Noname manuscript No.
(will be inserted by the editor)

An Algorithm for Nonsmooth Optimization by Successive Piecewise Linearization

Sabrina Fiege · Andrea Walther · Andreas Griewank

the date of receipt and acceptance should be inserted later

Keywords Piecewise smoothness, Nonsmooth optimization, Algorithmic differentiation, Abs-normal form, Clarke stationary
Abstract We present an optimization method for Lipschitz continuous, piecewise smooth (PS) objective functions based on successive piecewise linearization. Since, in many realistic cases, nondifferentiabilities are caused by the occurrence of **abs**(), **max**(), and **min**(), we concentrate on these nonsmooth elemental functions. The method's idea is to locate an optimum of a PS objective function by explicitly handling the kink structure at the level of piecewise linear models. This piecewise linearization can be generated in its *abs-normal-form* by minor extension of standard algorithmic, or automatic differentiation tools. This paper first presents convergence results for the minimization algorithm developed. Numerical results including comparisons with other nonsmooth optimization methods then illustrate the capabilities of the proposed approach.

1 Introduction

Even today only very few practical methods for the minimization of Lipschitzian piecewise smooth functions $f : \mathbb{R}^n \mapsto \mathbb{R}$ are available. On convex objectives, the use of subgradients in combination with merely square summable step lengths yields only a sublinear rate of convergence, see, e.g., [17, Chap. 2]. Another option is to adapt quasi-Newton methods for the nonsmooth case, as proposed in [12]. A more reasonable rate of convergence can be expected from bundle methods, see, e.g., [2,10,11,15], but their performance is somewhat erratic. In [7], we proposed a new algorithm for computing a stationary point of a piecewise linear (PL) function. Following the seminal work of Hiriart-Urruty and Lemaréchal [9], we could demonstrate finite convergence in the convex case and verify it numerically on a

Sabrina Fiege · Andrea Walther
Department of Mathematics, Paderborn University, Paderborn,
E-mail: sfiege@math.uni-paderborn.de

Andreas Griewank
School of Mathematical Science and Information Technology, Yachaytech, Urcuqui, Ecuador

few selected test problems. Moreover, the basic method was formulated for general nonconvex piecewise linear problems with an additional proximal term, and it was found to always reach a stationary point in preliminary numerical experiments. The performance compared favorably to an adapted BFGS method with gradient sampling [12] and a recent bundle method implementation [15].

As our efforts to theoretically exclude the possibility of Zenon-like zigzagging on nonconvex functions failed, we have modified the original method by replacing a line-search with a positive definite QP solve as sketched in Sect. 3. That modification reduces the number of outer iterations and immediately implies finite convergence to a stationary point for general piecewise linear functions with a proximal term. The resulting numerical performance of our code module PLMin on piecewise linear functions is reported in Sect. 4 of this paper.

As already foreshadowed in [4], the main thrust of our algorithmic development is the minimization of piecewise smooth (PS) and semi-smooth functions by successive piecewise linearization using PLMin as inner solver. For a conservative updating strategy of the proximal term it had already been shown that successive piecewise linearization generates a subsequence that converges to a stationary point from within a compact level set. After some numerical experiments we developed the more aggressive updating strategy described in Sect. 3, which maintains global convergence in the above sense.

The final Sect. 4 contains numerical results for a wide range of test problems. A direct comparison with other nonsmooth solvers is difficult because they utilize much less information about the objective than our approach. However, this additional structural information given by the abs-normal form of PL functions [4] is easy to come by, not only on all the usual test problems but also on large scale applications from scientific computing. It can be obtained by an extension of algorithmic differentiation (AD) to evaluation codes involving smooth elemental functions and the absolute value function as well as the min and max operator as shown in [4].

2 Notation and Background

Throughout the paper, we will consider only objective functions $f : \mathbb{R}^n \mapsto \mathbb{R}$ that can be described by a sequence of elementary functions. We assume that these elemental functions are either the absolute value function or Lipschitz continuously differentiable in the domain $D \subset \mathbb{R}^n$ of interest. Using the reformulations

$$\min(v,u) = (v + u - \mathrm{abs}(v-u))/2 \quad \text{and} \tag{1}$$

$$\max(v,u) = (v + u + \mathrm{abs}(v-u))/2, \tag{2}$$

a quite large range of piecewise differentiable and locally Lipschitz continuous functions are covered. It follows from this assumption that the resulting objective function $f(x)$ is piecewise smooth in the sense of Scholtes [16, Chap. 4].

Conceptually combining consecutive smooth elemental functions into larger smooth elemental functions ψ_i, one obtains the reduced evaluation procedure shown in Tab. 1, where all evaluations of the absolute value function can be clearly identified and exploited. Here, $j \prec i$ collects all variables v_j that influence v_i directly and $j \prec i$ implies just for simplicity that $j < i$. As can be seen from this

v_{i-n}	$=$	x_i	$i = 1 \dots n$
z_i	$=$	$\psi_i(v_j)_{j \prec i}$	
σ_i	$=$	$\mathrm{sign}(z_i)$	$i = 1 \dots s$
v_i	$=$	$\sigma_i z_i = \mathrm{abs}(z_i)$	
$y \equiv v_{s+1}$	$=$	$\psi_{s+1}(v_j)_{j \prec s+1}$	

Table 1: Reduced evaluation procedure

table, $s \in \mathbb{N}$ will denote the actual number of evaluations of the absolute value function. Since the intermediate value z_i is used as the argument of the absolute value and hence causes also the switches in the corresponding derivative values, the vector $z = (z_i) \in \mathbb{R}^s$ is called switching vector defining also the signature vector $\sigma = (\sigma_i(x))_{i=1,\dots,s} \equiv (\mathrm{sign}(z_i(x)))_{i=1,\dots,s} \in \mathbb{R}^s$.

Given the class of piecewise smooth functions considered in this paper, it follows that they can be represented as

$$f(x) \in \{f_\sigma(x) : \sigma \in \mathcal{E} \subset \{-1, 0, 1\}^s\} \quad \text{at} \quad x \in \mathbb{R}^n,$$

where the selection functions f_σ are continuously differentiable on neighborhoods of points where they are active, i.e., coincide with f, as described in [16]. We will assume that all f_σ with $\sigma \in \mathcal{E}$ are essential in that their coincidence sets $\{f(x) = f_\sigma(x)\}$ are the closures of their interiors. The particular form of the index set $\mathcal{E} \subset \{-1, 0, 1\}^s$ stems from our function evaluation model described in Tab. 1. One has for the generalized subdifferential ∂f that

$$\partial f(x) \equiv \mathbf{conv}(\partial^L f(x)) \quad \text{with} \quad \partial^L f(x) \equiv \{\nabla f_\sigma(x) : f_\sigma(x) = f(x)\},$$

where the elements of $\partial^L f(x)$ are called the limiting gradients of f at x. A *directionally active gradient* g is given by

$$g \equiv g(x; d) \in \partial^L f(x) \quad \text{such that} \quad f'(x; d) = g^\top d, \tag{3}$$

where $f'(x; d)$ is the directional derivative of f at x in direction d and $g(x; d)$ equals the gradient $\nabla f_\sigma(x)$ of a locally differentiable selection function f_σ that coincides with f on a set, whose tangent cone at x contains d and has a nonempty interior.

To obtain a piecewise linearization of the objective function f, one has to construct for each elemental function a tangent approximation. For a given argument x and a direction Δx, we will use the elemental linearizations

$$\Delta v_i = \Delta v_j \pm \Delta v_k \qquad \text{for } v_i = v_j \pm v_k, \tag{4}$$

$$\Delta v_i = v_j * \Delta v_k + v_k * \Delta v_j \qquad \text{for } v_i = v_j * v_k, \tag{5}$$

$$\Delta v_i = \varphi'(v_j)_{j \prec i} * \Delta(v_j)_{j \prec i} \qquad \text{for } v_i = \varphi_i(v_j)_{j \prec i} \neq \mathrm{abs}(v_j), \tag{6}$$

$$\Delta v_i = \mathrm{abs}(v_j + \Delta v_j) - v_i \qquad \text{for } v_i = \mathrm{abs}(v_j). \tag{7}$$

The linearizations (4)–(6) are well-known, whereas the linearization (7) was proposed in [4]. These linearizations can be used to compute the increment $\Delta f(x; \Delta x)$ and therefore also the piecewise linearization

$$f_{PL,x}(\Delta x) \equiv f(x) + \Delta f(x; \Delta x) \tag{8}$$

of the original PS function f at a given point x with the argument Δx.

As shown in [5], any piecewise linear function $y = f_{PL}(\Delta x)$ with

$$f_{PL} : \mathbb{R}^n \to \mathbb{R}^m$$

can be expressed using the argument Δx and the resulting switching vector $z \in \mathbb{R}^s$ in the *abs-normal* form given by

$$\begin{bmatrix} z \\ y \end{bmatrix} = \begin{bmatrix} c_z \\ c_y \end{bmatrix} + \begin{bmatrix} Z & L \\ Y & J \end{bmatrix} \begin{bmatrix} \Delta x \\ |z| \end{bmatrix}, \tag{9}$$

where $c_z \in \mathbb{R}^s$, $c_y \in \mathbb{R}^m$, $Z \in \mathbb{R}^{s \times n}$, $L \in \mathbb{R}^{s \times s}$, $Y \in \mathbb{R}^{m \times n}$ and $J \in \mathbb{R}^{m \times s}$. The matrix L is strictly lower triangular, i.e., each z_i is an affine function of absolute values $|z_j|$ with $j < i$ and the input values Δx_k for $1 \le k \le n$. The matrices Y and J are row vectors in this optimization context, since we consider functions with $m = 1$. Correspondingly, c_y is only a real number instead of a vector. Defining the signature matrix

$$\Sigma \equiv \Sigma(\Delta x) \equiv \operatorname{diag}(\sigma(\Delta x)) \in \{-1, 0, 1\}^{s \times s}$$

for the switching variables of the piecewise linearization, one obtains for a fixed $\sigma \in \{-1, 0, 1\}^s$ and $|z| \equiv \Sigma z$ for z as in Eq. (9) from the first equation in this equation that

$$(I - L\Sigma)z = c_z + Z\Delta x \quad \text{and} \quad z = (I - L\Sigma)^{-1}(c_z + Z\Delta x). \tag{10}$$

Notice that due to the strict triangularity of $L\,\Sigma$ the inverse $(I - L\Sigma)^{-1}$ is well defined and polynomial in the entries of $L\,\Sigma$. Substituting this expression into the last equation of Eq. (9), it follows for the function value that

$$f_\sigma(x) \equiv \gamma_\sigma + g_\sigma^\top \Delta x \tag{11}$$

with

$$\gamma_\sigma = c_y + J\Sigma(I - L\Sigma)^{-1} c_z \quad \text{and} \quad g_\sigma^\top = Y + J\Sigma(I - L\Sigma)^{-1} Z.$$

That is, the gradient evaluation for the piecewise linearization reduces to the solve of a linear system with a triangular matrix. This will be exploited for a cheap gradient calculation in the inner loop of the optimization algorithm presented in the next section.

3 Successive Piecewise Linearization

As sketched already in [4], we propose the following algorithm to minimize Lipschitzian piecewise smooth functions:

Algorithm 1 (LiPsMin)

LiPsMin(x, q^0, q^{lb}, κ) // *Precondition:* $x \in \mathbb{R}^n$, $\kappa > 0$, $q^0 > 0$ *sufficiently great*

$x^0 = x$

for $k = 0, 1, 2, \ldots$

1. *Generate a PL model* $f_{PL,x^k}(.)$ *at the current iterate* x^k.
2. *Use PLMin(*x^k, Δx^k, q^k*) to solve the overestimated local probem*

$$\Delta x^k = \underset{\Delta x \in \mathbb{R}^n}{\arg\min} f_{PL,x^k}(\Delta x) + \frac{1}{2}(1+\kappa) q^k \|\Delta x\|^2.$$

3. Set $x^{k+1} = x^k + \Delta x^k$ *if* $f(x^k + \Delta x^k) < f(x^k)$ *and* $x^{k+1} = x^k$ *else.*
4. Compute

$$\hat{q}^{k+1} \equiv \hat{q}(x^k, \Delta x^k) \equiv \frac{2|f(x^{k+1}) - f_{PL,x^k}(\Delta x^k)|}{\|\Delta x^k\|^2}$$

and set $q^{k+1} = \max\{\hat{q}^{k+1}, \mu\, q^k + (1-\mu)\, \hat{q}^{k+1}, q^{lb}\}$ *with* $\mu \in [0,1]$.

As can be seen, the main ingredient of the approach is the successive piecewise linearization which was introduced in the last section. The local model will always be generated in Step 1 of LiPsMin. In step 2 an overestimation of a local subproblem

$$\hat{f}_x(\Delta x) \equiv f_{PL,x}(\Delta x) + \frac{\breve{q}}{2}\|\Delta x\|^2, \tag{12}$$

where $\breve{q}$ embraces all coefficients such as the penalty coefficients q and the overestimation $(1+\kappa)$ is solved by an inner loop which is discussed explicitly in Sec. 3.1. The overestimation is necessary to ensure lower boundedness of the model and to obtain the required convergence behavior. In this algorithmic specification we have not yet given a termination criterion so that the conceptual algorithm generates an infinite sequence of iterates $\{x^k\}$ that can be examined in the convergence analysis. Naturally, we would like that Algorithm 1 generates cluster points that are minimizers or at least stationary not only for the PL models but for the underlying PS objective. Here, we find the following relations.

Lemma 1

i) If the piecewise smooth function f *is locally minimal at* x*, then the quadratic model* $\hat{f}_x$ *is locally minimal at* $\Delta x = 0$ *for all* $q \geq 0$*.*
ii) If the quadratic model $\hat{f}_x$ *is Clarke stationary at* $\Delta x = 0$ *for one* $q \geq 0$*, then the piecewise smooth function* f *is Clarke stationary at* x*.*

Proof Note that according to Proposition 9 in [4] the subdifferential of the piecewise smooth function f at x contains that of the piecewise linearization evaluated in x at $\Delta x = 0$, i.e.,

$$\partial f(x) \supset \partial f_{PL,x}(0).$$

We define $h : \mathbb{R}^n \to \mathbb{R}$ as $h(\Delta x) := \frac{q}{2}\|\Delta x\|^2$ which is a twice continuously differentiable function with a unique minimizer at $\Delta x = 0$. The subdifferential of h is given by $\partial h(\Delta x) = \{2\Delta x\}$. Then, the quadratic model can be written as $\hat{f}_x(\Delta x) = f_{PL,x}(\Delta x) + h(\Delta x)$.

i) Let us assume for simplicity that f is locally minimal at x with $f(x) = 0$ and hence $\hat{f}_x(0) = 0$. Suppose that $\hat{f}_x(\cdot)$ is not minimal at 0 for some $q \geq 0$. Then we have for some Δx and $t > 0$

$$\hat{f}_x(t\Delta x) = t g_\sigma^\top \Delta x + o(t) < 0,$$

where we have used the directional differentiability of the piecewise linear model and g_σ is a suitable generalized gradient. Then it follows by the generalized Taylor expansion [4] that for sufficiently small t also

$$f(x + t\Delta x) - f(x) = t g_\sigma^\top \Delta x + o(t) < 0,$$

yielding a contradiction to the minimality of f at x.

ii) If $\hat{f}_x$ is Clarke stationary in $\Delta x = 0$, it implies that

$$0 \in \partial \hat{f}_x(0) = \partial(f_{PL,x}(0) + h(0)) \subseteq \partial f_{PL,x}(0) + \partial h(0).$$

Since $\partial h(0) = \{0\}$ one obtains that $0 \in \partial f_{PL,x}(0)$. By using the inclusion relation of the subdifferentials noted above this implies that also $0 \in \partial f(x)$ and therewith f is Clarke stationary in x. □

Throughout the rest of the paper, we will use the following example to illustrate our approach.

Example 1 We consider the piecewise smooth and nonconvex function

$$f : \mathbb{R}^2 \to \mathbb{R}, \quad f(x_1, x_2) = (x_2^2 - (x_1)_+)_+ \quad \text{with} \quad y_+ \equiv \max(0, y), \tag{13}$$

that is shown in Fig. 1. All points of nondifferentiablility are marked both in the domain of the function and in its value set.

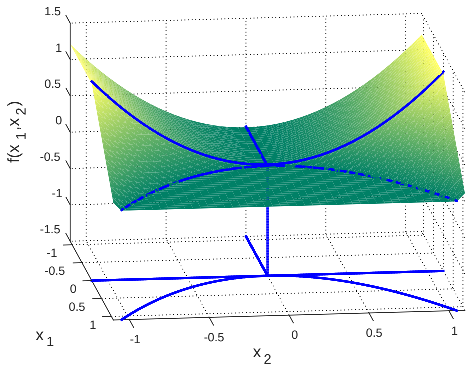

Fig. 1: PS function (13) with all points of nondifferentiability.

3.1 Minimization of the Piecewise Linear Subproblem with Proximal Term

In this section, we will analyze the structure of the local subproblems (12). Furthermore, we will present a method to compute stationary points of this subproblem. Since we consider the k-th iteration of Algo. 1 throughout this subsection, we use x, Δx, and q instead of x^k, Δx^k and q^k for simplicity.

The structure of the subproblem is given by its decomposition into polyhedra P_σ caused by the nondifferentiable points. For any PL local model, it follows by continuity that P_σ must be open but possibly empty if σ is *definite* in that all its components are nonzero. Generally we have for any nonempty P_σ

$$\dim(P_\sigma) \;\geq\; n + \|\sigma\|_1 - s = n - s + \sum_{i=1}^{s} |\sigma_i| \, . \tag{14}$$

When equality holds we call the signature σ *nondegenerate* and otherwise *critical.* In particular degenerate situations there may be some critical σ that are nevertheless *open* in that P_σ is open. The set of all polyhedra P_σ forms a directed acyclical graph, which is called a skeleton by Scholtes, see [16, Chapter 2].

We certainly have by definition of $\sigma = \sigma(x)$ for the closure of P_σ

$$\overline{P}_\sigma \;\subset\; \{x \in \mathbb{R}^n : f_{PL,x}(x) = f_\sigma(x)\},$$

where the selection functions f_σ are defined by Eq. (11). Note, that identity must hold in the convex case. In the nonconvex case f_σ may coincidentally be active, i.e., coincide with $f_{PL,x}$ at points in other polyhedra $P_{\tilde\sigma}$. In fact the coincidence sets may be the union of many polyhedral components but given the abs-normal form there is no need to deal with any of its arguments outside $\overline{P}_\sigma$. In particular f_σ is essentially active in the sense of Scholtes [16, Chapter 4.1] at all points in $\overline{P}_\sigma$ provided σ is open. Whether or not it is essentially active somewhere outside of $\overline{P}_\sigma$ is irrelevant and needs not be tested. To conform with the general concepts of piecewise smooth functions we may restrict f_σ to some open neighborhood of $\overline{P}_\sigma$ such that it cannot be essentially active outside P_σ. The corresponding signature vectors are given by

$$\mathcal{E} = \{\sigma \in \{-1,0,1\}^s : \emptyset \neq P_\sigma \text{ open}\}$$

and we will call them *essential.*

Example 2 We consider the piecewise smooth and nonconvex function defined in Ex. 1. Its piecewise linearization generated at a base point $\bar{x}$ and with the argument Δx is given by

$$f_{PL,\bar{x}}(\Delta x) = \frac{1}{2}\left(z_2 + \frac{1}{2}|z_2|\right) \tag{15}$$

$$\text{with} \quad z_1 = |\bar{x}_1 + \Delta x_1| \quad \text{and} \quad z_2 = \bar{x}_2^2 + 2\bar{x}_2 \Delta x_2 - \frac{1}{2}\left(\bar{x}_1 + \Delta x_1 + |z_1|\right).$$

The domain of the piecewise linearization (15) is decomposed by two absolut value functions into four open polyhedra with corresponding nonzero signatures $\sigma = (\pm 1, \pm 1)$ as can be seen in Fig. 2.

Generally, we will describe the polyhedral structure primarily in terms of the signature vectors σ, as shown in Ex. 2. They have a partial order, which is nicely reflected in the corresponding polyhedra as follows.

Proposition 1 (Polyhedral structure in terms of signature vectors)
(i) The signature vectors are partially ordered by the precedence relation

$$\sigma \preceq \tilde\sigma \;:\Longleftrightarrow\; \sigma_i^2 \leq \tilde\sigma_i\, \sigma_i \quad \text{for } 1 \leq i \leq s. \tag{16}$$

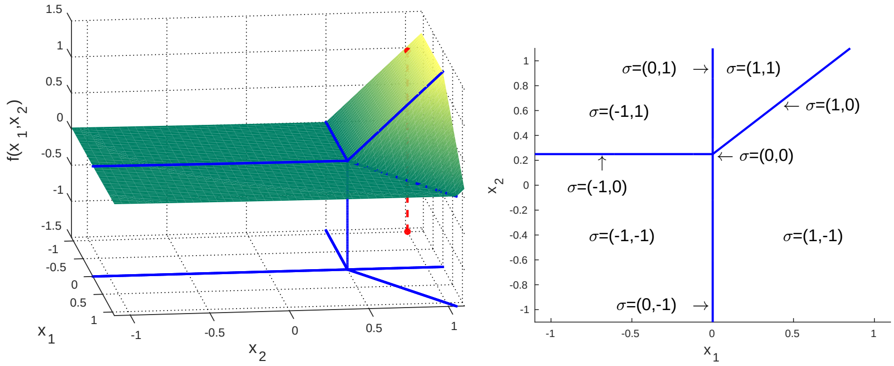

Fig. 2: Left: Piecewise linearization of Eq. (13) evaluated in $\bar{x} = (-1, 0.5)$, Right: The decomposition of its domain and its signatures

(ii) *The closure* $\overline{P}_\sigma$ *of any* P_σ *is contained in the* extended closure

$$\hat{P}_\sigma \equiv \{\tilde{x} \in \mathbb{R}^n : \sigma(x) \preceq \sigma\} \supset \overline{P}_\sigma \tag{17}$$

with equality holding unless $P_\sigma = \emptyset$.

(iii) *The essential signatures* $\mathcal{E}$ *are exactly the maximal elements amongst all nonempty signatures, i.e.,*

$$\mathcal{E} \ni \sigma \prec \tilde{\sigma} \implies P_{\tilde{\sigma}} = \emptyset \quad and \quad \hat{P}_\sigma = \hat{P}_{\tilde{\sigma}},$$

we will call such $\tilde{\sigma}$ extended essential.

(iv) *For any two signatures* σ *and* $\tilde{\sigma}$ *we have the equivalence*

$$\hat{P}_\sigma \subset \hat{P}_{\tilde{\sigma}} \iff \sigma \preceq \tilde{\sigma}.$$

(v) *Each polyhedron intersects only the extended closures of its successors*

$$P_\sigma \cap \hat{P}_{\tilde{\sigma}} \neq \emptyset \implies \sigma \preceq \tilde{\sigma}.$$

(vi) *The closures of the essential polyhedra form a polyhedral decomposition in that*

$$\bigcup_{\sigma \in \mathcal{E}} \hat{P}_\sigma = \mathbb{R}^n.$$

Proof See [7, Prop. 4.2] □

For the general scenario we have Algo. 2 to compute a stationary point for step 2 of Algo. 1. The essential difference to the *true descent algorithm* introduced in [7, Algo. 4] is the solution of a quadratic subproblem instead of computing a critical step multiplier for a given direction. An optimization strategy to find a minimizer of Eq. (12) solves a sequence of special QPs along a path of essential polyhedra.

Assume that an initial signature vector σ^0 corresponding to an essential polyhedron was identified, then the first quadratic subproblem is given by

$$\delta x_0 = \underset{\delta x \in \mathbb{R}^n}{\arg\min} f_{\sigma^0}(\delta x) + \frac{\breve{q}}{2}\|\delta x\|^2, \tag{18}$$

$$\text{s.t.} \quad z_i(x) + \nabla z_i(x)^\top \delta x \begin{cases} \leq 0 & \text{if } \sigma_i^0 < 0 \\ \geq 0 & \text{if } \sigma_i^0 > 0 \end{cases} \qquad \text{for} \quad i = 1, \ldots, s\ ,$$

where $z_i(x)$ is the i-th component of the switching variable of $z(x)$ and $\nabla z_i(x)$ the corresponding gradient. In order to solve the j-th subproblem the previous solutions Δx_l, $l = 0, ..., j-1$ have to be included such that the relationship between the current essential polyhedron P_{σ^j} and the base point x is maintained. Hence, one obtains the following subproblem

$$\delta x_j = \underset{\delta x \in \mathbb{R}^n}{\arg\min} f_{\sigma^j}\left(\Delta x_j + \delta x\right) + \frac{\breve{q}}{2}\left\|\Delta x_j + \delta x\right\|^2, \tag{19}$$

$$\text{s.t.} \quad z_i(x) + \nabla z_i(x)^\top \left(\Delta x_j + \delta x\right) \begin{cases} \leq 0 & \text{if } \sigma_i^j < 0 \\ \geq 0 & \text{if } \sigma_i^j > 0 \end{cases} \qquad \text{for} \quad i = 1, \ldots, s\ ,$$

with $\Delta x_j = \sum_{l=0}^{j-1} \delta x_l$.

By solving these special quadratic subproblems for a fixed σ^j, one can easily characterize the points $\tilde{x}$ in the extended closure $\hat{P}_{\sigma^j}$ that fulfill the system of inequalities. Both, the selection function f_{σ^j} defined in Eq. (11) and the gradient of the switching vector z can be evaluated via the abs-normal form as explained in the previous section. Minimizing f_{σ^j} subject to these constraints by any QP solver will yield the information whether P_{σ^j} is empty and otherwise whether f_{σ^j} is bounded on it, or not. Due to the proximal term added to the piecewise linear local model the objective function is positive definite quadratic on $\hat{P}_{\sigma^j}$. Nevertheless, the quadratic subproblem (19) can have a large number of constraints. Since typically many of these constraints are inactive, it is reasonable to develop a QP-solver including an appropiate warmstart.

The above approach can be summerized as follows, where we use the base point x and the step Δx as in Algo. 1 for clarity. The base point x and the quadratic coefficient q serve as input variables. The increment Δx is the output parameter.

Algorithm 2 (PLMin)

PLMin(x, Δx, q) // *Precondition:* $x, \Delta x \in \mathbb{R}^n$, $q \geq 0$

Set $\Delta x_0 = 0$. *Identify* $\sigma^0 = \sigma(x)$.
For $j = 0, 1, 2, ...$

1. *Determine solution* δx_j *of local QP* (19) *on current polyhedron* P_{σ^j}.
2. *Update* $\Delta x_{j+1} = \Delta x_j + \delta x_j$.
3. *Compute direction* d *by ComputeDesDir(*Δx_{j+1}, q, $G = \{g_{\sigma^j}\}$*).*
4. *If* $\|d\| = 0$: *STOP.*
5. *Identify new polyhedron* $P_{\sigma^{j+1}}$ *using direction* d.

return $\Delta x = \Delta x_{j+1}$

The main remaining challenge is to decide how to change the signature at a minimizer Δx_{j+1} in order to move to a neighboring polyhedron where the function value decreases in step 5 of Algo. 2. In other words we have to find a descent direction d at Δx_{j+1} and a signature σ^{j+1} such that $P_{\sigma^{j+1}}$ contains $\Delta x_{j+1} + \tau d$ for small positive τ. For that we can employ the following computation of a descent direction:

Algorithm 3 (Computation of Descent Direction)

ComputeDesDir(Δx_{j+1}**,** $\check{q}$**,** G**,** d**)**
// Precondition: $\Delta x_{j+1} \in \mathbb{R}^n$*,* $\check{q} \geq 0$*,* $\emptyset \neq G \subset \partial^L f_{PL,x}(\Delta x_{j+1})$

repeat
{ $d = -\mathbf{short}(\check{q}\Delta x_{j+1}, G)$
$g = g(\Delta x_{j+1}; d)$
$G = G \cup \{g\}$
}
until $(g + \check{q}\,\Delta x_{j+1})^\top d \leq -\beta \|d\|^2$
Set $G = \emptyset$*.*
return d

In this algorithm, G is a subset of the limiting subdifferential of the PL function $f_{PL,x}$ at the current iterate Δx_{j+1}. Initially it contains the gradient g_{σ^j} of the current selection function f_{σ^j}. The direction d is defined as $d = -\text{short}(qx, G)$ with

$$\text{short}(h, G) = \arg\min \left\{ \|d\| \;\middle|\; d = \sum_{j=1}^{m} \lambda_j g_j - h,\; g_j \in G,\; \lambda_j \geq 0,\; \sum_{j=1}^{m} \lambda_j = 1 \right\}. \tag{20}$$

Subsequently the bundle G gets augmented by further directionally active gradients $g(x; d)$ as defined in Eq. (3) corresponding to neighboring polyhedra. The solution of Eq. (20) is realized as the solution of a quadratic problem in the implementation of Algo. 2.

A very similar computation was already proposed in [7, Algo. 2], where also the finite termination of this algorithm is shown. However, for the general case considered here, we only want to identify a polyhedron $P_{\sigma^{j+1}}$ that provides descent compared to the current polyhedron. Hence, we introduce here the additional multiplier $\beta \in (0, 1)$ to relax the descent condition compared to [7, Algo. 2].

Algo. 2 converges to a stationary point Δx^* after finitely many steps, since the argument space is divided only into finitely many polyhedra, the local model $\hat{f}_x$ is bounded below and the function value is decreased each time we switch from one polyhedron to another.

3.2 Convergence Results for LiPsMin

To prove the convergence of Algo. 1 to a Clarke stationary point, we suppose that our piecewise smooth objective function $f : \mathbb{R}^n \to \mathbb{R}$ is bounded below and has a bounded level set $\mathcal{N}_0 \equiv \{x \in \mathbb{R}^n :\ f(x) \leq f(x^0)\}$ with x^0 the starting point of the generated sequence of iterates. Hence, the level set is compact. Furthermore, we assume that f satisfies all the assumptions of Sec. 2 on an open neighborhood $\tilde{\mathcal{N}}_0$ of $\mathcal{N}_0$. In [4] it was proven that the piecewise linearization $f_{PL,x}$ yields a second order approximation of the underlying function f. Therewith, it holds

$$\begin{aligned} f(x+\Delta x) &= f(x) + \Delta f(x;\Delta x) + \mathcal{O}(\|\Delta x\|^2) \\ &= f(x) + \Delta f(x;\Delta x) + c\|\Delta x\|^2 \end{aligned} \tag{21}$$

with the coefficient $c \in \mathbb{R}$. Subsequently, this coefficient is set as $c := \frac{1}{2}\tilde{q}$. The coefficient $\tilde{q}(x;\Delta x)$ can be computed for certain x and Δx. However, it is possible that $\tilde{q}(x;\Delta x)$ is negative and thus, the local quadratic model is not bounded below. Therefore, the coefficient $\hat{q}(x;\Delta x)$ is chosen as

$$\hat{q}(x;\Delta x) \equiv |\tilde{q}(x;\Delta x)| = \frac{2|f(x+\Delta x) - f(x) - \Delta f(x;\Delta x)|}{||\Delta x||^2}. \tag{22}$$

By doing so, one obtains from of Eq. (21) for all descent directions Δx the estimate

$$f(x+\Delta x) - f(x) \le \Delta f(x;\Delta x) + \frac{1}{2}\hat{q}(x;\Delta x)\|\Delta x\|^2 \le 0. \tag{23}$$

In [4, Prop. 1] it was proven as well that there exists a monotonic mapping $\bar{q}(\delta) : [0,\infty) \to [0,\infty)$ such that for all $x \in \mathcal{N}_0$ and $\Delta x \in \mathbb{R}^n$

$$\frac{2|f(x+\Delta x) - f(x) - \Delta f(x;\Delta x)|}{||\Delta x||^2} \le \bar{q}(||\Delta x||) \tag{24}$$

under the assumptions of this section. This holds on the one side because if the line segment $[x, x+\Delta x]$ is fully contained in $\tilde{\mathcal{N}}_0$, then the scalar $\bar{q}(\|\Delta x\|)$ denotes the constant of [4, Prop. 1]. On the other hand those steps Δx for which the line segment $[x, x+\Delta x]$ is not fully contained in $\tilde{\mathcal{N}}_0$ must have a certain minimal size, since the base points x are restricted to $\mathcal{N}_0$. Then the denominators in Eq. (24) are bounded away from zero so that $\bar{q}(||\Delta x||)$ exists.

Since $\bar{q}$ is a monotonic descending mapping which is bounded below, it converges to some limit $\bar{q}^* \in (0,\infty)$. Nevertheless $\bar{q}$ will generally not be known, so that we approximate it by estimates, referred to as quadratic coefficients throughout. From now on, we will mark elements of sequences with a superscript index. We generate the sequences of iterates $\{x^k\}_{k\in\mathbb{N}}$ with $x^k \in \mathcal{N}_0$ and corresponding steps $\{\Delta x^k\}_{k\in\mathbb{N}}$ with $\Delta x^k \in \mathbb{R}^n$ by Algo. 1 and consistently update the quadratic coefficient starting from some $q^0 > 0$ according to

$$q^{k+1} = \max\{\hat{q}^{k+1},\ \mu\, q^k + (1-\mu)\,\hat{q}^{k+1},\ q^{lb}\} \tag{25}$$

with $\hat{q}^{k+1} := \hat{q}(x^k;\Delta x^k)$, $\mu \in [0,1]$ and $q^{lb} > 0$ is a lower bound. Then the following lemma holds.

Lemma 2 *Under the general assumptions of this section, one has:*

a) The sequence of steps $\{\Delta x^k\}_{k\in\mathbb{N}}$ *exists.*
b) The sequences $\{\Delta x^k\}_{k\in\mathbb{N}}$ *and* $\{\hat{q}^k\}_{k\in\mathbb{N}}$ *are uniformly bounded.*
c) The sequence $\{q^k\}_{k\in\mathbb{N}}$ *is bounded.*

Proof a) By minimizing the supposed upper bound $\Delta f(x^k;\Delta x)+\frac{1}{2}q^k(1+\kappa)\|\Delta x\|^2$ on $f(x^k+\Delta x)-f(x^k)$ at least locally we always obtain a step

$$\Delta x^k \equiv \arg\min_s(\Delta f(x^k;s)+\frac{1}{2}q^k(1+\kappa)\|s\|^2).$$

A globally minimizing step Δx^k must exist since $\Delta f(x^k;s)$ can only decrease linearly so that the positive quadratic term always dominates for large $\|s\|$. Moreover, Δx^k vanishes only at first order minimal points x^k where $\Delta f(x^k;s)$ and $f'(x^k;s)$ have the local minimizer $s=0$. Of course, this is unlikely to happen.

b) It follows from $q^k \geq q^{lb} > 0$ and the continuity of all quantities on the compact set $\mathcal{N}_0$ that the step size $\delta \equiv \|\Delta x\|$ must be uniformly bounded by some $\bar{\delta}$. This means that the $\hat{q}$ are uniformly bounded by $\bar{q}\equiv\bar{q}(\bar{\delta})$.

c) Obviously, the sequence $\{q^k\}_{k\in\mathbb{N}}$ is bounded below by q^{lb}. Considering the first two arguments of Eq. (25), one obtains that $q^{k+1}=\hat{q}^{k+1}$ and $q^{k+1}>q^k$ if $\hat{q}^{k+1} > \mu\ q^k + (1-\mu)\ \hat{q}^{k+1}$. Respectively, if $\hat{q}^{k+1} \leq \mu\ q^k + (1-\mu)\ \hat{q}^{k+1}$, one obtains $q^{k+1}\geq\hat{q}^{k+1}$ and $q^{k+1}\leq q^k$. This means that the maximal element of the sequence is given by a $\hat{q}^j$ with $j\in\{1,...,k+1\}$ and thus bounded by $\bar{q}(\|\Delta x^j\|)$. Therefore, the sequence $\{q^k\}_{k\in\mathbb{N}}$ is bounded above. □

The proof of Lemma 2 c) gives us the important insight that $q^{k+1}\geq\hat{q}^{k+1}$ holds. With these results we can now prove the main convergence result of this paper.

Theorem 4 *Let $f:\mathbb{R}^n\to\mathbb{R}$ be a piecewise smooth function as described at the beginning of Sec. 2 which has a bounded level set $\mathcal{N}_0=\{x\in\mathbb{R}^n \mid f(x)\leq f(x^0)\}$ with x^0 the starting point of the generated sequence of iterates $\{x^k\}_{k\in\mathbb{N}}$.*

Then a cluster point x^* of the infinite sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by Algo. 1 exists. All cluster points of the infinite sequence $\{x^k\}_{k\in\mathbb{N}}$ are Clarke stationary.

Proof The sequence of steps $\{\Delta x^k\}_{k\in\mathbb{N}}$ is generated by solving the overestimated quadratic problem in step 2 of Algo. 1 of the form

$$\Delta x^k = \arg\min_s(\Delta f(x;s)+\frac{1}{2}(1+\kappa)q^k\|s\|^2).$$

Unless x^k satisfies first order optimality conditions the step Δx^k satisfies

$$\Delta f(x^k;\Delta x^k)+\frac{1}{2}(1+\kappa)q^k\|\Delta x^k\|^2<0. \tag{26}$$

Therewith, one obtains from Eq. (23)

$$f(x^k+\Delta x^k)-f(x^k)\leq\frac{1}{2}\left[q^{k+1}-(1+\kappa)q^k\right]\|\Delta x^k\|^2. \tag{27}$$

where $\hat{q}^{k+1}\leq q^{k+1}$ holds as a result to Eq. (25) and due to Eq. (26) one has $\Delta f(x^k;\Delta x^k)\leq-\frac{1}{2}q^k(1+\kappa)\|\Delta x^k\|^2$. The later inequality can be overestimated by applying the limit superior $\bar{q}=\limsup_{k\to\infty}q^{k+1}$ as follows

$$f(x^k+\Delta x^k)-f(x^k)\leq\frac{1}{2}\left[\bar{q}-(1+\kappa)q^k\right]\|\Delta x^k\|^2.$$

Considering a subsequence of $\{q^{k_j}\}_{j\in\mathbb{N}}$ converging to the limit superior, it follows that for each $\epsilon > 0$ a $\bar{j} \in \mathbb{N}$ exists such that for all $j \geq \bar{j}$ one obtains $\|\bar{q} - q^{k_j}\| < \epsilon$. Therewith the overestimated local problem provides that the term $\bar{q} - (1 + \kappa)q^{k_j} < 0$. Since the objective function f is bounded below on $\mathcal{N}_0$, infinitely many significant descent steps can not be performed and thus $f(x^{k_j} + \Delta x^{k_j}) - f(x^{k_j})$ has to converge to 0 as j tends towards infinity. As a consequence, the right hand side of Eq. (27) has to tend towards 0 as well. Therefore, the subsequence $\{\Delta x^{k_j}\}_{j\in\mathbb{N}}$ is a null sequence. Since the level set $\mathcal{N}_0$ is compact, the sequence $\{x^{k_j}\}_{j\in\mathbb{N}}$ has a subsequence that tends to a cluster point x^*. Hence, a cluster point x^* of the sequence $\{x^k\}_{k\in\mathbb{N}}$ exists.

Assume that the subsequence $\{x^{k_j}\}$ of $\{x^k\}$ converges to a cluster point. As shown above the corresponding sequence of penalty coefficients $\{\Delta x^{k_j}\}_{j\in\mathbb{N}}$ converges to zero if j tends to infinity. Therewith, one can apply Lemma 1 at the cluster point x^*, where it was proven that if $\hat{f}_x$ is Clarke stationary at $\Delta x = 0$ for one $q \geq 0$, then the piecewise smooth function f is Clarke stationary in x yielding the assertion. □

4 Numerical Results

The nonsmooth optimization method LiPsMin introduced in this paper will be tested in the following section. Therefore, we introduce piecewise linear and piecewise smooth test problems in the Sec. 4.1 and 4.2. In both cases the test set contains convex and nonconvex test problems. In Sec. 4.3 results of numerous optimization runs will be given and compared to other nonsmooth optimization software.

4.1 Piecewise linear test problems

The test set of piecewise linear problems comprises:

1. **Counterexample of HUL**

$$f(x) = \max\{-100,\ 3x_1 \pm 2x_2,\ 2x_1 \pm 5x_2\}, \quad (x_1,\ x_2)^0 = (9,\ -2).$$

2. **MXHILB**

$$f(x) = \max_{1\leq i\leq n} \left| \sum_{j=1}^{n} \frac{x_j}{i+j-1} \right|, \quad x_i^0 = 1, \quad \text{for all} \quad i = 1, ..., n.$$

3. **Max1**

$$f(x) = \max_{1\leq i\leq n} |x_i|, \quad x_i^0 = i, \quad \text{for all} \quad i = 1, ..., n.$$

4. **Second Chebyshev-Rosenbrock**

$$f(x) = \frac{1}{4}|x_1 - 1| + \sum_{i=1}^{n-1} |x_{i+1} - 2|x_i| + 1|$$

$x_i^0 = -0.5$, when $\mod(i, 2) = 1$, $i = 1, ..., n$ and
$x_i^0 = 0.5$, when $\mod(i, 2) = 0$, $i = 1, ..., n$.

Tab. 2 provides further information about the test problems such as the optimal value f^* of the function, the dimension n and the number of absolute value functions s occuring during the function evaluation depending on the dimension n. With s given in this way the relation of n and s can be given as well. Additionally, a reference is given for each test problem.

	n	s	$s \sim n$	f^*	properties	ref.
1	2	$2n$	$n < s$	-100	PL, convex	[9]
2	any	$2n-1$	$n \leq s$	0	PL, convex	[8]
3	any	$2n-1$	$n \leq s$	0	PL, convex	[15]
4	any	$2n-1$	$n \leq s$	0	PL, nonconvex	[3]

Table 2: Information about piecewise linear test problems

4.2 Piecewise smooth test problems

The test set of piecewise smooth problems is listed below.

5. **MAXQ**

$$f(x) = \max_{1 \leq i \leq n} x_i^2,$$

$$x_i^0 = i, \text{ for } i = 1, ..., n/2 \quad \text{and}$$
$$x_i^0 = -i, \text{ for } i = n/2+1, ..., n.$$

6. **Chained LQ**

$$f(x) = \sum_{i=1}^{n-1} \max \left\{ -x_i - x_{i+1},\ -x_i - x_{i+1} + (x_i^2 + x_{i+1}^2 - 1) \right\}$$

$$x_i^0 = -0.5, \text{ for all } i = 1, ..., n.$$

7. **Chained CB3 II**

$$f(x) = \max \{f_1(x), f_2(x), f_3(x)\},$$

$$\text{with} \quad f_1(x) = \sum_{i=1}^{n-1} \left(x_i^4 + x_{i+1}^2 \right), \quad f_2(x) = \sum_{i=1}^{n-1} \left((2 - x_i)^2 + (2 - x_{i+1})^2 \right)$$

$$\text{and} \quad f_3(x) = \sum_{i=1}^{n-1} \left(2e^{-x_i + x_{i+1}} \right),$$

$$x_i^0 = 2, \quad \text{for all} \quad i = 1, ..., n.$$

8. **MAXQUAD**

$$f(x) = \max_{1 \leq i \leq 5} \left(x^\top A^i x - x^\top b^i \right)$$

$$A_{kj}^i = A_{jk}^i = e^{j/k} \cos(jk) \sin(i), \quad \text{for} \quad j < k, \quad j, k = 1, ..., 10$$
$$A_{jj}^i = \tfrac{j}{10} |\sin(i)| + \textstyle\sum_{k \neq j} \left| A_{jk}^i \right|,$$
$$b_j^i = e^{j/i} \sin(ij),$$

$$x_i^0 = 0, \text{ for all } i = 1, ..., 10.$$

9. **Number of active faces**

$$f(x) = \max_{1\leq i\leq n} \left\{ g\left(-\sum_{j=1}^{n} x_j\right),\ g(x_i) \right\}, \quad \text{where } \ g(y) = \ln\left(|y|+1\right).$$

$x_i^0 = 1$, for all $i = 1, ..., n$.

10. **Chained Cresent I**

$$f(x) = \max\left\{f_1(x), f_2(x)\right\},$$

$$\text{with} \quad f_1(x) = \sum_{i=1}^{n-1} \left(x_i^2 + (x_{i+1}-1)^2 + x_{i+1} - 1\right)$$

$$\text{and} \quad f_2(x) = \sum_{i=1}^{n-1} \left(-x_i^2 - (x_{i+1}-1)^2 + x_{i+1} + 1\right),$$

$x_i^0 = -1.5$, when $\mod(i,2) = 1$, $i = 1, ..., n$ and
$x_i^0 = 2$, when $\mod(i,2) = 0$, $i = 1, ..., n$.

11. **Chained Cresent II**

$$f(x) = \sum_{i=1}^{n-1} \max\left\{f_{1,i}(x), f_{2,i}(x)\right\},$$

with $f_{1,i}(x) = x_i^2 + (x_{i+1}-1)^2 + x_{i+1} - 1,$

and $f_{2,i}(x) = -x_i^2 - (x_{i+1}-1)^2 + x_{i+1} + 1,$

$x_i^0 = -1.5$, when $\mod(i,2) = 1$, $i = 1, ..., n$ and
$x_i^0 = 2$, when $\mod(i,2) = 0$, $i = 1, ..., n$.

12. **First Chebyshev-Rosenbrock**

$$f(x) = \frac{1}{4}(x_1-1)^2 + \sum_{i=1}^{n-1} \left|x_{i+1} - 2x_i^2 + 1\right|$$

$x_i^0 = -0.5$, when $\mod(i,2) = 1$, $i = 1, ..., n$ and
$x_i^0 = 0.5$, when $\mod(i,2) = 0$, $i = 1, ..., n$.

In Tab. 3 further information about the test problems are given, compare Tab. 2.

	n	s	$s \sim n$	f^*	properties	ref.
5	any	$n-1$	$s < n$	0	PS, convex	[8]
6	any	$n-1$	$s < n$	$-(n-1)2^{1/2}$	PS, convex	[8]
7	any	2	$s \leq n$	$2(n-1)$	PS, convex	[8]
8	10	$5(n^2-n+1)$	$n < s$	-0.8414083	PS, convex	[13]
9	any	$n+1$	$n < s$	0	PS, nonconvex	[8]
10	any	2	$s \leq n$	0	PS, nonconvex	[8]
11	any	$n-1$	$s \leq n$	0	PS, nonconvex	[8]
12	any	$n-1$	$s < n$	0	PS, nonconvex	[3]

Table 3: Information about piecewise smooth test problems

4.3 Performance results of LiPsMin and comparison with other nonsmooth optimization methods

In the following, the introduced routine LiPsMin is compared with the nonsmooth optimization routines MPBNGC, i.e., a proximal bundle method described in [15], and the quasi-Newton type method HANSO described in [12].

The idea of the bundle method MPBNGC is to approximate the subdifferential of the objective function at the current iterate by collecting subgradients of previous iterates and storing them into a bundle. Thus, more information about the local behavior of the function is available. To reduce the required storage the amount of stored subgradients has to be restricted. Therefore an aggregated subgradient is computed from several previous subgradients so that these subgradients can be removed without losing their information. For a detailed description see [14], [15].

The quasi-Newton type method HANSO combines the BFGS method with an inexact line search and the gradient sampling approach described in [1]. The gradient sampling approach is a stabilized steepest descent method. At each iterate the corresponding gradient and additional gradients of nearby points are evaluated. The descent direction is choosen as the vector with the smallest norm in the convex hull of these gradients.

As stopping criterions of routines LiPsMin and PLMin we used $\epsilon = 1e - 8$ and the maximal iteration number $maxIter = 1000$. In the implementation of Algo. 1 we chose the parameter $\mu = 0.9$. Under certain conditions, e.g., a high amount of active constraints at the optimal point, it is reasonable to add a termination criterion that considers the reduction of the function value in two consecutive iterations, i.e., $|f(x_k) - f(x_{k+1})| < \epsilon$. For the bundle method MPBNGC we choose the following parameter settings. The maximal bundle size equals the dimension n. If the considered test function is convex then the parameter `gam` equlas 0 otherwise `gam` is set to 0.5. Further stopping criterions are the number of iterations $nIter = 10000$ (Iter), the number of function and gradient evaluations NFASG= 10000 ($\#f = \#\nabla f$) and the final accuracy `EPS`= $1e{-}8$. For HANSO we choose $normtol = 1e - 8$, $evaldist = 1e - 4$, $maxit = 10000$ (Iter) and the sampling radii $[10\ 1\ 0.1] * evaldist$.

Performance results of piecewise linear test problems

First, we consider the problems of the piecewise linear test set, see subsection 4.1. The results of the piecewise linear and convex problems are presented in Tab. 4–6. Each table contains all results of a single test problem generated by the three different optimization routines mentioned. The columns of the tables give the dimension n of the problem, the final function value f^*, the number of function evaluations $\#f$, the number of gradient evaluations $\#\nabla f$ and number of iterations (HANSO: number of BFGS iterations + number of gradient sampling (GS)). Additionally, the inital penalty coefficient q^0 of LiPsMin is given. Since we assume that all our test problems are bounded below and we consider piecewise linear problems, the initial penalty coefficient is choosen $q^0 = 0$. For the test problem MXHILB the additional stopping criterion considering the function value reduction was added.

	n	q^0	f^*	$\#f$	$\#\nabla f$	Iter
LiPsMin	2	0	-100	3	14	2
HANSO	2	–	-100	9	9	3
MPBNGC	2	–	-100	7	7	6

Table 4: Results for Counterexample of HUL

In Fig. 3 a comparison of the behavior of optimization runs generated by LiPsMin, HANSO, and MPBNGC is illustrated. As intended LiPsMin uses the additional information of the polyhedral decomposed domain efficiently in order to minimize the number of iterations. As a consequence the optimization run computed by LiPsMin is more predictable and purposeful than the runs computed by HANSO and MPBNGC, which seem to be rather erratic. This behavior is characteristic for all piecewise linear problems solved by LiPsMin. In contrast to the first three test problems the 2nd Chebyshev-Rosenbrock function is nonconvex. The corresponding results are given in Tab. 7. Most optimization routines failed to detect the unique minimizer. The detected points are Clarke stationary points.

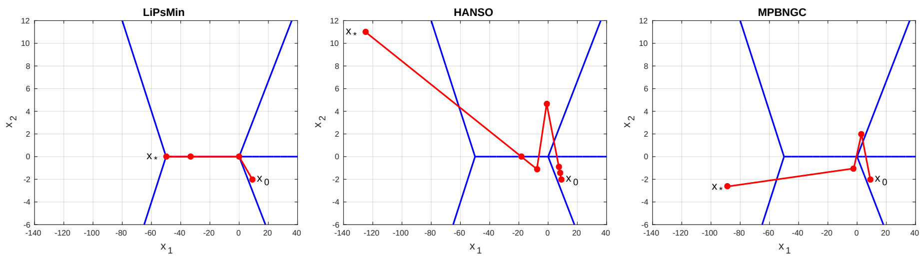

Fig. 3: Optimization runs of test problem HUL performed by LiPsMin, HANSO and MPBNGC.

	n	q^0	f^*	$\#f$	$\#\nabla f$	Iter
LiPsMin	2	0	5.6e-17	3	7	2
	5	0	2.7e-10	3	35	2
	10	0	5.6e-10	3	26	2
	20	0	4.7e-9	3	34	2
	50	0	3.0e-9	3	20	2
	100	0	2,1e-12	3	8	2
HANSO	2	–	1.6e-2	10191	10191	5 + 3GS
	5	–	5.7e-3	11678	11678	4 + 3GS
	10	–	8.8e-3	14320	14320	2 + 3GS
	20	–	1.2e-1	17953	17953	3 + 3GS
	50	–	1.8e-1	26841	26841	3 + 3GS
	100	–	4.4e-2	38484	38484	3 + 3GS
MPBNGC	2	–	4.1e-15	40	40	37
	5	–	1.4e-1	10000	10000	103
	10	–	1.5e-3	10000	10000	3347
	20	–	1.2e-2	10000	10000	5010
	50	–	3.3e-1	10000	10000	3338
	100	–	4.0e-1	10000	10000	3338

Table 5: Results for MXHILB

To distinguish minimizers and Clarke stationary points that are not optimal, new optimiality conditions were established in [6]. These optimality conditions are based on the *linear independent kink qualification* (LIKQ) which is a generalization of LICQ familiar from nonlinear optimization. It is shown in the mentioned article that the 2nd Chebyshev-Rosenbrock function satisfies LIKQ globally, i.e., throughtout $\mathbb{R}^n$. Herewith we can adapt Algo. 2 for piecewise linear functions which satisfy LIKQ globally as given in Algo. 5. The algorithm is merely based on the LIKQ conditions and does not yet use the mentioned optimality conditions. The call of the routine ComputeDesDir() in Algo. 2 is replaced by a reflection of the signature vector σ of the current polyhedron into the opposing polyhedron by switching all active signs from 1 to -1 or vice versa.

	n	q^0	f^*	$\#f$	$\#\nabla f$	Iter
LiPsMin	2	0	0	3	7	2
	5	0	0	3	10	2
	10	0	0	3	15	2
	20	0	0	3	25	2
	50	0	0	3	203	2
	100	0	0	3	404	2
HANSO	2	–	1.5e-5	21	21	17
	5	–	7.6e-6	28	28	21
	10	–	7.6e-6	38	38	26
	20	–	3.8e-6	59	59	37
	50	–	3.8e-6	119	119	67
	100	–	1.9e-6	220	220	118
MPBNGC	2	–	3.4e-9	67	67	42
	5	–	1.7e-16	17	17	15
	10	–	1.2e-13	29	29	27
	20	–	3.8e-11	56	56	54
	50	–	1.0e-12	123	123	121
	100	–	1.7e-9	176	176	165

Table 6: Results for Max1

Algorithm 5 (PLMin_Reflection)

PLMin_Reflection(x,Δx) // *Precondition:* $x,\ \Delta x \in \mathbb{R}^n,\ \Delta x = 0$

1. *Determine solution* δx *of local QP* (19) *on current* P_σ.
2. *If* $\|\delta x\| > \epsilon$
   Compute new σ *by switching all active signs.*
   Update $\Delta x = \Delta x + \delta x$.
   Go to 1.
   else
   STOP.

return Δx

The results of Algo. 5 applied on the 2nd Chebyshev-Rosenbrock function are given in Tab. 8. As stopping criterion we used the accuracy $\epsilon = 1e-8$. The minimizer was reached for $n = 2, 5, 10, 20$. Each evaluated gradient ∇f represents an open polyhedron. From Tab. 2 we know that $s = 2n-1$ absolute value functions occure during the evaluation of the 2nd Chebyshev-Rosenbrock function. That means that for $n = 20$ there are up to $2^s = 2^{39}$ open polyhedron and therewith only a vanishing amount of these possible polyhedra was entered by Algo. 5 as indicated by the number of gradient evaluations.

	n	q^0	f^*	$\#f$	$\#\nabla f$	Iter
LiPsMin	2	0	1,29e-11	4	11	3
	5	0	1.9e-1	5	53	4
	10	0	4.0e-1	4	42	3
	20	0	4.0e-1	3	45	2
	50	0	4.0e-1	3	57	2
	100	0	4.0e-1	3	120	2
HANSO	2	–	3.8e-7	211	211	61
	5	–	2.5e-1	13169	13169	2521 + 3GS
	10	–	4.0e-1	1691	1691	515 + 3GS
	20	–	4.0e-1	23269	23269	2177 + 3GS
	50	–	4.0e-1	29608	29608	463 + 3GS
	100	–	4.0e-1	40955	40955	290 + 3GS
MPBNGC	2	–	1.7e-16	80	80	52
	5	–	2.5e-1	10000	10000	3561
	10	–	4.0e-1	10000	10000	9807
	20	–	4.0e-1	66	66	65
	50	–	4.0e-1	188	188	187
	100	–	3.5e-1	251	251	249

Table 7: Results for 2nd Chebyshey-Rosenbrock

n	q^0	f^*	# f	$\#\nabla f$	Iter
2	0	1.3e-11	3	6	2
5	0	1.1e-16	3	17	2
10	0	1.6e-14	3	414	2
20	0	1.1e-16	4	419438	3

Table 8: Numerical results for 2nd Chebyshev-Rosenbrock with reflection

Performance results of piecewise smooth test problems

In the following, we consider the problems of the piecewise smooth test set, as introduced in Sec. 4.2. The results of the piecewise smooth problems are presented in Tab. 9–15. Each table contains all results of a single test problem generated by the three different optimization routines mentioned above. The columns of the tables are the same as before. The inital penalty coefficient is chosen as $q_0 = 0.1$ in most cases. In the other cases it is chosen as $q_0 = 1$.

Since the stopping criteria of HANSO are quite different from the criteria of the other two methods, the accuracy of the final function value f^* differs, i.e., in some cases it is strongly increased. Because of that the iteration count is disproportionately high in these cases. We mark these cases by (*Iter) where Iter indicates the number of iterations that was necessary to obtain a final function value with comparable accuracy.

	n	q^0	f^*	$\#f$	$\#\nabla f$	Iter
LiPsMin	2	0.1	2.3e-9	27	156	26
	5	0.1	1.8e-9	36	486	35
	10	0.1	2.7e-9	34	939	33
	20	0.1	1.9e-9	36	1980	35
	50	0.1	1.4e-8	58	8011	57
	100	0.1	3.5e-8	117	32359	116
HANSO	2	–	3.2e-19	18	18	16 (*9)
	5	–	3.0e-19	242	242	116 (*47)
	10	–	6.2e-17	787	787	352 (*88)
	20	–	1.1e-16	1362	1362	637 (*221)
	50	–	2.1e-16	4409	4409	1906 (*494)
	100	–	3.0e-16	8922	8922	3991 (*1023)
MPBNGC	2	–	7.6e-9	15	15	14
	5	–	3.1e-9	60	60	49
	10	–	3.4e-9	126	126	34
	20	–	2.6e-9	244	244	222
	50	–	3.8e-9	577	577	549
	100	–	4.5e-9	1118	1118	1083

Table 9: Results for MAXQ

	n	q^0	f^*	$\#f$	$\#\nabla f$	Iter
LiPsMin	2	0.1	-1.41421	10	26	9
	5	0.1	-5.65685	47	471	46
	10	0.1	-12.7278	15	128	18
	20	0.1	-26.8701	15	258	14
	50	0.1	-69.2965	15	646	14
	100	0.1	-140.007	15	1341	14
HANSO	2	–	-1.41421	370	370	51 + 3GS
	5	–	-5.65685	778	778	52 + 3GS
	10	–	-12.7279	3920	3920	100 + 3GS
	20	–	-26.8701	18548	18548	165 + 3GS
	50	–	-69.2965	28239	28239	274 + 3GS
	100	–	-140.007	41354	41354	416 + 3GS
MPBNGC	2	–	-1.41421	9	9	8
	5	–	-5.65685	34	34	30
	10	–	-12.7279	40	40	33
	20	–	-26.8701	63	63	61
	50	–	-69.2965	143	143	108
	100	–	-140.007	468	468	273

Table 10: Results for Chained LQ

In Tab. 9–12 the results of the piecewise smooth and convex test problems are presented. The additional stopping criterion of LiPsMin considering the function value reduction was applied for test problem MAXQ. A large number of optimitzation runs detected successfully minimal solutions. The bundle method MPBNGC stopped once because the maximal number of function and gradient evaluations was reached, see Tab. 11. HANSO stopped once with a quite low accuracy of the final function value f^*, see Tab. 12 for $n = 10$. The number of function value and gradient evaluations are hardly comparable as a consequence of the varying underlying information. However, several optimization runs performed by HANSO, see, e.g., Tab. 10, and LiPsMin, see, e.g., Tab. 9 resulted in a higher number of

	n	q^0	f^*	$\#f$	$\#\nabla f$	Iter
LiPsMin	2	1	2.00000	12	72	11
	5	1	8.00000	69	530	68
	10	1	18.0000	67	515	66
	20	1	38.0000	63	482	62
	50	1	98.0000	61	465	60
	100	1	198.000	59	449	58
HANSO	2	–	2.00000	407	407	65 + 3 GS
	5	–	8.00000	655	655	96 + 3GS
	10	–	18.0000	756	756	71 + 3GS
	20	–	38.0000	4023	4023	80 + 3GS
	50	–	98.0000	380	380	79 + 3GS
	100	–	198.000	894	894	69 + 3GS
MPBNGC	2	–	2.00001	10000	10000	9999
	5	–	8.00000	35	35	34
	10	–	18.0000	46	46	45
	20	–	38.0000	42	42	41
	50	–	98.0000	60	60	59
	100	–	198.000	41	41	40

Table 11: Results for Chained CB3 II

	n	q^0	f^*	$\#f$	$\#\nabla f$	Iter
LiPsMin	10	0.1	-8.414083	48	426	47
HANSO	10	–	-8.413940	2955	2955	1 + 3GS
MPBNGC	10	–	-8.414083	40	40	39

Table 12: Results for MAXQUAD

function and gradient evalutions than the other two respective routines, whereas the iteration numbers of all three routines are of comparable order of magnitude.

In Tab. 13–16 the results of the piecewise smooth and nonconvex test problems are presented. From general theory one can expect that these problems are difficultly to solve and indeed the results are not that clear as those results of the previously considered test problems. The results of test problems Number of active faces and Chained Cresent I given in Tab. 13 and Tab. 14 are encouraging. However, the final function values generated by HANSO are less accurate than expected and MPBNGC terminates several times because the maximum number of function and gradient evaluation was reached. Comparing test problems Chained Cresent I and II one can see how minor changes in the objective function influence the optimization results. The required number of iterations of all three optimizations methods increased.

As in the piecewise linear case the Chebyshev-Rosenbrock function seems to be more difficult than the other test problems. Only a few optimization runs detected the minimizer $f^* = 0$ with sufficiant accuracy. Most runs locate points that are no minimizers, but only Clarke stationary points. Consequently, these points fulfill termination criteria. To distinguish between those stationary but not minimal points it is again necessary to adapted the termination critera by using the new optimality conditions introduced in [6]. This adaption will be part of our future work.

The above results illustrate that in the piecewise smooth case the introduced optimization method LiPsMin compares well with the state of the art optimization

	n	q^0	f^*	$\#f$	$\#\nabla f$	Iter
LiPsMin	2	0.1	6.7e-16	3	7	2
	5	0.1	2.2e-16	4	8	3
	10	0.1	4.2e-15	4	8	3
	20	0.1	8.2e-15	5	9	4
	50	0.1	2.5e-14	9	13	8
	100	0.1	6.8e-14	14	18	13
HANSO	2	–	2.1e-6	23	23	8
	5	–	1,3e-5	24	24	11
	10	–	8.4e-5	23	23	11
	20	–	3.2e-5	25	25	9
	50	–	2.4e-5	27	27	11
	100	–	1.3e-4	29	29	11
MPBNGC	2	–	8.2e-13	20	20	14
	5	–	2.2e-10	12	12	10
	10	–	7.6e-9	20	20	15
	20	–	2.0e-7	10000	10000	9995
	50	–	6.4e-7	10000	10000	6428
	100	–	4.4e-5	10000	10000	9993

Table 13: Results for Number of Active Faces

	n	q^0	f^*	$\#f$	$\#\nabla f$	Iter
LiPsMin	2	1	7.0e-13	56	327	55
	5	1	8.0e-13	61	347	60
	10	1	9.1e-13	64	375	63
	20	1	9.5e-13	65	381	64
	50	1	1.1e-13	149	875	148
	100	1	62.1e-13e-13	92	543	91
HANSO	2	–	0	1409	1409	51 + 3GS (*28)
	5	–	0	198	198	57 + 3GS (*26)
	10	–	0	229	229	49 + 3GS (*27)
	20	–	0	266	266	48 + 3GS (*29)
	50	–	1.7e-15	358	358	52 + 3GS (*28)
	100	–	1.1e-15	703	703	47 + 3GS (*29)
MPBNGC	2	–	1.4e-8	52	52	45
	5	–	4.2e-9	78	78	57
	10	–	1.1e-8	49	49	40
	20	–	4.5e-9	65	65	50
	50	–	5.7e-9	167	167	83
	100	–	4.2e-9	96	96	66

Table 14: Results for Chained Cresent I

software as HANSO and MPBNGC. These results can be confirmed by Fig. 4–5 that give a first hint of the convergence rate of LiPsMin. Each figure correspondes to one test problem and it shows how the function value $f(x_k)$ of the k-th iteration decreases during an optimization run for $n = 10$ and $n = 100$. Considering the results of Fig. 4 LiPsMin seems to converge quadratic under certain conditions. These conditions will be analyzed in detail in future work.

5 Conclusion and Outlook

In [7] we proposed a method for the optimization of Lipschitzian piecewise smooth functions by successive piecewise linearization. The central part of that previous

	n	q^0	f^*	$\#f$	$\#\nabla f$	Iter
LiPsMin	2	0.1	6.4e-13	53	304	52
	5	0.1	8.3e-13	62	695	61
	10	0.1	5.8e-13	64	1228	63
	20	0.1	9.1e-13	64	2231	63
	50	0.1	7.0e-13	65	5260	64
	100	0.1	7.9e-13	65	10251	64
HANSO	2	–	0	4262	4262	51 + 3GS (*28)
	5	–	8.9e-16	2854	2854	346 + 3GS (*69)
	10	–	5.2e-15	8665	8665	238 + 3GS (*131)
	20	–	1.1e-14	19156	19156	292 + 3GS (*154)
	50	–	7.1e-7	27450	27450	92 + 3GS
	100	–	4.2e-7	39458	39458	111 + 3GS
MPBNGC	2	–	1.4e-8	52	52	45
	5	–	4.1e-9	80	80	79
	10	–	5.1e-9	196	196	195
	20	–	7.0e-9	488	488	485
	50	–	7.1e-9	519	519	518
	100	–	8.0e-9	733	733	684

Table 15: Results for Chained Cresent II

	n	q^0	f^*	$\#f$	$\#\nabla f$	Iter
LiPsMin	2	0.1	2.2e-14	308	1814	307
	5	0.1	6.4e-2	1001	15095	1000
	10	0.1	0.81744	1001	30371	1000
	20	0.1	0.81814	1001	42150	1000
	50	0.1	0.81814	6	256	5
	100	0.1	0.81814	6	542	5
HANSO	2	–	3.8e-7	211	211	61
	5	–	0.221199	11633	11633	2521 + 3GS
	10	–	0.399414	6441	6441	515 + 3 GS
	20	–	0.40000	23269	23269	2177 + 3GS
	50	–	0.40000	29608	29608	463 + 3GS
	100	–	0.40000	40915	40915	290 + 3GS
MPBNGC	2	–	8.7e-9	143	143	75
	5	–	0.169945	10000	10000	3509
	10	–	0.626470	10000	10000	9786
	20	–	0.818136	201	201	200
	50	–	0.630266	353	353	352
	100	–	0.061094	2283	2283	1842

Table 16: Results for 1st Chebyshev-Rosenbrock

article was the concept of piecewise linearization and the exploitation of the therewith gained information. An approach of the outer routine LiPsMin was introduced and tested. In the current work we gave an overview of a refined version of the LiPsMin method.

While computing the results of the piecewise linear examples by the inner routine PLMin in [7], it became obvious that the computation of the critical step multiplier was not as numerically accurate as required. Therefore, it was replaced by the more efficient solution of the quadratic subproblem introduced in Sec. 3.1. For this adapted inner routine of LiPsMin we confirmed convergence in finitely many iteration. A first version of the LiPsMin algorithm was introduced in [4]. In the current version a more aggressive updating strategy of the penalty multiplier

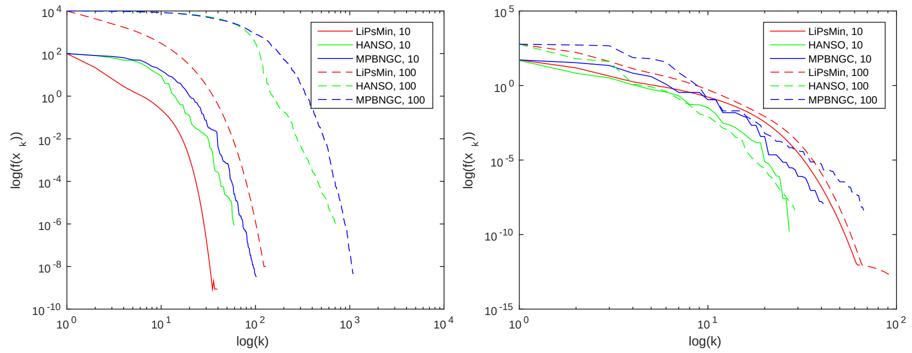

Fig. 4: Convergence behavior of LiPsMin, HANSO and MPBNGC for MAXQ (Left) and Chained Cresent I (Right) with $n = 10$ and $n = 100$.

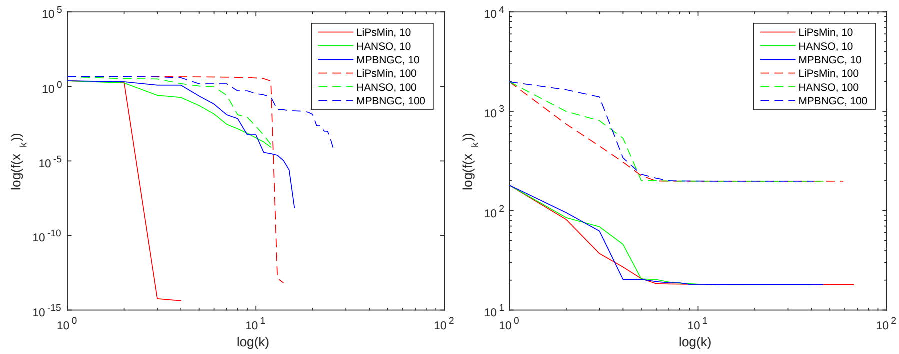

Fig. 5: Convergence behavior of LiPsMin, HANSO and MPBNGC for Number of Active Faces (Left) and Chained CB3 II (Right) with $n = 10$ and $n = 100$.

q was applied. In Sec. 3.2 it was proven that LiPsMin combined with the new updating strategy maintains its global convergence to a stationary point.

The performance results of LiPsMin in the piecewise linear case, see Sec. 4.3, confirmed well our expectation. The incorporation of additional information gained by the abs-normal form leads to a pointed and predictable descent trajectory. In the piecewise smooth case the performance results generated by LiPsMin also compared well with the state of the art optimitzation software tools MPBNGC and HANSO. Both the tables and the figures of Sec. 4 affirmed this conclusion. The results also indicated that LiPsMin converges quadratically under certain condition. These conditions have to be analyzed in future work.

Nevertheless, the results also illustrated that it is useful to check if stationary points are also minimal points. That is why it should be an important question of future work how to incorporate the optimality conditions of [6] such that minimizers can be uniquely identified. Hence, it would be benefical to gain more information about the polyhedral decomposition of the domain, such as convexity properties of the function. The additional information can be used to identify the subsequent polyhedron more efficiently which is especially relevant when the considered function is high dimensional or includes a higher number of absolute value function evalutions.

References

1. J. Burke, A. Lewis, and M. Overton. A robust gradient sampling algorithm for nonsmooth nonconvex optimization. *SIAM J. Optim.*, 15(3):751–779, 2005.
2. W. de Oliveira and C. Sagastizábal. Bundle methods in the XXIst century: A birds'-eye view. *Pesquisa Operacional*, 34(3):647–670, 2014.
3. J.-L. Goffin. Subgradient optimization in nonsmooth optimization (including the Soviet Revolution). *Documenta Mathematica*, Extra Vol.:277–290, 2012.
4. A. Griewank. On stable piecewise linearization and generalized algorithmic differentiation. *Opt. Meth. and Softw.*, 28(6):1139–1178, 2013.
5. A. Griewank, J.-U. Bernt, M. Radons, and T. Streubel. Solving piecewise linear systems in abs-normal form. *Linear Algebra and its Applications*, 471:500–530, 2015.
6. A. Griewank and A. Walther. First and second order optimality conditions for piecewise smooth objective functions. *Optimization Methods and Software*, 31(5):904–930, 2016.
7. A. Griewank, A. Walther, S. Fiege, and T. Bosse. On Lipschitz optimization based on gray-box piecewise linearization. *Mathematical Programming, Series A*, pages 1–33, 2015.
8. M. Haarala, K. Miettinen, and M.M. Mäkelä. New limited memory bundle method for large-scale nonsmooth optimization. *Optimization Methods and Software*, 19(6):673–692, 2004.
9. J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I.* Springer, 1993.
10. N. Karmitsa and M.M. Mäkelä. Limited memory bundle method for large bound constrained nonsmooth optimization: convergence analysis. *Optimization Methods and Software*, 25(6):895–916, 2010.
11. C. Lemaréchal and C. Sagastizábal. Variable metric bundle methods: from conceptual to implementable forms. *Math. Program.*, 76(3):393–410, 1997.
12. A. Lewis and M. Overton. Nonsmooth optimization via quasi-Newton methods. *Math. Program.*, 141(1-2):135–163, 2013.
13. L. Lukšan and J. Vlček. Test problems for nonsmooth unconstrained and linearly constrained optimization. Technical Report 798, Institute of Computer Science, Academy of Sciences of the Czech Republic, 2000.
14. M. M. Mäkelä. Multiobjective proximal bundle method for nonconvex nonsmooth optimization: Fortran subroutine MPBNGC 2.0. Reports of the Department of Mathematical Information Technology, Series B, Scientific computing No. B 13/2003, University of Jyväskylä, 2003.
15. M.M. Mäkelä and P. Neittaanmäki. *Nonsmooth Optimization: Analysis and Algorithms with Applications to Optimal Control.* World Scientific Publishing Co., 1992.
16. S. Scholtes. *Introduction to Piecewise Differentiable Functions.* Springer, 2012.
17. N.Z. Shor. *Nondifferentiable optimization and polynomial problems.* Kluwer, 1998.