)	(0.0, 4.0, 1.0)	(0, 1, 1)	(-1, 0, -1, 0, -1, 0, 0, -1, -1)
4	(0.0, 3.0)	(0, 0)	(0.0, 4.0, 1.0)	(0, 1, 1)	(-1, 0, -1, 0, -1, 0, -1, -1, -1)
5	(0.0, 3.0)	(0, 0)	(0.0, 4.0, 1.0)	(1, 1, 1)	(-1, 0, -1, 0, -1, 0, -1, -1, -1)
6	(0.0, 3.0)	(0, 0)	(0.5, 4.0, 0.0)	(1, 1, 0)	(-1, 0, -1, 0, -1, 0, -1, -1, 0)

Table 3: Optimization history of Algo. 1 for Ex. 5.5.

6 Summary and outlook

In this paper, we considered optimization problems with a piecewise linear target function and piecewise linear constraints as they arise for example in linear complementarity problems or certain bi-level optimization problems.

Using the approach of abs-linearization, we have shown an alternative proof in comparison to [14, 16] that we can verify the optimality of a given point with polynomial effort. This is in contrast to most optimality conditions available for nonsmooth optimization.

Furthermore, starting from the already known Active Signature Method to solve unconstrained piecewise linear optimization problems, we developed an extension for the constrained case. For this purpose, we adapted the idea of decomposing the $\mathbb{R}^n$ into polyhedra such that the constraints are taken into account. On one such polyhedron, the objective function was additionally penalized by a quadratic term ensuring the existence of a minimizer on each polyhedron. This minimizer can be determined using an adapted method to solve smooth quadratic problems. Employing the optimality conditions derived before, a switching strategy between the polyhedra was derived that ensures finite convergence of the overall algorithm. Numerical results for several test cases illustrate the performance of the resulting Constrained Active Signature Method.

The optimization problems solved in this paper have been of a purely academic nature. In the future, we want to apply the algorithm to larger problems stemming from realistic applications. For example, solution approaches for the optimization of gas networks yield constrained piecewise linear subproblems, cf. [20, 21, 1]. First promising results in this direction were already obtained, see [19]. The optimization problems considered there are of much larger dimension than the test examples in this paper having more than 500 optimization variables, 1000 constraints and almost 2000 switches.

One remaining challenge is the determination of a feasible starting point. For some real-world applications, such as the gas networks just mentioned, there are sometimes simple ways to find such a feasible starting point, cf. [19]. However, for other problems, such as general bi-level problem considered in the section on the numerical examples, determining a feasible starting point has turned out to be complicated. There, the reformulation of the lower level problem leads to new optimization variables corresponding to the Lagrange multipliers for which there are no intuitive starting values. The development of a suitable Phase-I method could help to overcome this challenge. An already established Phase-I method, as known for linear optimization problems [24], is usually not easily applicable, since the linear problems can be considered only on the polyhedra. Thus it can happen that on some polyhedra no feasible point exists at all. A very costly approach would then be to examine each polyhedron during a Phase-I method.

Furthermore, the Constrained Active Signature Method proposed in this paper could be used as solver for the inner loop problem of a SALMIN approach [4] extended for constrained piecewise smooth problems,where a local piecewise linear model is considered. However, similar to the smooth situation this might lead to non-feasible iterates in the outer loop dealing with the nonlinear problem. Hence, suitable strategies to handle this infeasibility have to be designed.

Acknowledgments

The authors thank the Deutsche Forschungsgemeinschaft for their support within Project B10 in the Sonderforschungsbereich/Transregio 154 *Mathematical Modelling, Simulation and Optimization using the Example of Gas Networks* (project ID: 239904186). In addition, the authors would like to express their sincere thanks to Marc Steinbach for his constructive and detailed feedback after reading this paper. The data that support the findings of this study are available from the corresponding author upon request. Conflicts of Interest: The authors declare no conflict of interest.

References

[1] D. Aßmann, F. Liers, and M. Stingl. “Decomposable robust two-stage optimization: an application to gas network operations under uncertainty”. In: *Networks* 74.1 (2019), pp. 40–61.

[2] I. Ben Gharbia and J. C. Gilbert. "Nonconvergence of the plain Newton-min algorithm for linear complementarity problems with a P-matrix". In: *Math. Program.* 134.2, Ser. A (2012), pp. 349–364.

[3] F. E. Curtis, T. Mitchell, and M. L. Overton. "A BFGS-SQP method for nonsmooth, nonconvex, constrained optimization and its evaluation using relative minimization profiles". In: *Optimization Methods and Software* 32.1 (2017), pp. 148–181.

[4] S. Fiege, A. Walther, and A. Griewank. "An algorithm for nonsmooth optimization by successive piecewise linearization". In: *Math. Program.* 177.1-2, Ser. A (2019), pp. 343–370.

[5] F. Fischer and C. Helmberg. "Dynamic graph generation and dynamic rolling horizon techniques in large scale train timetabling". In: *ATMOS'10.* 2010, pp. 45–60.

[6] X. Glorot, A. Bordes, and Y. Bengio. "Deep sparse rectifier neural networks". In: *Proceedings of the fourteenth international conference on artificial intelligence and statistics.* JMLR Workshop and Conference Proceedings. 2011, pp. 315–323.

[7] A. Griewank and A. Walther. "First- and second-order optimality conditions for piecewise smooth objective functions". In: *Optim. Methods Softw.* 31.5 (2016), pp. 904–930.

[8] A. Griewank. "On stable piecewise linearization and generalized algorithmic differentiation". In: *Optim. Methods Softw.* 28.6 (2013), pp. 1139–1178.

[9] A. Griewank and A. Walther. "Finite convergence of an active signature method to local minima of piecewise linear functions". In: *Optim. Methods Softw.* 34.5 (2019), pp. 1035–1055.

[10] A. Griewank and A. Walther. "Polyhedral DC decomposition and DCA optimization of piecewise linear functions". In: *Algorithms (Basel)* 13.7 (2020), Paper No. 166, 25.

[11] A. Griewank and A. Walther. "Relaxing kink qualifications and proving convergence rates in piecewise smooth optimization". In: *SIAM J. Optim.* 29.1 (2019), pp. 262–289.

[12] A. Griewank, A. Walther, S. Fiege, and T. Bosse. "On Lipschitz optimization based on gray-box piecewise linearization". In: *Math. Program.* 158.1-2, Ser. A (2016), pp. 383–415.

[13] M. Gürbüzbalaban and M. L. Overton. "On Nesterov's nonsmooth Chebyshev-Rosenbrock functions". In: *Nonlinear Anal.* 75.3 (2012), pp. 1282–1289.

[14] L. C. Hegerhorst-Schultchen. "Optimality conditons for abs-normal NLPs". PhD thesis. Gottfried Wilhelm Leibniz Universität, 2020.

[15] L. C. Hegerhorst-Schultchen, C. Kirches, and M. C. Steinbach. "Relations between Abs-Normal NLPs and MPCCs. Part 1: Strong Constraint Qualifications". In: *Journal of Nonsmooth Analysis and Optimization* Volume 2 (Feb. 2021).

[16] L. C. Hegerhorst-Schultchen and M. C. Steinbach. "On first and second order optimality conditions for abs-Normal NLP". In: *Optimization* 69.12 (2020), pp. 2629–2656. eprint: https://doi.org/10.1080/02331934.2019.1626386.

[17] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms. I.* Springer-Verlag, Berlin, 1993.

[18] T. Kreimeier. "Solving Constrained Piecewise Linear Optimization Problems by Exploiting the Abs-linear Approach". PhD thesis. Humboldt-Universität zu Berlin, 2023.

[19] T. Kreimeier, M. Kuchlbauer, F. Liers, M. Stingl, and A. Walther. *Towards the Solution of Robust Gas Network Optimization Problems Using the Constrained Active Signature Method*. Online: https://opus4.kobv.de/opus4-trr154/frontdoor/index/index/docId/448. 2021.

[20] M. Kuchlbauer, F. Liers, and M. Stingl. "Adaptive bundle methods for nonlinear robust optimization". In: *Informs Journal on Computing* (2021). Accepted.

[21] M. Kuchlbauer, F. Liers, and M. Stingl. *Outer approximation for mixed-integer nonlinear robust optimization*. Online: https://opus4.kobv.de/opus4-trr154/frontdoor/index/index/docId/414. 2021.

[22] F. Liers and M. Merkert. "Structural investigation of piecewise linearized network flow problems". In: *SIAM J. Optim.* 26.4 (2016), pp. 2863–2886.

[23] M. M. Mäkelä, N. Karmitsa, and O. Wilppu. "Proximal bundle method for nonsmooth and nonconvex multiobjective optimization". In: *Mathematical modeling and optimization of complex structures*. Vol. 40. Comput. Methods Appl. Sci. Springer, Cham, 2016, pp. 191–204.

[24] J. Nocedal and S. J. Wright. *Numerical optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006.

[25] W. de Oliveira. "Short Paper-A note on the Frank–Wolfe algorithm for a class of nonconvex and nonsmooth optimization problems". In: *Open Journal of Mathematical Optimization* 4 (2023), pp. 1–10.

[26] S. Scholtes. *Introduction to Piecewise Differentiable Equations*. SpringerBriefs in Optimization. New York, NY : Springer New York, 2012.

[27] Q. Tao, L. Li, X. Huang, X. Xi, S. Wang, and J. A. Suykens. "Piecewise linear neural networks and deep learning". In: *Nature Reviews Methods Primers* 2.1 (2022), pp. 1–17.

[28] H. Tuy, A. Migdalas, and P. Värbrand. "A quasiconcave minimization method for solving linear two-level programs". In: *J. Global Optim.* 4.3 (1994), pp. 243–263.