# An Exact Penalty Approach for General $\ell_0$-Sparse Optimization Problems

Christian Kanzow* Felix Weiß†

December 25, 2023

**Abstract**

We consider the general nonlinear optimization problem where the objective function has an additional term defined by the $\ell_0$-quasi-norm in order to promote sparsity of a solution. This problem is highly difficult due to its nonconvexity and discontinuity. We generalize some recent work and present a whole class of reformulations of this problem consisting of smooth nonlinear programs. This reformulated problem is shown to be equivalent to the original $\ell_0$-sparse optimization problem both in terms of local and global minima. The reformulation contains a complementarity constraint, and exploiting the particular structure of this reformulated problem, we introduce several problem-tailored constraint qualifications, first- and second-order optimality conditions and develop an exact penalty-type method which is shown to work extremely well on a whole bunch of different applications.

**Keywords.** Sparse optimization; global minima; local minima; strong stationarity; second-order conditions; approximate KKT conditions; exact penalty function
**Mathematical Science Classification.** 65K05 90C26 90C30 90C46

## 1 Introduction

In this paper, we consider the sparse optimization problem of the form

$$\min_x f(x) + \rho \|x\|_0, \quad x \in X, \tag{SPO}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$ is a smooth function, $X \subseteq \mathbb{R}^n$ a nonempty and closed set, $\rho > 0$ a given scalar, and

$$\|x\|_0 := \text{number of nonzero components } x_i \text{ of } x.$$

Note that $\|\cdot\|_0$ is not a norm, though it is often referred to as the $\ell_0$-*norm* in the literature. We call (SPO) also the *sparsest* optimization problem since we really want to solve this problem with the $\ell_0$-norm, and do not approximate this expression by some weaker version like in the standard approach, where the term $\|x\|_0$ gets approximated by $\|x\|_1$ or some other (nicer)

*University of Würzburg, Institute of Mathematics, Campus Hubland Nord, Emil-Fischer-Str. 30, 97074 Würzburg, Germany; kanzow@uni-wuerzburg.de

†University of Würzburg, Institute of Mathematics, Campus Hubland Nord, Emil-Fischer-Str. 30, 97074 Würzburg, Germany; felix.weiss@uni-wuerzburg.de

function. In the first part of the paper, we deal with an abstract feasible set $X$, whereas later, in the algorithmic part, we will assume that $X$ is described suitably by some equality and inequality constraints.

The solution of the sparsest optimization problem (SPO) is highly difficult due to the nonconvexity and discontinuity of the $\ell_0$-term in the objective function. According to [20], existing solution methods for (SPO) can be divided into the following categories: (a) convex approximations, (b) nonconvex approximations, and (c) nonconvex exact reformulations.

The convex approximation schemes typically replace the $\ell_0$-norm by the $\ell_1$-norm. This is the most standard approach which works very successfully in many applications. Furthermore, it has the major advantage that the resulting optimization problem is convex provided that the objective function $f$ and the feasible set $X$ are convex. The $\ell_1$-norm makes this problem nondifferentiable, nevertheless, there are plenty of methods which can be applied to this nonsmooth convex problem, see, e.g., the excellent monograph [4] for many examples of this kind.

The class of nonconvex approximation schemes usually replaces the $\ell_0$-term in (SPO) by a nonconvex penalty function. One possibility is to use the $\ell_p$-quasi-norm for $p \in (0, 1)$, see, e.g., [11, 12], which has nicer properties than the $\ell_0$-norm, e.g., it is continuous. However, despite its nonconvexity, it also fails to be Lipschitz continuous. There exist several other nonconvex penalty functions with the aim to approximate the $l_0$-norm in a suitable way and to keep some nicer smoothness assumptions like SCAD (= smoothly clipped absolute deviation) [14], MCP (= minimax concave penalty) [35], PiE (= piecewise exponential) [28] or the transformed $\ell_1$ approach [36]. The penalty decomposition algorithm from [22] is another approximation scheme for the solution of (SPO) and based on the quadratic (inexact) penalty function. Note that many of these techniques are investigated only for particular classes of problems covered by (SPO).

Here, we are particularly interested in the third class, the exact (nonconvex) reformulations of the sparsest optimization problem. There exist exact reformulations of (SPO) as mixed-integer programs, see [6] and the recent survey article [33] for further references. At least for convex quadratic programs with an additional $\ell_0$-term, this allows to compute a global minimum by suitable solvers if they get enough time. Another type of reformulation has been developed on the back of DC-approaches (DC = difference of convex) in [20], where DC-functions were used to approximate the $\ell_0$-norm, and in [16], where an exact reformulation of the $\ell_0$-norm was featured (though mainly in the context of cardinality-constrained problems, see below). The paper [15] considers an approach where the $\ell_0$-term is replaced by a suitable complementarity constraint. Subsequently, the latter was shown to be equivalent to (SPO) both in terms of local and global minima by the authors in [18].

The current work generalizes the recent contribution from [18] by introducing and investigating a whole class of reformulations of (SPO). The main idea presented here is somewhat related to a similar technique for cardinality-constrained optimization problems discussed in [10, 9] where the $\ell_0$-term in the objective function is replaced by a constraint of the form $\|x\|_0 \leq s$ for some given $s \in \mathbb{N}$. Note, however, that it is not possible to reformulate cardinality-constrained problems into a sparsest optimization problem, see [33] for a counterexample.

The class of reformulations presented here is also related to optimization problems called mathematical programs with equilibrium or switching constraints (MPEC and MPSC, for short). The stationarity conditions developed here are mostly in the same vein. Globalization approaches for these types of problems include, for instance, relaxation and (exact) penalty methods, cf. [17, 30, 21]. Due to the special (almost separable) structure the equililibrium (complementarity) or switching constraints arise in our class of reformulations, we are able to

prove relatively strong results which go far beyond those which are known for general MPECs or MPSCs.

The paper is organized as follows: Section 2 presents some background material from optimization and variational analysis. In Section 3, we introduce our class of reformulations of problem (SPO) and show that both the local and global minima coincide with the global and local minima of (SPO) (note that this is in contrast to the related reformulation of cardinality constraints discussed in [10, 9] where the reformulated problem might have additional local minima). We then introduce several problem-tailored constraint qualifications in Section 4 and present the resulting first- and second-order conditions for problem (SPO). An approximate stationarity concept will be discussed in Section 5. We then present our exact penalty method in 6 and provide several (strong) exactness and convergence results. In Section 7, we then investigate the numerical behaviour of our methods applied to a variety of different applications, which indicates that our method usually gets high-quality solutions, in many cases equal to the global minimum (for those problems where the global minimum is known or computable). We conclude with some final remarks in Section 8.

We close with some remarks on our notation in use: In the various parts of this paper, we address via

$$I_0(x) := \{i \,|\, x_i = 0\}$$

the set of indices for which $x$ vanishes. Furthermore, we write $x \circ y$ for the Hadamard-product of $x$ and $y$, i.e. the component wise multiplication of the two vectors. We abbreviate the canonical unit vector by $e_i \in \mathbb{R}^n$, indicating that the single 1 is in the $i$-th position, and additionally write $e := (1, 1, ..., 1)^T \in \mathbb{R}^n$. Since we will introduce sign constraints to our variables, we also denote with $\mathbb{R}^n_+$ the cone of vectors with only non-negative entries in $\mathbb{R}^n$.

## 2 Mathematical Background

This section provides some background from mathematical optimization and variational analysis, see, e.g., the monographs [5, 29] and [27, 32], respectively, for more details and corresponding proofs.

Consider the optimization problem

$$\min \; f(x) \qquad \text{s.t.} \qquad x \in X$$

with a continuously differentiable objective function $f\colon \mathbb{R}^n \to \mathbb{R}$ and a nonempty, closed set $X \subseteq \mathbb{R}^n$. For a feasible point $x \in X$, the *(Bouligand) tangent cone* or *contingent cone* of $x$ with respect to $X$ is defined by

$$\mathscr{T}_X(x) := \Big\{ d \in \mathbb{R}^n \,\Big|\, \exists \{x^k\} \to_X x, \exists \{t_k\} \downarrow 0 : d = \lim_{k\to\infty} \frac{x^k - x}{t_k} \Big\},$$

where the notation $x^k \to_X x$ indicates a sequence $\{x^k\}$ converging to $x$ such that $x^k \in X$ for all $k \in \mathbb{N}$. Furthermore, if the feasible set $X$ has a representation of the form

$$X = \big\{ x \in \mathbb{R}^n \,\big|\, g_i(x) \le 0 \; (i = 1, \dots, m), \; h_j(x) = 0 \; (j = 1, \dots, p) \big\}$$

for continuously differentiable functions $g_i, h_j : \mathbb{R}^n \to \mathbb{R}$, the corresponding *linearization cone* of $x \in X$ is defined by

$$\mathscr{L}_X(x) := \big\{ d \in \mathbb{R}^n \,\big|\, \nabla g_i(x)^T d \le 0 \; (i \in I_g(x)), \; \nabla h_j(x)^T d = 0 \; (j = 1, \dots, p) \big\},$$

where

$$I_g(x) := \big\{ i \in \{1,\dots,m\} \,\big|\, g_i(x) = 0 \big\}$$

denotes the set of active inequality constraints at the feasible point $x$. Note that the linearization cone depends on the particular representation of $X$, whereas the tangent cone is a purely geometric object, independent of any representation.

Given a nonempty cone $C \subseteq \mathbb{R}^n$, we denote by

$$C^\circ := \big\{ v \in \mathbb{R}^n \,\big|\, v^T d \le 0 \text{ for all } d \in C \big\}$$

the *polar cone* of $C$. We then say that the *Abadie constraint qualification* (Abadie CQ or ACQ for short) holds at $x \in X$ if

$$\mathscr{T}_X(x) = \mathscr{L}_X(x)$$

holds (the inclusion $\mathscr{T}_X(x) \subseteq \mathscr{L}_X(x)$ is automatically true, hence the opposite inclusion is the central requirement). Moreover, the *Guignard constraint qualification* (Guignard CQ or simply GCQ) is satisfied at $x \in X$ if the corresponding polar cones coincide, i.e., if

$$\mathscr{T}_X(x)^\circ = \mathscr{L}_X(x)^\circ$$

holds. Note that ACQ implies GCQ, whereas the converse is not true in general.

For a nonempty and closed set $X \subseteq \mathbb{R}^n$, we call

$$\hat{\mathscr{N}}_X(\bar{x}) := \mathscr{T}_X(\bar{x})^\circ$$

the *Fréchet normal cone* of $\bar{x} \in X$. Furthermore,

$$\begin{aligned} \mathscr{N}_X(\bar{x}) &:= \operatorname{Limsup}_{x \to_X \bar{x}} \hat{\mathscr{N}}_X(x) \\ &:= \big\{ v \in \mathbb{R}^n \,\big|\, \exists \{x^k\}, \exists \{v^k\} : x^k \to_X \bar{x},\ v^k \to v,\ v^k \in \hat{\mathscr{N}}_X(x^k)\ \forall k \in \mathbb{N} \big\} \end{aligned}$$

denotes the *Mordukhovich normal cone* or *limiting normal cone* of $\bar{x} \in X$. For the sake of completeness, we set $\hat{\mathscr{N}}_X(\bar{x}) := \mathscr{N}_X(\bar{x}) := \emptyset$ for each point $\bar{x} \notin X$. Note that we always have the inclusion $\hat{\mathscr{N}}_X(\bar{x}) \subseteq \mathscr{N}_X(\bar{x})$, whereas for convex sets $X$, both normal cones coincide and are equal to the usual normal cone from convex analysis, i.e., the equalities

$$\hat{\mathscr{N}}_X(\bar{x}) = \mathscr{N}_X(\bar{x}) = \mathscr{N}_X^{conv}(\bar{x}) := \big\{ v \in \mathbb{R}^n \,\big|\, v^T(x - \bar{x}) \le 0\ \forall x \in X \big\}$$

hold for convex $X$.

Next, let us write $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ for the extended real line (excluding the value $-\infty$). For $\varphi\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ being proper, we call

$$\operatorname{epi}(\varphi) := \big\{ (x,\alpha) \in \mathbb{R}^n \times \mathbb{R} \,\big|\, \varphi(x) \le \alpha \big\}$$

the *epigraph* of $\varphi$. Based on the previously introduced normal cones, we can define two corresponding subdifferentials for the nonsmooth function $\varphi$, namely the *Fréchet subdifferential* of $\bar{x} \in \operatorname{dom}(\varphi) := \{x \in \mathbb{R}^n \,|\, \varphi(x) < \infty\}$, given by

$$\hat{\partial}\varphi(\bar{x}) := \big\{ s \in \mathbb{R}^n \,\big|\, (s,-1) \in \hat{\mathscr{N}}_{\operatorname{epi}(\varphi)}\big(\bar{x}, \varphi(\bar{x})\big) \big\}$$

and the *Limiting or Mordukhovich subdifferential*

$$\partial\varphi(\bar{x}) := \big\{ s \in \mathbb{R}^n \,\big|\, (s,-1) \in \mathscr{N}_{\operatorname{epi}(\varphi)}\big(\bar{x}, \varphi(\bar{x})\big) \big\}.$$

From the corresponding relation between the normal cones, we get the inclusion $\hat{\partial}\varphi(\bar{x}) \subseteq \partial\varphi(\bar{x})$, whereas both subdifferentials coincide and are equal to the standard subdifferential from convex analysis, i.e.,

$$\hat{\partial}\varphi(\bar{x}) = \partial\varphi(\bar{x}) = \partial^{conv}\varphi(\bar{x}) := \left\{ s \in \mathbb{R}^n \,\middle|\, \varphi(x) \geq \varphi(\bar{x}) + s^T(x - \bar{x}) \;\forall x \in \mathbb{R}^n \right\}$$

for $\varphi$ being a convex function.

Using these subdifferentials, we introduce the following notion.

**Definition 2.1.** *Consider the optimization problem*

$$\min \;\; \psi(x) \qquad s.t. \qquad x \in X \tag{1}$$

*for some proper and lower semicontinuous function $\psi\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ and a nonempty, closed set $X \subseteq \mathbb{R}^n$. We then call $\bar{x} \in X$*

*(a) an* M-stationary point *(Mordukhovich stationary point) of (1) if $0 \in \partial\psi(\bar{x}) + \mathcal{N}_X(\bar{x})$;*

*(b) an* S-stationary point *(strongly stationary point) of (1) if $0 \in \hat{\partial}\psi(\bar{x}) + \hat{\mathcal{N}}_X(\bar{x})$.*

Since the Fréchet normal cone is (in general) smaller than the limiting normal cone, S-stationarity is a stronger stationary concept than M-stationarity.

We next restate a sum rule for the above two subdifferentials, see [27, Prop. 1.30] for a proof.

**Theorem 2.2.** *Let $\psi := f + \varphi$ with $\varphi\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ proper and lower semicontinuous, $f\colon \mathbb{R}^n \to \mathbb{R}$, and $\bar{x} \in dom(\varphi)$ be given. Then the following statements hold:*

*(a) If $f$ is differentiable in $\bar{x}$, then $\hat{\partial}\psi(\bar{x}) = \nabla f(\bar{x}) + \hat{\partial}\varphi(\bar{x})$ holds.*

*(b) If $f$ is continuously differentiable in a neighbourhood of $\bar{x}$, then $\partial\psi(\bar{x}) = \nabla f(\bar{x}) + \partial\varphi(\bar{x})$ holds.*

Now, consider the (constrained composite) optimization problem

$$\min \;\; f(x) + \varphi(x) \qquad \text{s.t.} \qquad x \in X \tag{2}$$

with $f\colon \mathbb{R}^n \to \mathbb{R}$ continuously differentiable, $\varphi\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ proper and lower semicontinuous, and $X \subseteq \mathbb{R}^n$ nonempty and closed. Writing $\psi := f + \varphi$ for the objective function, we obtain from Theorem 2.2 that

$$\hat{\partial}\psi(\bar{x}) = \nabla f(\bar{x}) + \hat{\partial}\varphi(\bar{x}) \quad \text{and} \quad \partial\psi(\bar{x}) = \nabla f(\bar{x}) + \partial\varphi(\bar{x}).$$

Consequently, using the notion from Definition 2.1, we see that a feasible point $\bar{x} \in X$ is M-stationary for (2) if

$$0 \in \nabla f(\bar{x}) + \partial\varphi(\bar{x}) + \mathcal{N}_X(\bar{x})$$

holds, whereas it is S-stationary for (2) if we have

$$0 \in \nabla f(\bar{x}) + \hat{\partial}\varphi(\bar{x}) + \hat{\mathcal{N}}_X(\bar{x}).$$

Note that $\varphi(x) := \rho \|x\|_0$ is a proper and lower semicontinuous function, hence our sparse optimization problem (SPO) is a special instance of the formulation (2). The previous M- and S-stationary conditions then require the corresponding subdifferentials of this function. The answer is given in the following result, which follows from [19, 13] and lower semicontinuity of the $\ell_0$-norm.

**Lemma 2.3.** *Consider the function* $\varphi(x) := \rho \|x\|_0$ *for some* $\rho > 0$. *Then*

$$\hat{\partial}\varphi(x) = \partial\varphi(x) = \left\{ s \in \mathbb{R}^n \,\middle|\, s_i = 0 \text{ for all } i \text{ with } x_i \neq 0 \right\}$$

*for all* $x \in \mathbb{R}^n$.

Hence the limiting and Fréchet subdifferentials of $\varphi(x) = \rho \|x\|_0$ coincide and are independent of the particular value of the parameter $\rho > 0$. In order to apply the above M- and S-stationarity conditions of problem (2) to our setting from (SPO), it remains to compute the corresponding normal cones $\mathcal{N}_X(\bar{x})$ and $\hat{\mathcal{N}}_X(\bar{x})$. This will be done by using some (problem-tailored) constraint qualifications, see Section 4.

# 3 Reformulation of Sparse Optimization Problem

In the previous work [18], we established an equivalence regarding local and (up to a scaling of $\rho$) global minima between problem (SPO) and the following reformulation of (SPO) based on an auxiliary variable $y$:

$$\min_{x,y} f(x) + \frac{\rho}{2} y^T (y - 2e) \qquad \text{s.t.} \qquad x \in X,\ x \circ y = 0, \tag{3}$$

where $e = (1, 1, ..., 1)^T \in \mathbb{R}^n$. The aim of this section is to generalize this approach.

To this end, we introduce a penalty function $p^\rho \colon \mathbb{R}^n \to \mathbb{R}$ (usually depending on the parameter $\rho > 0$) given by

$$p^\rho(y) = \sum_{i=1}^{n} p_i^\rho(y_i) \tag{4}$$

with each $p_i^\rho \colon \mathbb{R} \to \mathbb{R}$ being such that it satisfies the following conditions:

(P.1) $p_i^\rho$ is convex and attains a unique minimum (possibly depending on $\rho$) at some point $s_i^\rho > 0$;

(P.2) $p_i^\rho(0) - p_i^\rho(s_i^\rho) = \rho$;

(P.3) $p_i^\rho$ is sufficiently smooth.

Assumption (P.1) simply states that $p_i^\rho$ is a convex function which attains its unique minimum in the open interval $(0, \infty)$. We denote this minimum by $s_i^\rho > 0$. Furthermore, we write

$$m_i^\rho := p_i^\rho(s_i^\rho) \quad \text{and} \quad M^\rho := \sum_{i=1}^{n} m_i^\rho \tag{5}$$

for the corresponding minimal function values of $p_i^\rho$ and $p^\rho$, respectively. Condition (P.2) is a scaling assumption that can always be guaranteed by multiplication of $p_i^\rho$ with a suitable factor, whereas condition (P.3) is a smoothness condition, with the degree of smoothness depending on the particular situation which should be clear from the corresponding context. In particular, for the reformulation of the sparse optimization problem (SPO) within this section, it will be enough to have each $p_i^\rho$ continuous (which is automatically satisfied by the convexity assumption). The subsequent discussion of suitable constraint qualifications and stationarity concepts requires each $p_i^\rho$ to be continuously differentiable, whereas in the second-order theory, $p_i^\rho$ needs to be twice continuously differentiable.

In the following, we provide some examples of suitable functions $p_i^\rho$.

**Example 3.1.** The following functions $p_i^\rho\colon \mathbb{R} \to \mathbb{R}$ satisfy conditions (P.1)–(P.3):

(a) The function $p_i^\rho(y_i) := \rho y_i(y_i - 2)$ is convex (in fact, uniformly convex), satisfies all smoothness requirements, and attains a unique minimum at $s_i^\rho := 1$ (which, in this case, is independent of $\rho$).

(b) The function $p_i^\rho(y_i) = \frac{1}{2}(y_i - \sqrt{2\rho})^2$ also satifies all of the above requirements and can be seen as a somewhat natural choice, since we want $y_i^*$ to attain some positive value $s_i^\rho$ for $x_i^*$ to vanish. This particular choice simply penalizes the deviation in $y_i^*$ from $s_i^\rho = \sqrt{2\rho}$, where $s_i^\rho$ was selected in accordance to (P.2).

(c) The shifted absolute-value function $p_i^\rho(y_i) := \rho|y_i - 1|$ also satisfies (P.1)–(P.3), though (P.3) only holds with continuity, which is sufficient for the considerations within this section. Using a Huber-type smoothing (together with a suitable scaling so that (P.2) holds), we can easily construct a continuously differentiable version of this function satisfying (P.1)–(P.3).

It is clear that several other examples satisfying (P.1)–(P.3) can be constructed easily. In the following, we assume that $p^\rho$ is given by (4) with each term $p_i^\rho$ satisfying conditions (P.1)–(P.3), where only continuity is required in (P.3) within this section. We then consider the reformulation

$$\min_{x,y} f(x) + p^\rho(y) \qquad \text{s.t.} \qquad x \in X,\ x \circ y = 0 \tag{SPOref}$$

of the sparse optimization problem (SPO) (the acronym "SPOref" stands for "SPO-reformulation"). Note that, for the choice of $p_i^\rho$ as in Example 3.1 (a), we reobtain the previous formulation from (3) (except for the factor $\frac{1}{2}$ which would destroy property (P.2)).

The aim of this section is to show that problem (SPOref) is indeed a reformulation of the sparse optimization problem in the sense that it has the same local and global minima. For this purpose, we begin with the following preliminary observation.

**Lemma 3.2.** *Let $p^\rho$ be given by (4) with each $p_i^\rho$ satisfying properties (P.1)–(P.3), and let $M^\rho$ be defined by (5). Then the following statements hold:*

*(a) The inequality $\rho \|x\|_0 \le p^\rho(y) - M^\rho$ holds for any feasible point $(x, y)$ of (SPOref).*

*(b) Equality $\rho \|x\|_0 = p^\rho(y) - M^\rho$ holds for a feasible point $(x, y)$ of (SPOref) if and only if $y_i = s_i^\rho$ for all $i \in I_0(x)$.*

*(c) If $(x^*, y^*)$ is a local minimum of (SPOref), we have $y_i^* = s_i^\rho$ for all $i \in I_0(x^*)$.*

*Proof.* (a) The claim follows from

$$\begin{aligned}
\rho \|x\|_0 &= \sum_{i \notin I_0(x)} \rho = \sum_{i \notin I_0(x)} \big(p_i^\rho(0) - p_i^\rho(s_i^\rho)\big) = \sum_{i \notin I_0(x)} \big(p_i^\rho(y_i) - p_i^\rho(s_i^\rho)\big) \\
&\le \sum_{i \notin I_0(x)} \big(p_i^\rho(y_i) - p_i^\rho(s_i^\rho)\big) + \sum_{i \in I_0(x)} \big(p_i^\rho(y_i) - p_i^\rho(s_i^\rho)\big) \\
&= \sum_{i=1}^n \big(p_i^\rho(y_i) - p_i^\rho(s_i^\rho)\big) = p^\rho(y) - M^\rho,
\end{aligned}$$

where the first identity results from the definition of $\|x\|_0$ together with the one of the index set $I_0(x)$, the second equation exploits the scaling property (P.2), the third equation comes from the fact that we necessarily have $y_i = 0$ for all $i \notin I_0(x)$ due to the constraints $x \circ y = 0$, the inequality takes into account that $p_i^\rho(y_i) - p_i^\rho(s_i^\rho) \geq 0$ due to the minimality of $s_i^\rho$, and the remaining part is simply the definition of $p^\rho$ and $M^\rho$.

(b) Observe that the previous chain of equations and inequalities holds with equation if and only if $p_i^\rho(y_i) = p_i^\rho(s_i^\rho)$ for all $i \in I_0(x)$. Since the minimum $s_i^\rho$ is unique by condition (P.1), this holds if and only if $y_i = s_i^\rho$ for all $i \in I_0(x)$, hence statement (b) holds.

(c) This statement follows from the observation that, for any $i \in I_0(x^*)$, the auxiliary variable $y_i^*$ has to solve the problem

$$\min_{y_i} \; p_i^\rho(y_i)$$

(note that $y_i = 0$ is fixed for each $i \notin I_0(x^*)$ due to the complementarity-type constraint $x \circ y = 0$, and that the objective function is separable in each $y_i$). □

Note that Lemma 3.2 together with the constraint $x \circ y = 0$ implies that, for $(x^*, y^*)$ being a local minimum of (SPOref), we necessarily have

$$y_i^* = \begin{cases} s_i^\rho, & \text{for } i \in I_0(x^*), \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

In particular, $y^*$ is uniquely defined by $x^*$. Exploiting this observation, we are able to formulate an equivalence result between the local minima of the two problems (SPO) and (SPOref).

**Theorem 3.3.** *A feasible $x^*$ for (SPO) is a local minimum of (SPO) if and only if $(x^*, y^*)$ with $y^*$ given by (6) is a local minimum of (SPOref).*

*Proof.* The proof is very similar to the corresponding one in our previous work [18], and is presented here for the sake of completeness and because part of the proof will be used also in the proof of the subsequent result

Let $x^*$ be local minimum of (SPO), and let $y^*$ be defined by (6). We then obtain

$$f(x^*) + p^\rho(y^*) = f(x^*) + \rho \|x^*\|_0 + M^\rho \leq f(x) + \rho \|x\|_0 + M^\rho \leq f(x) + p^\rho(y) \tag{7}$$

for all feasible $(x, y)$ with $x$ sufficiently close to $x^*$, where the first equality results from Lemma 3.2 (b), the subsequent inequality exploits the local minimality of $x^*$ for the program (SPO), and the final inequality follows from Lemma 3.2 (a).

Conversely, let $(x^*, y^*)$ be a local minimum of (SPOref). Recall that $y^*$ is then given by (6), cf. Lemma 3.2 (c). Assume that $x^*$ is not a local minimum of (SPO). Then there exists a sequence $\{x^k\} \subseteq X$ such that $x^k \to x^*$ and

$$f(x^k) + \rho \left\|x^k\right\|_0 < f(x^*) + \rho \|x^*\|_0 \quad \forall k \in \mathbb{N}. \tag{8}$$

Note that $\left\|x^k\right\|_0 \geq \|x^*\|_0$ holds for all $k$ sufficiently large. Hence we either have a subsequence $\{x^k\}_K$ such that $\left\|x^k\right\|_0 = \|x^*\|_0$ for all $k \in K$, or $\|x^*\|_0 + 1 \leq \left\|x^k\right\|_0$ is true for almost all

$k \in \mathbb{N}$. In the former case, it follows that $(x^k, y^*)$ is feasible for (SPOref), hence we obtain from Lemma 3.2 (b) and the minimality of $(x^*, y^*)$ for (SPOref) that

$$\begin{aligned} f(x^k) + \rho \left\|x^k\right\|_0 + M^\rho &= f(x^k) + \rho \left\|x^*\right\|_0 + M^\rho \\ &= f(x^k) + p^\rho(y^*) \\ &\geq f(x^*) + p^\rho(y^*) \\ &= f(x^*) + \rho \left\|x^*\right\|_0 + M^\rho, \end{aligned}$$

which contradicts (8). Otherwise, we have $\left\|x^*\right\|_0 + 1 \leq \left\|x^k\right\|_0$ and, by continuity, also $f(x^*) \leq f(x^k) + \rho$ for all $k \in \mathbb{N}$ sufficiently large, which, in turn, gives

$$f(x^k) + \rho \left\|x^k\right\|_0 \geq f(x^k) + \rho + \rho \left\|x^*\right\|_0 \geq f(x^*) + \rho \left\|x^*\right\|_0 .$$

Hence, also in this situation, we have a contradiction to (8). □

Note that the full equivalence of the set of local minima is quite interesting, especially since a similar result does not hold for a somewhat related reformulation of optimization problems with cardinality constraints, see [9].

The next result states the equivalence between global minima of the two problems (SPO) and (SPOref). This result, however, is less surprising than the previous one regarding local minima (and holds, in particular, also for the previously mentioned cardinality-constrained problems discussed in [9]).

**Theorem 3.4.** *A feasible $x^*$ for (SPO) is a global minimum of (SPO) if and only if $(x^*, y^*)$ with $y^*$ given by (6) is a global minimum of (SPOref).*

*Proof.* First assume that $x^*$ is a global minimum of (SPO). Then the chain of inequalities from (7) holds for all $(x, y)$ feasible for (SPOref), showing that $(x^*, y^*)$ is a global minimum of (SPOref).

Conversely, let $(x^*, y^*)$ solve problem (SPOref) globally, and let $x \in X$ be an arbitrary feasible point of (SPO). We then define a vector $y \in \mathbb{R}^n$ similar to (6) so that $(x, y)$ is feasible for (SPOref). Using the optimality of $(x^*, y^*)$ and exploiting Lemma 3.2 (b) twice, we then obtain

$$f(x^*) + \rho \left\|x^*\right\|_0 + M^\rho = f(x^*) + p^\rho(y^*) \leq f(x) + p^\rho(y) = f(x) + \rho \left\|x\right\|_0 + M^\rho.$$

Subtracting the constant $M^\rho$ from both sides shows that $x^*$ is a global minimum of problem (SPO). □

Altogether, the results from this section show that we can reformulate the nonsmooth and even discontinuous sparse optimization problem (SPO) as a continuous problem without any loss of information regarding local or global minima. The main difficulty of the reformulation (SPOref) is, of course, the constraint $x \circ y = 0$ which is not easy to deal with, in particular, it violates most of the standard constraint qualifications like the Abadie CQ and, therefore, all constraint qualifications which are stronger than Abadie. We are going to deal with this difficulty in our subsequent sections by introducing some problem-tailored CQs and by considering a particular method for the solution of problem (SPOref) which exploits the structure of the underlying problem.

# 4 Constraint Qualifications and Optimality Conditions

The aim of this section is to introduce some suitable (problem-tailored) constraint qualifcations which are then used to obtain corresponding optimality conditions. Note that this requires that the feasible set $X$ has an explicit representation by equality and/or inequality constraints. In view of our subsequent algorithmic approach for solving the reformulation (SPOref) of the sparse optimization problem (SPO), we assume from now on that the feasible set is given by

$$X := \{x \in \mathbb{R}^n \mid g(x) \leq 0,\ h(x) = 0,\ x \geq 0\} \tag{9}$$

with some smooth functions $g\colon \mathbb{R}^n \to \mathbb{R}^m$ and $h\colon \mathbb{R}^n \to \mathbb{R}^p$. Hence, in addition to some standard constraints, we assume explicitly that the inequalities contain nonnegativity constraints, and that these nonnegativity constraints are separated from the remaing inequalities $g(x) \leq 0$. This representation turns out to be very useful in our subsequent exact penalty approach.

In order to derive some problem-tailored constraint qualifications, we follow a standard approach that is also used in the context of mathematical programs with equilibrium constraints, see [23], or optimization problems with switching constraints, see [17, 26]. To this end, let $x^*$ be a local minimum of problem (SPO) with $X$ defined by (9). This implies that $x^*$ is also a local minimum of the *tightend nonlinear program*

$$\min_x f(x) \qquad \text{s.t.} \qquad g(x) \leq 0,\ h(x) = 0,\ x_i = 0,\ i \in I_0(x^*) \tag{TNLP($x^*$)}$$

since, locally, the feasible set of (TNLP($x^*$)) is a subset of $X$ and $x^*$, by definition, is still feasible for this problem. Note, however, that (TNLP($x^*$)) depends on $x^*$ (via the index set $I_0(x^*)$) and can therefore be used only as a theoretical tool. Here, we exploit this observation to formulate the subsequent constraint qualifications.

**Definition 4.1.** *Let $x^* \in X$ with $X$ being defined by (9). We say that $x^*$ satisfies*

*(a)* SP-LICQ *(sparse LICQ) if the gradients*

$$\nabla g_i(x^*),\ i \in I_g(x^*), \quad \nabla h_j(x^*),\ j = 1, ..., p, \quad e_i,\ i \in I_0(x^*)$$

*are linearly independent.*

*(b)* SP-MFCQ *(sparse MFCQ) if the gradients*

$$\nabla g_i(x^*),\ i \in I_g(x^*), \quad \nabla h_j(x^*),\ j = 1, ..., m, \quad e_i,\ i \in I_0(x^*),$$

*are positively linearly independent.*

*(c)* SP-RCPLD *(sparse RCPLD) if there is a neighborhood $U$ of $x^*$ such that, for any index sets $I_1 \subseteq \{1, ..., p\}$, $I_2 \subseteq I_0(x^*)$ with the gradients $\{\nabla h_i(x^*)\}_{I_1}$, $\{e_i\}_{i \in I_2}$ forming a basis of the subspace generated by all gradients $\nabla h_i(x^*)$ $(i = 1, \ldots, p)$, $e_i$ $(i \in I_0(x^*))$, the following holds for all $x \in U$.*

1. *$\left(\{\nabla h_i(x)\}_{i=1}^p,\ \{e_i\}_{i \in I_0(x^*)}\right)$ is of constant rank for all $x \in U$*
2. *For every $J \subseteq I_g(x^*)$, if $(\{\nabla h_i(x^*)\}_{I_1}, \{e_i\}_{I_2}, \{\nabla g_i(x^*)\}_J)$ is positive-linearly dependent, then $(\{\nabla h_i(x)\}_{I_1}, \{e_i\}_{I_2}, \{\nabla g_i(x)\}_J)$ is linearly dependent for every $x \in U$.*

Observe that SP-LICQ, SP-MFCQ and SP-RCPLD correspond to standard LICQ, standard MFCQ and standard RCPLD for the tightened problem (TNLP($x^*$)) (with RCPLD being a constraint qualification introduced in [1]). Hence, it is clear how to formulate further constraint qualifications based on this relation. In particular, standard theory on constraint qualifications therefore show that each of the problem-tailored constraint qualifications from Definition 4.1 imply that (standard) Abadie CQ holds for the tightened problem (TNLP($x^*$)). We now use this observation in order to derive a suitable optimality condition for the sparse optimization problem (SPO).

To this end, let $x^*$ be a local minimum of (SPO), and consider the corresponding tightened problem (TNLP($x^*$)). Assume further that any of the problem-tailored constraint qualifications from Definition 4.1 hold. Let us denote by

$$X_{\text{TNLP}(x^*)} \text{ the feasible set of } (\text{TNLP}(x^*)),$$

and by

$$\mathcal{T}_{X_{\text{TNLP}(x^*)}}(x^*), \quad \mathcal{L}_{X_{\text{TNLP}(x^*)}}(x^*), \quad \hat{\mathcal{N}}_{X_{\text{TNLP}(x^*)}}(x^*), \quad \text{and} \quad \mathcal{N}_{X_{\text{TNLP}(x^*)}}(x^*)$$

the corresponding tangent cone, linearization cone, Fréchet normal cone, and limiting normal cone of $x^* \in X_{\text{TNLP}(x^*)}$. Since each of the SP-CQs from Definition 4.1 implies that the standard Abadie and, hence, the standard Guignard CQ holds at $x^* \in X_{\text{TNLP}(x^*)}$, we obtain

$$\mathcal{T}_{X_{\text{TNLP}(x^*)}}(x^*)^\circ = \mathcal{L}_{X_{\text{TNLP}(x^*)}}(x^*)^\circ.$$

Now, the linearization cone $\mathcal{L}_{X_{\text{TNLP}(x^*)}}(x^*)$ is a polyhedral convex cone, and standard results from convex analysis imply that its polar is given by

$$\mathcal{L}_{X_{\text{TNLP}(x^*)}}(x^*)^\circ = \Big\{ d \,\Big|\, d = \sum_{i \in I_g(x^*)} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{p} \mu_j \nabla h_j(x^*) + \sum_{i \in I_0(x^*)} \gamma_i e_i,\ \lambda_i \geq 0 \ (i \in I_0(x^*)) \Big\}.$$

Since, by definition, we have

$$\hat{\mathcal{N}}_{X_{\text{TNLP}(x^*)}}(x^*) = \mathcal{T}_{X_{\text{TNLP}(x^*)}}(x^*)^\circ = \mathcal{L}_{X_{\text{TNLP}(x^*)}}(x^*)^\circ,$$

this shows that

$$\hat{\mathcal{N}}_{X_{\text{TNLP}(x^*)}}(x^*) = \Big\{ d \,\Big|\, d = \sum_{i \in I_g(x^*)} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{p} \mu_j \nabla h_j(x^*) + \sum_{i \in I_0(x^*)} \gamma_i e_i,\ \lambda_i \geq 0 \ (i \in I_0(x^*)) \Big\}.$$

Now, using the expression of the Fréchet normal cone for the tightened nonlinear program (TNLP($x^*$)), the notation of an S-stationary point from Section 2 for a general optimization problem as in (2), and taking into account the formula for the Fréchet subdifferential of the function $\varphi(x) := \rho \|x\|_0$ from Lemma 2.3, it follows that the local minimum $x^*$ satisfies the following S-stationary conditions under any of the SP-CQs from Definition 4.1, where, for simplicity of notation,

$$L^{SP}(x, \lambda, \mu) := f(x) + \lambda^T g(x) + \mu^T h(x) \tag{10}$$

denotes a mapping that we call the *SP-Lagrangian* of problem (SPO) with the feasible set $X$ given by (9) (note that this SP-Lagrangian neither includes the $\ell_0$-term of the original objective function nor any term resulting from the nonnegativity constraints).

**Definition 4.2.** *Let $x^* \in X$ be feasible for the sparse optimization problem (SPO), where $X$ is given by (9). Then we call $x^*$ an* S-stationary point *of (SPO) if there exist multipliers $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ such that*

$$\begin{aligned} \nabla_{x_i} L^{SP}(x^*, \lambda^*, \mu^*) = 0, \quad i \notin I_0(x^*), \\ h(x^*) = 0, \\ \lambda^* \geq 0,\ g(x^*) \leq 0,\ \lambda^* \circ g(x^*) = 0. \end{aligned}$$

Note that the previous derivation shows that S-stationarity is a necessary optimality condition for a local minimum $x^*$ of problem (SPO) provided that a suitable (problem-tailored) constraint qualification holds. In a similar way, one can also derive an M-stationary condition. However, in this particular case, there is no difference between M- and S-stationarity due to the fact that the Fréchet and limiting subdifferentials of the function $\varphi(x) = \rho \|x\|_0$ coincide, cf. Lemma 2.3.

Before providing another interpretation of S-stationary points, we consider a slightly different reformulation of problem (SPO) with $X$ given by (9). Since, by (6), any local minimum $(x^*, y^*)$ of the reformulated problem (SPOref) automatically satisfies $y^* \geq 0$, it follows that (SPOref) and, hence, (SPO) itself is totally equivalent to the program

$$\min_{x,y} f(x) + p^\rho(y) \qquad \text{s.t.} \qquad g(x) \leq 0,\ h(x) = 0,\ x \circ y = 0,\ x \geq 0,\ y \geq 0 \qquad \text{(SPOcp)}$$

in terms of local and global minima. We call (SPOcp) the complementarity reformulation of problem (SPO) (hence the acronym "cp") due to the complementarity constraints $x \geq 0, y \geq 0, x^T y = 0$. Note that (SPOcp) will be the basis of our algorithmic approach for the solution of the sparse optimization problem (SPO).

Assuming that the function $p^\rho$ is continuously differentiable, the two reformulations (SPOref) and (SPOcp) are smooth optimization problems. Hence, we can write down the corresponding KKT conditions. It turns out they are equivalent to the S-stationarity conditions from Definition 4.2. This is summarized in the following result.

**Theorem 4.3.** *Consider the sparse optimization problem (SPO) with feasible set $X$ given by (9). Furthermore, for a feasible point $x^* \in X$, let $y^*$ denote the corresponding vector defined by (6). Then the following statements are equivalent:*

*(a) $x^*$ is an S-stationary point of (SPO).*

*(b) The KKT conditions of (SPOref) are satisfied at $(x^*, y^*)$.*

*(c) The KKT conditions of (SPOcp) are satisfied at $(x^*, y^*)$.*

We skip the proof of Theorem 4.3 since it is rather elementary. We only stress the following two facts: Definition (6) of $y^*$ implies that the bi-active set of the solution pair $(x^*, y^*)$ is empty, i.e., there is no index $i$ with $(x_i^*, y_i^*) = (0, 0)$. Furthermore, for $x_i^* = 0$, we have that $y_i^*$ is equal to the unique minimum $s_i^\rho$ of the function $p_i^\rho$, hence $[\nabla p^\rho(y^*)]_i = 0$ follows.

The following observation is simple, but quite interesting and important for our subsequent theory, hence we state it explicitly in the following remark.

**Remark 4.4.** Let $(x^*, y^*)$ be a stationary point of (SPOcp). We then claim that the bi-active set

$$\mathscr{B}(x^*, y^*) := \{i \mid (x_i^*, y_i^*) = (0, 0)\}$$

is automatically empty. In fact, from the stationarity conditions of (SPOcp), we, in particular, obtain

$$\nabla p_i^\rho(y_i^*) + \eta x_i^* - \nu_i^x = 0 \quad \forall i = 1, \ldots, m$$

for some corresponding multipliers $\eta \in \mathbb{R}$ and $\nu^x \geq 0$. If there were an index $i$ with $(x_i^*, y_i^*) = (0, 0)$, the term in the middle vanishes, and the first term is strictly negative by the convexity of $p_i^\rho$ together with the assumption that this function attains a unique minimum at the positive number $s_i^\rho$. Hence, it follows that $\nabla p_i^\rho(y_i^*) + \eta x_i^* - \nu_i^x < 0$, and this contradiction completes the proof.

Theorem 4.3 shows that the given sparse optimization problem (SPO) and its two reformulations (SPOref) and (SPOcp) are not only equivalent with respect to local and global minima, but also in terms of their first-order optimality conditions.

We next demonstrate that also the corresponding second-order conditions coincide. To this end, we assume that all functions are twice continuously differentiable. Furthermore, let $L^{SP}$ be the SP-Lagrangian of (SPO). We then define the *SP-critical cone*

$$\begin{aligned} \mathscr{C}^{SP}(x^*, \lambda^*) = \big\{ d \,\big|\, & \nabla h_j(x^*)^T d = 0, \ j = 1, \ldots, p, \\ & \nabla g_i(x^*)^T d = 0, \ i \in I_g(x^*), \ \lambda_i^* > 0, \\ & \nabla g_i(x^*)^T d \leq 0, \ i \in I_g(x^*), \ \lambda_i^* = 0, \\ & d_i = 0, \ i \in I_0(x^*) \big\} \end{aligned}$$

at some S-stationary point $x^*$ with corresponding multipliers $(\lambda^*, \mu^*)$.

**Definition 4.5.** *Given an S-stationary point $x^*$ with multipliers $(\lambda^*, \mu^*)$, and using the notion of the SP-Lagrangian and the SP-critical cone as before, we say that the triple $(x^*, \lambda^*, \mu^*)$ satisfies the*

*(a)* SP-SOSC *(sparse second-order sufficiency condition) if*

$$d^T \nabla_{xx}^2 L^{SP}(x^*, \lambda^*, \mu^*) d > 0 \quad \forall d \in \mathscr{C}^{SP}(x^*, \lambda^*) \setminus \{0\}$$

*(b)* SP-SONC *(sparse second-order necessary condition) if*

$$d^T \nabla_{xx}^2 L^{SP}(x^*, \lambda^*, \mu^*) d \geq 0 \quad \forall d \in \mathscr{C}^{SP}(x^*, \lambda^*)$$

*holds.*

These second-order conditions turn out to be equivalent to the standard second-order conditions of the two reformulations (SPOref) and (SPOcp). This observation is summarized in the following result.

**Theorem 4.6.** *Let $x^*$ be an S-stationary point with multipliers $(\lambda^*, \mu^*)$, and let $y^*$ be given by (6). Assume that $p^\rho$ is twice continuously differentiable. Then the following statements are equivalent.*

*(a) SP-SONC holds at $x^*$ for (SPO).*

*(b) SONC holds at $(x^*, y^*)$ for (SPOref).*

*(c) SONC holds at $(x^*, y^*)$ for (SPOcp).*

*The same equivalences also hold for SP-SOSC and SOSC provided that the (diagonal) matrix $\nabla p^\rho(y^*)$ is positive definite.*

Note that, due to the separable structure of the function $p^\rho$, the Jacobian $\nabla p^\rho(y^*)$ is indeed a diagonal matrix. Due to the assumed convexity of each function $p_i^\rho$, this diagonal matrix is automatically positive semidefinite. For the equivalence of the second-order sufficiency conditions, however, we require the slightly stronger assumption that this matrix is positive definite. Note that this holds automatically if $p_i^\rho$ is strongly convex around $y^*$, an assumption that is satisfied for all instances from Example 3.1.

We skip the details of the proof, but provide at least some hints. The equivalence of statements (b) and (c) is based on the following two facts: (i) the Hessian of the corresponding two Lagrangians coincide since the Lagrangian of (SPOcp) contains only one more linear term which disappears in the second-order derivatives, (ii) the criticial cones of the two sets $\{(x, y) \,|\, x \geq 0, x \circ y = 0\}$ and $\{(x, y) \,|\, x \geq 0, y \geq 0, x \circ y = 0\}$ coincide at $(x, y) = (x^*, y^*)$ since the biactive set is empty, cf. Remark 4.4 (otherwise, this statement would be wrong).

Furthermore, a simple calculation shows that the Hessian of the Lagrangian of either (SPOref) or (SPOcp) is given by

$$H = \begin{pmatrix} \nabla^2_{xx} L^{SP}(x^*, \lambda^*, \mu^*) & \operatorname{diag}(\gamma) \\ \operatorname{diag}(\gamma) & \nabla^2_{yy} p^\rho(y^*) \end{pmatrix}$$

for some multiplier $\gamma$. This implies

$$\begin{pmatrix} d_x \\ d_y \end{pmatrix}^T H \begin{pmatrix} d_x \\ d_y \end{pmatrix} = d_x^T \nabla^2_{xx} L^{SP}(x^*, \lambda^*, \mu^*) d_x + 2 d_y^T \nabla^2_{yy} p^\rho(y^*) d_y + \sum_{i=1}^{n} \gamma_i \left(d_x\right)_i \left(d_y\right)_i .$$

Observe that $d_y^T \nabla^2_{yy} p^\rho(y) d_y \geq 0$ holds due to the convexity of the (separable) function $p^\rho$, and that $\left(d_x\right)_i \left(d_y\right)_i = 0$ for all $i$ and all vectors $(d_x, d_y)$ from the critical cone of either (SPOref) or (SPOcp). Taking these observations into account, the equivalence of statement (a) to assertions (b) or (c) is easy to verify.

Note that, in general, we prefer to deal with the above two sparse second-order conditions since they are defined directly in terms of the sparse optimization problem (SPO), hence they are independent of the auxiliary variable $y$ and the function $p^\rho$ introduced in order to obtain the desired reformulations.

# 5 Approximate S-Stationarity

This section considers a sequential optimality condition which is the counterpart of the *approximate KKT conditions* (AKKT conditions for short) originally introduced for standard nonlinear programs of the form

$$\min_x \; f(x) \qquad \text{s.t.} \qquad g(x) \leq 0, \; h(x) = 0 \tag{11}$$

by [2] (with the name *cone continuity property*), see also [7]: A feasible point $x^*$ of (11) is called an *AKKT point* if there are sequences $\{x^k\} \subset \mathbb{R}^n$, $\lambda^k \subset \mathbb{R}^m_+$, and $\mu^k \subset \mathbb{R}^p$ such that

$$x^k \to x^*, \quad \nabla_x L(x^k, \lambda^k, \mu^k) \to 0, \quad \min\left\{-g_i(x^k), \lambda_i^k\right\} \to 0 \;\; \forall i = 1, \ldots, m,$$

where $L$ denotes the (ordinary) Lagrangian of (11).

The notion of an AKKT point has been generalized in different ways to optimization problems having a special and/or difficult structure, often coined *approximate M-stationarity* (AM-stationarity) since it is based on a sequential version of the M-stationary optimality conditions, see, e.g., [25] for a corresponding discussion in a very general setting.

In the following, we introduce a sequential optimality condition for the sparse optimization problem (SPO) which is based on the notion of an S-stationary point from Definition 4.2 and, therefore, takes into account the particular structure of this problem.

**Definition 5.1.** *Consider the sparse optimization problem* (SPO) *with feasible set* $X$ *being defined by* (9). *We then call a feasible point* $x^* \in X$ *an* approximate S-stationary point *(AS-stationary point, for short) if there exist sequences* $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m_+$, *and* $\{\mu^k\} \subset \mathbb{R}^p$ *such that*

$$x^k \to x^*, \quad \nabla_{x_i} L^{SP}(x^k, \lambda^k, \mu^k) \to 0 \ \ \forall i \notin I_0(x^*), \quad \min\{-g_i(x^k), \lambda_i^k\} \to 0 \ \ \forall i = 1, ..., m,$$

*where* $L^{SP}$ *denotes, once again, the SP-Lagrangian from* (10).

Similar to existing results on AKKT- and AM-stationary points, we can also derive several useful properties for our notion of an AS-stationary point in the context of sparse optimization problems. The first result in this context shows that any local minimum is automatically an AS-stationary point. Note that this statement holds without assuming any constraint qualification.

**Theorem 5.2.** *Let* $x^*$ *be a local minimum of the sparse optimization problem* (SPO) *with* $X$ *given by* (9). *Then* $x^*$ *is an AS-stationary point of* (SPO).

*Proof.* First recall that the local minimum $x^*$ of (SPO) is also a local minimum of the corresponding tightened nonlinear program from (TNLP($x^*$)). Therefore, standard results on AKKT points imply that there exists sequences $x^k \to x^*$ as well as $\{\lambda^k\} \subset \mathbb{R}^m_+$, $\{\mu^k\} \subset \mathbb{R}^p$, and $\{\gamma^k\} \subset \mathbb{R}^{|I_0(x^*)|}$ satisfiying

$$\nabla_x L^{SP}(x^k, \lambda^k, \mu^k) + \sum_{i \in I_0(x^*)} \gamma_i^k e_i \to 0 \quad \text{and} \quad \min\{-g_i(x^k), \lambda_i^k\} \to 0 \ \ \forall i = 1, \dots, m.$$

Hence, for all $i \notin I_0(x^*)$, this implies $\nabla_{x_i} L^{SP}(x^k, \lambda^k, \mu^k) \to 0$, and the claim follows. □

Theorem 5.2 shows that AS-stationarity is a necessary optimality condition. Of course, this does not automatically imply that AS-stationarity is a suitable (strong) optimality condition. In fact, it is known that, for certain classes of optimization problems like cardinality-constrained problems, the standard AKKT-conditions hold at every feasible point, cf. [31]. The following example shows that this unfortunate situation does not hold in our setting with AS-stationarity.

**Example 5.3.** Consider the (sparse) optimization problem

$$\min_x \sum_{i=1}^n x_i + \rho \|x\|_0 \qquad \text{s.t.} \qquad \|x\|_2^2 \le 1, \ x \ge 0.$$

The origin $x = 0$ is the only local and global minimum of this problem and, hence, also an AS-stationary point in view of Theorem 5.2. We claim that it is also the only AS-stationary point. In fact, suppose there exists an AS-stationary point $x^*$ with $x_i^* > 0$ for at least one component $i$. Since $x^*$ is an AS-stationary point, there exist suitable sequences $\{x^k\} \to x^*$ and $\{\lambda^k\} \in [0, \infty)$ such that, in particular, for the component $i$, we have $1 + 2\lambda^k x_i^k \to 0$, which is impossible for all $k$ sufficiently large since $\lambda^k \ge 0$ and $x_i^k \to x_i^* > 0$. Hence, $x = 0$ is the only AS-stationary point. Notice that the origin is also S-stationary in this example.

We next want to provide conditions under which an AS-stationary point is already S-stationary. The following example shows that this implication does not hold in general.

**Example 5.4.** Consider the (sparse) optimization problem

$$\min_x \; x_1 + \rho \|x\|_0, \qquad \text{s.t.} \qquad \frac{1}{2}\|x-e\|_2^2 = 0, \; x \geq 0.$$

The only feasible point and therefore local and global minimum is $x^* = e$. Theorem 5.2 implies that $x^*$ is AS-stationary. This can also be verified directly using Definition 5.1 and the sequences $x^k = (1 - \frac{1}{k}, 1, ..., 1)$, $\mu^k = k$, for which we obtain the desired limit

$$\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + k \cdot \begin{pmatrix} 1 - \frac{1}{k} - 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = 0$$

(note that the sequence $\{\mu^k\}$ is unbounded). We claim, however, that $x^*$ is not an S-stationary point. This is clear since otherwise we would have

$$0 = \nabla_{x_1} L^{SP}(x^*, \mu^*) = 1 + \mu(x_1^* - 1) = 1$$

for some multiplier $\mu \in \mathbb{R}$, which is impossible.

For an AS-stationary point to be S-stationary, we therefore require a suitable constraint qualification. The following is the natural counterpart of what is usually called AKKT-regularity or AM-regularity in the context of standard nonlinear programs or certain structured optimization problems, see [3, 2, 7] as well as [25] and references therein.

**Definition 5.5.** *Consider the sparse optimization problem* (SPO) *with feasible set $X$ being defined by* (9)*. Furthermore, let $x^* \in X$ be any given feasible point. We say that $x^*$ satisfies the* AS-regularity condition *if the cone*

$$K(x) := \left\{ d \,\middle|\, d_j = \Big[ \sum_{i=1}^{p} \mu_i \nabla h_i(x) + \sum_{i \in I_g(x^*)} \lambda_i \nabla g_i(x) \Big]_j, \; \mu_i \in \mathbb{R}, \; \lambda_i \geq 0, \; j \notin I_0(x^*) \right\}$$

*is outer semicontinuous at $x^*$, i.e., $Limsup_{x \to x^*} K(x) \subseteq K(x^*)$, where*

$$Limsup_{x \to x^*} K(x) := \left\{ d \,\middle|\, \exists \{x^k\} \to x^*, \exists d^k \to d : d^k \in K(x^k) \; \forall k \in \mathbb{N} \right\}$$

*denotes the upper or outer limit of the set-valued mapping $K(\cdot)$.*

Note that, in the previous definition of $K(x)$, the index set $I_0(x^*)$ is fixed at $x^*$. Moreover, we note that no condition is required for the components $d_j$ with $j \in I_0(x^*)$.

The following result shows that, in a certain sense, AS-regularity is a necessary and sufficient condition for an AS-stationary point to be S-stationary. In particular, this means that AS-regularity is a constraint qualification.

**Theorem 5.6.** *Let $x^*$ be feasible for* (SPO) *with feasible set $X$ defined in* (9)*. Then the following statements hold:*

1. *If $x^*$ is an AS-stationary point of (SPO) satisfying AS-regularity, then $x^*$ is an S-stationary point of (SPO).*

2. *Conversely, if for every continuously differentiable objective function $f$, the implication*

$$x^* \text{ is an AS-stationary point of (SPO)} \Longrightarrow x^* \text{ is an S-stationary point of (SPO)}$$

*holds, then $x^*$ is AS-regular.*

*Proof.* (a): Since $x^*$ is an AS-stationary point, there exist sequences $\{x^k\} \subset \mathbb{R}^n, \{\lambda^k\} \subset \mathbb{R}^m$, and $\{\mu^k\} \subset \mathbb{R}^p$ such that $x^k \to x^*$, $\nabla_{x_i} L^{SP}(x^k, \lambda^k, \mu^k) \to 0$ for all $i \notin I_0(x^*)$, and $\min\{-g_i(x^k), \lambda_i^k\} \to 0$ for all $i = 1, \ldots, m$. The latter condition implies that we may assume without loss of generality that $\lambda_i^k = 0$ for all $i \notin I_g(x^*)$ and all $k \in \mathbb{N}$, cf. [7, Thm. 3.2] for a formal proof. Then, writing

$$w^k := \sum_{i \in I_g(x^*)} \lambda_i^k g_i(x^k) + \sum_{j=1}^{p} \mu_j^k \nabla h_j(x^k)$$

we see that $w^k \in K(x^k)$ for all $k \in \mathbb{N}$, and that $\xi_i^k := [\nabla f(x^k) + w^k]_i \to 0$ for all $i \notin I_0(x^*)$. Define $\xi_i^k := 0$ for the remaining components $i \in I_0(x^k)$ and set $v^k := \xi^k - \nabla f(x^k)$. We then obtain $v^k \in K(x^k)$ for all $k \in \mathbb{N}$ since, for the relevant components $i \notin I_0(x^*)$, we have $v_i^k = \xi_i^k - [\nabla f(x^k)]_i = w_i^k$. The assumed AS-regularity of $x^*$ then implies

$$-\nabla f(x^*) = \lim_{k \to \infty} v^k \in \operatorname{Limsup}_{x \to x^*} K(x) \subseteq K(x^*),$$

which shows that $x^*$ is S-stationary.

(b): Take $w^* \in \operatorname{Limsup}_{x \to x^*} K(x)$ arbitrarily. Then there is $\{(x^k, w^k)\} \to (x^*, w^*)$ with $w^k \in K(x^k)$ for all $k \in \mathbb{N}$. Hence, there exist sequences $\{\lambda^k\} \subset \mathbb{R}_+^{|I_g(x^*)|}$ and $\{\mu^k\} \subset \mathbb{R}^p$ such that

$$w_i^k = \sum_{j \in I_g(x^*)} \lambda_j^k \nabla g_j(x^k)_i + \sum_{j=1}^{p} \mu_j^k \nabla h_j(x^k)_i \quad \forall i \notin I_0(x^*).$$

Define the function $f(x) := \sum_{i=1}^n -x_i w_i^*$ and choose $\lambda_i^k := 0$, for $i \notin I_g(x^*)$. Then clearly $\nabla f(x^k)_i + w_i^k \to 0$, and $x^*$ is an AS-stationary point. By assumption, $x^*$ is already S-stationary, which is equivalent to $w^* = -\nabla f(x^*) \in K(x^*)$. □

Having identified AS-regularity as a constraint qualifcation, the question is how this property is related to other SP-CQs. Among those given in Definition 4.1, the weakest one is SP-RCPLD. The following result shows that this condition still implies AS-regularity.

**Theorem 5.7.** *Let $x^*$ be feasible for (SPO) with feasible set $X$ defined by (9). Assume SP-RCPLD is satisfied at $x^*$. Then AS-regularity holds at $x^*$.*

*Proof.* By construction, SP-RCPLD implies standard RCPLD of for the tightened program (TNLP($x^*$)) at $x^*$. It is known from [2] that RCPLD yields AKKT-regularity for this program. This, however, is exactly our AS-regularity condition. □

# 6 An Exact Penalty Algorithm

Throughout this section, we assume that all functions $f, g$, and $h$ involved in the sparse optimization problem are at least continuously differentiable. Then (SPOcp) represents a smooth reformulation of the nonsmooth problem (SPO). Consequently, the reformulation (SPOcp) allows the application of a variety of different methods known from smooth optimization in order to solve the given sparse problem (SPO).

On the other hand, a suitable choice for solving (SPOcp) requires some care. For example, taking into account the almost separable structure of (SPOcp) in terms of the two variables $x$ and $y$, it is very tempting to apply an alternating minimization approach to this problem which uses a separate minimization with respect to the variables $x$ and $y$. This approach has the major advantage that the resulting subproblems are (usually) very easy to solve. However, this technique then terminates after the first cylce with an S-stationary point and, afterwards, does not make any further progress. This is unfortunate since the corresponding objective function value is typically very poor. In fact, this method often gets stuck at a local minimum with a relatively large function value, so that the method terminates with a point that is far away from being globally optimal.

In view of our experience, and in order to obtain good candidates for a global minimum, it is advantageous to apply a technique which might be feasible or approximately feasible with respect to the standard constraints $g(x) \leq 0$ and $h(x) = 0$, but with the complementarity term not approaching zero too fast, because this leaves some freedom in reducing the remaining objective function.

The aim of this section is therefore to present an exact penalty approach for the solution of the sparse optimization problem with feasible set $X$ given by (9). Our exact penalty method is based on the reformulation (SPOcp) and penalizes the (difficult) complementarity term only, whereas the remaining restrictions stay as constraints in the penalized problem. Using the standard $\ell_1$-penalty function, the penalized objective function then reads

$$f(x) + p^\rho(y) + \alpha \left| x^T y \right| = f(x) + p^\rho(y) + \alpha \left\| x \circ y \right\|_1 .$$

The $\ell_1$-term usually leads to a nonsmoothness of this penalty approach, which is the major drawback of this technique. In our particular situation, however, we have sign contraints on $x$ and $y$, cf. the reformulated problem (SPOcp) once again. Hence, we may remove the absolute value and obtain the following (smooth!) penalized version of (SPOref):

$$\min_{x,y} \; f(x) + p^\rho(y) + \alpha x^T y \quad \text{s.t.} \quad g(x) \leq 0, \; h(x) = 0, \; x \geq 0, \; y \geq 0. \tag{Pen($\alpha$)}$$

This motivates the following exact penalty-type algorithm.

**Algorithm 6.1.** *(Exact Penalty Method for Sparse Optimization)*

1. *Choose a non-negative sequence* $\varepsilon_k \searrow 0$ *and parameters* $\alpha_0 > 0$, $\beta > 1$, *and* $\delta \geq 0$.
2. *For* $k = 0, 1, 2, ...,$ *compute* $(x^{k+1}, y^{k+1}, \lambda^{k+1}, \mu^{k+1}, \nu_x^{k+1}, \nu_y^{k+1}) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+ \times \mathbb{R}^p_+ \times \mathbb{R}^m \times$

$\mathbb{R}^n_+ \times \mathbb{R}^n_+$ *such that*

$$\begin{aligned}
\|\nabla_x L^{SP}(x^{k+1}, \mu^{k+1}, \lambda^{k+1}) - \sum_{i=1}^{n} (\nu_x^{k+1})_i e_i + \alpha_k y^{k+1}\| &\leq \varepsilon_k \\
\|\nabla p^\rho(y^{k+1}) - \sum_{i=1}^{n} (\nu_y^{k+1})_i e_i + \alpha_k x^{k+1}\| &\leq \varepsilon_k \\
\min |\{-g_i(x^{k+1}), \lambda_i^{k+1}\}| &\leq \varepsilon_k, \quad i = 1, ..., p \\
\|h(x^{k+1})\| &\leq \varepsilon_k, \\
\min\{x_i^{k+1}, (\nu_x^{k+1})_i\} &\leq \varepsilon_k, \quad i = 1, ..., n \\
\min\{y_i^{k+1}, (\nu_y^{k+1})_i\} &\leq \varepsilon_k, \quad i = 1, ..., n
\end{aligned}$$

*3. If*

$$\varepsilon_k \leq \delta \quad \textit{and} \quad \sum_{i=1}^{n} x_i^{k+1} y_i^{k+1} \leq \delta$$

*then STOP. Otherwise set* $\alpha_{k+1} = \alpha_k \cdot \beta$ *and go to 2.*

The main computational burden, of course, is in step 2. Note that we do not require to solve the penalized subproblems exactly. The tests within step 2 only check whether we have an approximate KKT point of the penalized problem (Pen($\alpha$)), with $\varepsilon_k$ denoting the measure of inexactness, and with $\lambda^{k+1}, \mu^{k+1}, \nu_x^{k+1}$, and $\nu_y^{k+1}$ being the Lagrange multipliers associated to the constraints $g(x) \leq 0, h(x) = 0, x \geq 0$, and $y \geq 0$, respectively. This general framework allows plenty of methods in order to deal with the penalized subproblems, which is an important feature of the overall method since a suitable choice depends on the particular problem under consideration. In addition, we note that some methods might deal with the nonnegativity constraints $x \geq 0, y \geq 0$ explicitly, so that these methods do not generate corresponding multiplier estimates $\nu_x^{k+1}, \nu_y^{k+1}$. In this situation, we can simply delete the two final tests in step 2, and replace the first two by the related (multiplier-free) tests

$$\begin{aligned}
\|P_{[0,\infty)}\left(x^{k+1} - (\nabla_x L^{SP}(x^{k+1}, \mu^{k+1}, \lambda^{k+1}) + \alpha_k y^{k+1})\right) - x^{k+1}\| &\leq \varepsilon_k, \\
\|P_{[0,\infty)}\left(y^{k+1} - (\nabla p^\rho(y^{k+1}) + \alpha_k x^{k+1})\right) - y^{k+1}\| &\leq \varepsilon_k.
\end{aligned}$$

The subsequent theory remains true with this kind of test, too. Our analysis, however, concentrates on the inexactness measure from step 2.

The final step 3 represents the stopping criterion for the outer iteration. Based on the termination parameter $\delta \geq 0$, we simply check whether we are (approximately) feasible both with respect to the standard constraints and with respect to the penalized complementarity constraints. For the (theoretical) choice $\delta = 0$, it follows immediately that we terminate with a feasible point. The following result shows that, in this case, we even have an S-stationary point.

**Lemma 6.2.** *Let* $\delta = 0$ *and assume that Algorithm 6.1 terminates after a finite number of steps in a point* $(x^*, y^*)$*. Then* $(x^*, y^*)$ *is a KKT point of* (SPOref)*, and* $x^*$ *is S-stationary.*

*Proof.* Let $\lambda^*, \mu^*, \nu_x^*, \nu_y^*$ be the associated multipliers of the iterate $(x^*, y^*)$. We first show that $x^*$ is S-stationary. Since $\delta = 0$ by assumption, we also have from step 3 that $\varepsilon_k = 0$ for

the corresponding iteration $k$. Therefore, step 2 immediately implies that $x^*$ is feasible for the sparse optimization problem (SPO), and $(x^*, \lambda^*)$ satisfies the complementarity conditions. Now, consider an index $i$ with $x_i^* \neq 0$. Then steps 2 and 3 yield $y_i^* = 0$ and $(\nu_x^*)_i = 0$. Using step 2 again, this yields $\nabla_{x_i} L^{SP}(x^*, \lambda^*, \mu^*) = 0$. By definition, this shows that $x^*$ is an S-stationary point.

Next, consider an index $i$ with $x_i^* = 0$. Then we have $\nabla p_i^\rho(y_i^*) - (\nu_y^*)_i = 0$ from step 2, which implies $\nabla p_i^\rho(y_i^*) \geq 0$ and hence $y_i^* \geq s_i^\rho > 0$ by convexity of $p_i^\rho$ and uniqueness of $s_i^\rho$. As a result $(\nu_y^*)_i = 0$ and $y_i^* = s_i^\rho$. Consequently, the vector $y^*$ satisfies the relation (6). Theorem 4.3 therefore shows that $(x^*, y^*)$ is a KKT point of (SPOref). □

We next want to show an exactness result for our penalty approach. In principle, there are two directions one is interested in, namely that a stationary point of the penalized problem (Pen($\alpha$)) yields a stationary point of the sparse optimization problem (SPO), and vice versa. Exactness results usually concentrate on the opposite direction only. In fact, this direction is usually the easier one and, for our particular setting, contained in the following result.

**Theorem 6.3.** *Let $(x^*, y^*)$ be a stationary point of (SPOcp). Then there exists an $\alpha^* > 0$ such that $(x^*, y^*)$ is a stationary point of (Pen($\alpha$)) for all $\alpha \geq \alpha^*$.*

*Proof.* By stationarity of $(x^*, y^*)$ for (SPOcp), we obtain, after a simple transformation, that

$$0 = \begin{pmatrix} \nabla f(x^*) + h'(x^*)^T \mu^* + g'(x^*)^T \lambda^* + \sum_{I_0(x^*)} \alpha_i^x e_i \\ \nabla p^\rho(y^*) + \sum_{I_0(y^*)} \alpha_i^y e_i \end{pmatrix} \tag{12}$$

holds for some multipliers $\mu^*, \lambda^*, \alpha^x, \alpha^y$ with sign constraints only w.r.t. to $\lambda^*$, and further $y_i^* = s_i^\rho$ if and only if $i \in I_0(x^*)$. On the other hand, the point $(x^*, y^*)$ is stationary for (Pen($\alpha$)) if there exists multipliers $\mu$, $\lambda \geq 0$, $\nu^x \geq 0$, and $\nu^y \geq 0$ such that

$$0 = \begin{pmatrix} \nabla f(x^*) + h'(x^*)^T \mu + g'(x^*)^T \lambda - \sum_{I_0(x^*)} \nu_i^x e_i + \alpha \sum_{I_0(x^*)} s_i^\rho e_i \\ \nabla p^\rho(y^*) - \sum_{I_0(y^*)} \nu_i^y e_i + \alpha \sum_{I_0(y^*)} x_i^* e_i \end{pmatrix} \tag{13}$$

holds. Now, setting $\mu = \mu^*$, $\lambda = \lambda^*$, $\nu_i^x = \alpha s_i^\rho - \alpha_i^x$ $(i \in I_0(x^*))$ and $\nu_i^y = \alpha x_i^* - \alpha_i^y$ $(i \in I_0(y^*))$ for an arbitrary $\alpha > 0$, we see that (13) is a direct consequence of (12). Moreover, for

$$\alpha \geq \alpha^* := \max \left\{ \frac{|\alpha_i^x|}{s_i^\rho} \ (i \in I_0(x^*)), \frac{|\alpha_i^y|}{x_i^*} \ (i \in I_0(y^*)) \right\},$$

we also have $\nu_i^x \geq 0$ $(i \in I_0(x^*))$ and $\nu_i^y \geq 0$ $(i \in I_0(y^*))$. It follows that $(x^*, y^*)$ is a stationary point of (Pen($\alpha$)) for all $\alpha \geq \alpha^*$. □

Note that Theorem 6.3 does not require any constraint qualification. Typical exactness results of this kind need an MFCQ-type assumption which, here, is not necessary for two reasons: First, the (potentially simple) standard constraints are still in the constraints of the penalized problem, and second, the penalized (difficult) complementarity constraint has a very simple structure such that no constraint qualification is needed to verify the exactness statement from Theorem 6.3. In fact, this complementarity constraint alone satisfies automatically any constraint qualification.

Before presenting an exactness result for the other direction, we first consider a simple example.

**Example 6.4.** Consider the one-dimensional sparse optimization problem

$$\min_x \ -x + \|x\|_0 \quad \text{s.t.} \quad x \geq 0$$

(note that we take $\rho = 1$ only for the sake of simplicity). Then $x^* := 0$ is both a local minimum and an S-stationary point. Using the function $p_i^\rho$ from Example 3.1 (b), the corresponding penalized problem (Pen($\alpha$)) reads

$$\min_{x,y} \ -x + \frac{1}{2}(y - \sqrt{2})^2 + \alpha xy \quad \text{s.t.} \quad x \geq 0,\ y \geq 0$$

with an arbitrary $\alpha > 0$. Then, for any $\alpha \geq 1/\sqrt{2}$,

$$(x, y) := \frac{1}{\alpha}\left(\sqrt{2} - \frac{1}{\alpha}, 1\right)$$

are stationary points of problem (Pen($\alpha$)), but none of these points satisfy the complementarity conditions. On the other hand, observe that, for $\alpha \to \infty$, the corresponding sequence of stationary points converges to $(x^*, y^*) = (0, 0)$, hence the complementarity condition $x^* y^* = 0$ is satisfied in the limit. Note, however, that this limit point does not satisfy the relation (6) which, in particular, guarantees that the bi-active set is empty.

The following result contains an exactness statement for the other direction. This result may be viewed as a generalization of a related theorem given in [30] in the context of mathematical programs with equilibrium contraints (MPECs), though our assumptions are weaker.

**Theorem 6.5.** *Let $(x^*, y^*)$ be a stationary point of (SPOcp) such that SP-MFCQ holds at $x^*$. Then there exists an $\alpha^* > 0$ and a neighborhood $U$ of $(x^*, y^*)$ such that for all $\alpha \geq \alpha^*$, every stationary point of (Pen($\alpha$)) in $U$ is a stationary point of (SPOcp).*

*Proof.* Assume, by contradiction, that there is a sequence $\alpha_k \to \infty$ and a sequence $\{(x^k, y^k)\}$ such that $(x^k, y^k) \to (x^*, y^*)$, where $(x^k, y^k)$ is stationary for (Pen($\alpha$)) with $\alpha = \alpha_k$, but not stationary for (SPOcp). Recall that stationarity of $(x^k, y^k)$ for (Pen($\alpha$)) with $\alpha = \alpha_k$ implies the existence of multipliers $(\mu^k, \lambda^k, \nu_x^k, \nu_y^k)$ such that

$$0 = \begin{pmatrix} \nabla f(x^k) + h'(x^k)^T \mu^k + \sum_{i \in I_g(x^k)} \lambda_i^k \nabla g_i(x^k) + \alpha_k y^k - \sum_{i \in I_0(x^k)} (\nu_x^k)_i e_i \\ \nabla p^\rho(y^k) + \alpha_k x^k - \sum_{i \in I_0(y^k)} (\nu_y^k)_i e_i \end{pmatrix}, \tag{14}$$

where $x^k \in X$ and $y^k \geq 0$. We divide the proof into several steps.

(a) Since $(x^*, y^*)$ is a stationary point of (SPOcp), we recall from Remark 4.4 that the biactive set $\mathscr{B}(x^*, y^*) = \{i \mid (x_i^*, y_i^*) = (0, 0)\}$ is empty. We stress that this observation plays a central role for the subsequent proof.

(b) Since $x^k \to x^*$, the continuity of $g$ implies that $I_g(x^k) \subseteq I_g(x^*)$ for all $k$ sufficiently large. Because $\lambda_i^k = 0$ for all $i \notin I_g(x^k)$, we may therefore replace the index set $I_g(x^k)$ in (14) with the constant set $I_g(x^*)$ for all $k$ large enough. Hence, we have both

$$\nabla f(x^k) + h'(x^k)^T \mu^k + \sum_{i \in I_g(x^*)} \lambda_i^k \nabla g_i(x^k) + \alpha_k y^k - \sum_{i \in I_0(x^k)} (\nu_x^k)_i e_i = 0 \tag{15}$$

and

$$\nabla p^{\rho}(y^k) + \alpha_k x^k - \sum_{i \in I_0(y^k)} (\nu_y^k)_i e_i = 0 \tag{16}$$

for all $k$ sufficiently large.

(c) We claim that $y_i^k = 0$ holds for all $i \in I_0(y^*)$ and all $k$ sufficiently large. To this end, assume there is an index $i \in I_0(y^*)$ and a subsequence $\{y_i^k\}_K$ such that $0 < y_i^k \to_K y_i^* = 0$. It then follows that $(\nu_y^k)_i = 0$ for all $k \in K$. Hence (16) implies

$$0 = \nabla p^{\rho}(y_i^k) + \alpha_k x_i^k \quad \forall k \in K.$$

On the other hand, the right-hand side tends to $+\infty$ for $k \to_K \infty$ since $y_i^k \to_K y_i^*$ and the continuity of $\nabla p^{\rho}$ implies the convergence of the first term, whereas the second term is unbounded since $\alpha_k \to_K \infty$ and $x_i^k \to x_i^*$ with $x_i^* > 0$ due to (a).

(d) We claim that

$$\left[\nabla f(x^k) + h'(x^k)^T \mu^k + \sum_{i \in I_g(x^*)} \lambda_i^k \nabla g_i(x^k)\right]_i = 0 \quad \forall i \in I_0(y^*) \tag{17}$$

for all $k$ sufficiently large. This follows from (15) together with the fact that $y_i^k = 0$ for all $i \in I_0(y^*)$ and $k$ sufficiently large by part (c), and since $I_0(x^k) \subseteq I_0(x^*)$ with $I_0(x^*) \cap I_0(y^*) = \emptyset$ by part (a).

(e) In view of (a) and (17), we can find scalars $\gamma_i^k \in \mathbb{R}$ for $i \in I_0(x^*)$ such that

$$\nabla f(x^k) + h'(x^k)^T \mu^k + \sum_{i \in I_g(x^*)} \lambda_i^k \nabla g_i(x^k) + \sum_{i \in I_0(x^*)} \gamma_i^k e_i = 0$$

holds for all $k$ large enough. Due to the assumed SP-MFCQ condition, a standard argument then shows that the sequence of multipliers $\left\{\left(\lambda_i^k (i \in I_g(x^*)), \mu^k, \gamma_i^k (i \in I_0(x^*))\right)\right\}$ remains bounded.

(f) We also claim that $x_i^k = 0$ for all $i \in I_0(x^*)$ and all $k$ sufficiently large. Assume, by contradiction, that there is a subsequence $\{x_i^k\}_K$ with $0 < x_i^k \to_K x_i^*$. Then $(\nu_x^k)_i = 0$ holds for all $k \in K$. Consequently, we obtain from (15) that

$$0 = \nabla_{x_i} L^{SP}(x^k, \lambda^k, \mu^k) + \alpha_k y_i^k.$$

Now, the first term on the right-hand side remains bounded by continuity of $\nabla_{x_i} L^{SP}$ as well as the fact that $x^k \to x^*$ and the boundedness of the multiplier sequences $\{\lambda^k\}$ and $\{\mu^k\}$, cf. part (e). On the other hand, the second term converges to $+\infty$ since $\alpha_k \to \infty$ and $y_i^k \to_K y_i^* > 0$ for $i \in I_0(x^*)$, see part (a).

(g) In view of parts (c) and (f), we, in particular, have $x_i^k y_i^k = 0$ for all $i = 1, \ldots, n$ and all $k$ sufficiently large. This shows that $(x^k, y^k)$ is, at least, feasible for (SPOcp). Furthermore, since $(x^k, y^k) \to (x^*, y^*)$ and $\mathscr{B}(x^*, y^*) = \emptyset$ by part (a), we also have $\mathscr{B}(x^k, y^k) = \emptyset$ for all $k$ large enough. We can therefore define the multipliers

$$(\alpha_i^x)^k := \alpha_k y_i^k - (\nu_x^k)_i \quad (i \in I_0(x^k)) \quad \text{and} \quad (\alpha_i^y)^k := \alpha_k x_i^k - (\nu_y^k)_i \quad (i \in I_0(y^k)),$$

so that the stationary conditions of the penalized problem yield

$$0 = \begin{pmatrix} \nabla f(x^k) + h'(x^k)^T \mu^k + \sum_{i \in I_g(x^k)} \lambda_i^k \nabla g_i(x^k) + \sum_{I_0(x^k)} (\alpha_i^x)^k e_i \\ \nabla p^\rho(y^k) + \sum_{I_0(y^k)} (\alpha_i^y)^k e_i \end{pmatrix}$$

for all $k$ sufficiently large, cf. (14). Altogether, this implies that $(x^k, y^k)$ is a stationary point of (SPOref) for all $k$ sufficiently large, and this contradiction completes the proof □

Observe that part (d) of the previous proof already shows that $x^k$, for all $k$ sufficiently large, is an S-stationary point of (SPO), and that this holds without any constraint qualification. The SP-MFCQ assumption is mainly used to show that the pair $(x^k, y^k)$ is eventually feasible for (SPOcp), i.e., satisfies the complementarity condition $x \circ y = 0$.

Note further that Example 6.4 does not contradict the statement of Theorem 6.5. Though SP-MFCQ holds for this example in $x^* = 0$, the sequence of stationary points of the corresponding penalized problems converges to $(0, 0)$, which is not a stationary point (SPOcp), as assumed in Theorem 6.5.

The following result provides a relation between the second-order condition of the penalized problem (Pen($\alpha$)) and SP-SOSC for the sparse optimization problem.

**Theorem 6.6.** *Let $(x^*, y^*)$ be stationary for (Pen($\alpha$)) and feasible for (SPOcp). Assume that $p^\rho$ is twice continuously differentiable and that standard SOSC holds at (Pen($\alpha$)). Then SP-SOSC holds at $x^*$.*

*Proof.* An elementary calculation shows that standard SOSC for (Pen($\alpha$)) is given by

$$\begin{aligned} & \begin{pmatrix} dx \\ dy \end{pmatrix}^T \begin{pmatrix} \nabla^2_{xx} L^{SP}(x^*, \mu^*, \lambda^*) & \alpha I \\ \alpha I & \nabla^2_{yy} p^\rho(y) \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} > 0 \\ \iff \quad & (dx)^T \nabla^2_{xx} L^{SP}(x^*, \mu^*, \lambda^*) dx + 2\alpha (dx)^T dy + (dy)^T \nabla^2_{yy} p^\rho(y) dy > 0, \end{aligned} \tag{18}$$

for all $(dx, dy) \neq (0, 0)$ such that

$$\begin{aligned} dx_i &= 0, \ (i \in I_0(x^*), \ \nu_i^x > 0), \\ dx_i &\geq 0, \ (i \in I_0(x^*), \ \nu_i^x = 0), \\ \nabla h_i(x^*)^T dx &= 0, \ (i = 1, ..., m), && (19) \\ \nabla g_i(x^*)^T dx &= 0, \ (i \in I_g(x^*), \ \lambda_i^* > 0), && (20) \\ \nabla g_i(x^*)^T dx &\leq 0, \ (i \in I_g(x^*), \ \lambda_i^* = 0), && (21) \\ dy_i &= 0, \ (i \in I_0(y^*), \nu_i^y > 0), \\ dy_i &\geq 0, \ (i \in I_0(y^*), \nu_i^y = 0), \end{aligned}$$

where $\nu_i^x, \nu_i^y$ and $\lambda_i^*$ are the Lagrangian multipliers associated to the sign constraints on $x$ and $y$ and to the inequality constraints governed by $g$, respectively. Now, choose $dy = 0$ and $dx$ such that $dx_i = 0$ for all $i \in I_0(x^*)$ and conditions (19), (20) and (21) hold. The claim then follows directly from (18). □

A difficulty with (exact) penalty approaches for general optimization problems is that accumulation points are not guaranteed to be feasible. In our setting, this feasibility issue arises for the complementarity constraints only, and it turns out that, due to the particular structure of our reformulated problem, these complementarity conditions are satisfied even if Algorithm 6.1 does not terminate after finitely many iterations.

**Theorem 6.7.** *Let $\delta = 0$, and let $\{(x^k, y^k)\}$ be an infinite sequence generated by Algorithm 6.1 such that $x^{k+1} \to_K x^*$ on some subsequence $K$. Then there is a subset $K' \subseteq K$ such that $y^{k+1} \to_{K'} y^*$ and $x_i^* y_i^* = 0$ holds for all $i = 1, ..., n$.*

*Proof.* We first show that the corresponding subsequence $\{y^{k+1}\}_K$ remains bounded. By contradiction, assume that there is an index $i$ such that $\{y_i^{k+1}\}$ is unbounded. Due to the nonnegativity constraint, we may therefore assume, without loss of generality, that $y_i^{k+1} \to_K \infty$. In particular, we then have $y_i^{k+1} \geq 2 \cdot s_i^\rho$ for infinitely many $k \in K$. The convexity of $p_i^\rho$ then implies

$$\nabla p_i^\rho(y_i^{k+1}) \geq \nabla p_i^\rho(2s_i^\rho) =: c > \nabla p_i^\rho(s_i^\rho) = 0.$$

In particular, we then have

$$\nabla p_i^\rho(y_i^{k+1}) + \alpha_k x_i^{k+1} \geq c$$

and therefore $(\nu_y^{k+1})_i \geq c/2$ for infinitely many $k \in K$ due to the second termination check in step 2. On the other hand, by the final condition in step 2, we have

$$\min\{y_i^{k+1}, (\nu_y^{k+1})_i\} \to 0,$$

and this contradiction shows that $\{y^{k+1}\}_K$ is indeed a bounded sequence.

Consequently, there is a subset $K' \subseteq K$ such that $\{y^{k+1}\}_{K'}$ converges to some point $y^*$. We claim that $x_i^* y_i^* = 0$ holds for this limit for all $i = 1, \ldots, n$. For $x_i^* = 0$, there is nothing to prove. Hence consider an index $i$ with $x_i^* > 0$. Then clearly $\alpha_k x_i^{k+1} \to_K \infty$. Since $p_i^\rho$ is convex by assumption, its derivative $\nabla p_i^\rho$ is monotone. Taking into account the sign restriction $y_i^{k+1} \geq 0$, we therefore obtain

$$\nabla p_i^\rho(y_i^{k+1}) \geq \nabla p_i^\rho(0).$$

This implies

$$\nabla p_i^\rho(y_i^{k+1}) + \alpha_k x_i^{k+1} \to_{K'} \infty,$$

and the second termination check in step 2 of Algorithm 6.1 therefore yields

$$(\nu_y^{k+1})_i \to_{K'} \infty.$$

Hence, the sixth condition in step 2 immediately gives $y_i^{k+1} \to_{K'} 0 = y_i^*$, and this completes the proof. □

The above theorem shows that every accumulation point of Algorithm 6.1 is indeed (approximately) feasible for (SPOref). Note that this is not as surprising as it seems in the beginning. In fact, when looking at the original problem SPO, the feasible set is given by $X$ and depends on the variables $x$ alone. Moving to an auxiliary variable $y$ should not increase the difficulty to find feasible points for the reformulation.

In general, we cannot guarantee to obtain approximate stationary points if the algorithm does not terminate after a finite number of iterations. We may, however, choose $\alpha_k$ and $\varepsilon_k$ in dependence to recover such a result.

**Theorem 6.8.** *Let $\delta = 0$, and let $\{(x^k, y^k)\}$ be an infinite sequence generated by Algorithm 6.1 such that $x^{k+1} \to_K x^*$ on a subsequence $K$. Then the following statements hold:*

*(a) If $y_i^{k+1}\alpha_k \to_K 0$ for all $i \notin I_0(x^*)$, then $x^*$ is an AS-stationary point.*

*(b) If $\varepsilon_k \alpha_k \to 0$, then $y_i^{k+1}\alpha_k \to 0$ for all $i \notin I_0(x^*)$.*

*Proof.* (a) First recall that $\{x^{k+1}\}_K \to x^*$ by assumption, and that $\min\{-g_i(x^{k+1}), \lambda_i^{k+1})\} \to_K 0$ follows from the third test in Algorithm 6.1. Hence, it remains to show that

$$\nabla_{x_i} L^{SP}(x^{k+1}, \lambda^{k+1}, \mu^{k+1}) \to_K 0 \quad (i \notin I_0(x^*)) \tag{22}$$

holds. Therefore, consider an arbitrary index $i \notin I_0(x^*)$, so that $x_i^* > 0$. Then it follows from the fifth test in step 2 that $\{(\nu_x^{k+1})_i\} \to_K 0$. Together with the assumption $y_i^{k+1}\alpha_k \to_K 0$, we see that (22) follows from the first test in step 2.

(b) Consider an index $i \notin I_0(x^*)$, so that $x_i^* > 0$. We then have $\alpha_k x_i^{k+1} \to_K \infty$. Then the second condition in step 2 of Algorithm 6.1 implies $(\nu_y^{k+1})_i \to_K \infty$ since $\nabla p_i^\rho$ takes its (finite) minimum at 0. Multiplying condition 6 by $\alpha_k$ yields

$$\min\{\alpha_k y_i^{k+1}, \underbrace{\alpha_k (\nu_y^{k+1})_i}_{\to\infty}\} \leq \alpha_k \varepsilon_k \to 0$$

and hence $\alpha_k y_i^{k+1} \to 0$. This completes the proof. □

Statement (a) of Theorem 6.8 provides a condition under which an arbitrary limit point of the sequence $\{x^k\}$ is an AS-stationary point and, hence, an S-stationary point under any of the SP-type constraint qualifications discussed in Section 5. Statement (b) then gives a sufficient condition under which the assumption from part (a) holds. Note that this sufficient condition can be realized. In fact, in iteration $k$, we have a penalty parameter $\alpha_k$, and then choose a termination parameter $\varepsilon_k$ such that $\varepsilon_k = o(1/\alpha_k)$ holds. From a practical point of view, however, this means that we might have to choose $\varepsilon_k$ small, possibly even at a relatively early stage of the entire method, hence it is not clear whether such a choice is always desirable.

# 7 Numerical Experiments

The aim of this section is to present a variety of applications where our exact penalty approach can be applied to. We recall that all these applications are extremely difficult due to the $\ell_0$-term (in the original formulation of the sparse optimization problem), nevertheless, the numerical results indicate that we find very good candidates for a solution of the underlying problem, quite often even the global minimum.

## 7.1 Sparse Portfolio Optimization

### 7.1.1 Preliminaries

The sparse portfolio optimization can be stated in the form

$$\min_x \; x^T Q x + \rho \|x\|_0 \quad \text{s.t.} \quad e^T x = 1, \; \mu^T x \geq s, \; 0 \leq x \leq u, \tag{23}$$

with a positive (semi-) definite covariance matrix $Q$, $\mu \in \mathbb{R}^n$ the mean of $n$ possible assets, $s > 0$ the minimum amount of (expected) return, $e = (1, 1, ..., 1)^T$, and $u_i$ an upper bound for the variable $x_i$ which represents the percentage of our total investment into asset $i$. Hence, the economic interpretation of the portfolio model (23) is, basically, as follows: The customer is willing to spend a certain amount of money in a few (due to the $\ell_0$-term) possible assets in such a way that he minimizes the risk (represented by the objective function) and has at least a

minimum return. Note that adding an $\ell_1$-term instead of an $\ell_0$-term does not yield any sparsity due to the constraints of this problem.

Note that (23) can also be written as a mixed integer quadratic program. In fact, using an auxiliary variable $z$, problem (23) is equivalent (in terms of global solutions) to

$$\min_{(x,z)} \; x^TQx + \rho e^Tz, \quad \text{s.t.} \quad e^Tx = 1,\; \mu^Tx \geq s,\; u \circ z \geq x \geq 0,\; z \in \{0,1\}^n, \tag{24}$$

cf. [6]. Since there exists commercially available software to tackle these types of problems like CPLEX or Gurobi, it is, in principle, possible to find the global minimum. This, in turn, allows to compare the quality of solutions obtained by our exact penalty technique.

It is useful to point out that the constraints in (23) are all polyhedral. This directly implies for the SP-RCPLD and therefore AS-regularity to hold.

Application of our technique yields the corresponding penalized problem (Pen($\alpha$))

$$\min_{x,y} \; x^TQx + p^\rho(y) + \alpha x^Ty \quad \text{s.t.} \quad e^Tx = 1,\; \mu^Tx \geq s,\; 0 \leq x \leq u,\; 0 \leq y. \tag{25}$$

If we follow Example 3.1 (b) and choose $p^\rho(y) := 1/2 \left\| y - \sqrt{2\rho}e \right\|_2^2$, we may rewrite (25) in the form

$$\min_{x,y} \; \begin{pmatrix} x \\ y \end{pmatrix}^T \hat{Q} \begin{pmatrix} x \\ y \end{pmatrix} - \sqrt{2\rho}e^Ty \quad \text{s.t.} \quad e^Tx = 1,\; \mu^Tx \geq s,\; 0 \leq x \leq u,\; 0 \leq y, \tag{26}$$

where

$$\hat{Q} = \frac{1}{2} \begin{pmatrix} 2Q & \alpha I \\ \alpha I & I \end{pmatrix}.$$

Note that this is a quadratic program with a positive definite Matrix $\hat{Q}$ as long as $\alpha < \sqrt{2\lambda_{\min}}$, where $\lambda_{\min} := \min\{\lambda \in \sigma(Q)\}$ denotes the minimum eigenvalue of the matrix $Q$. Hence, for this choice of $\alpha$, solutions of (25) are unique and easy to compute, but not necessarily feasible for (SPOref). However, this motivates an initial choice $\alpha_0 = \sqrt{2\lambda_{\min}} \cdot c$, $c \in (0,1)$ for the penalty parameter. In our set of test instances, all $\lambda_{\min}$ happened to be strictly positive. We further stress that, in case of a positive definite matrix $Q$, it is easy to see from equation (18) that, at a feasible point $(x^*, y^*)$, SOSC is satisfied for (SPOref).

### 7.1.2 Numerical Test

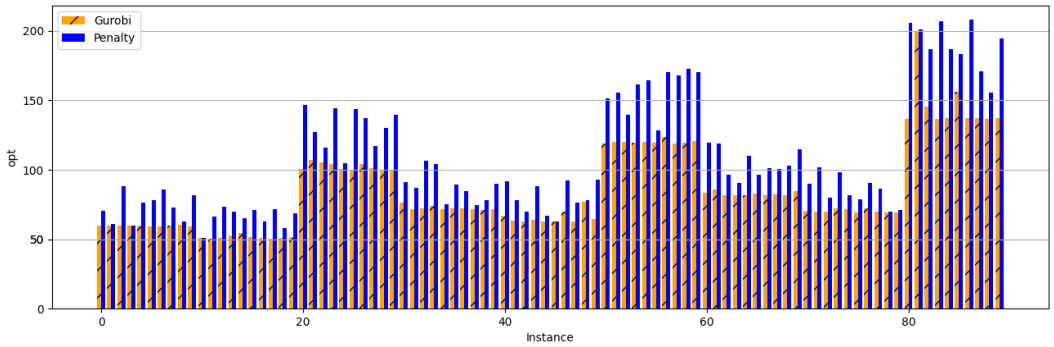
For our numerical tests, we have chosen an instance of problems provided by Frangioni and Gentile[1], with the constraints $(1-y_i)l_i \leq x_i \leq (1-y_i)u_i$, corresponding to $x_i = 0$ or $x_i \in [l_i, u_i]$, relaxed to

$$0 \leq x_i \leq (1-y_i)u_i.$$

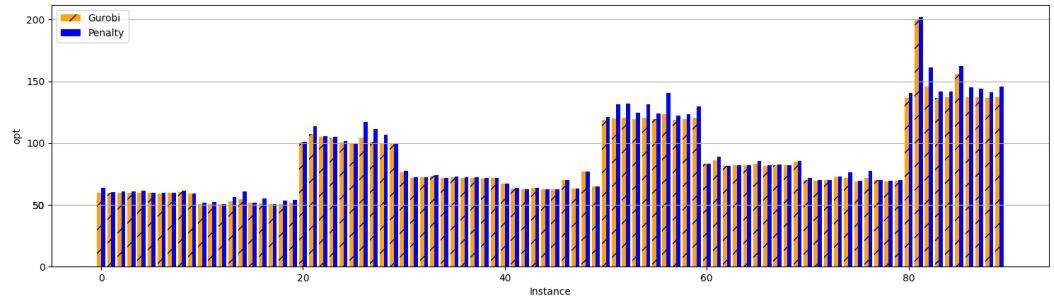
We first applied the branch-and-bound type algorithm by Gurobi[2] to the mixed-integer reformulation (24) to approximate a global minimum. In accordance to the previous subsection, we then took $p^\rho(y) = 1/2 \left\| y - \sqrt{2\rho}e \right\|_2^2$, $\rho = 1$, and computed via Python $\lambda_{\min}$ and $x^0$ as solution to the quadratic programm (26) for an $\alpha$ just below $\sqrt{2\lambda_{\min}}$ with a call to the corresponding Gurobi-module. This process took up 21 seconds of CPU-time for all of the 90 test instances.

[1] http://groups.di.unipi.it/optimize/Data/MV.html

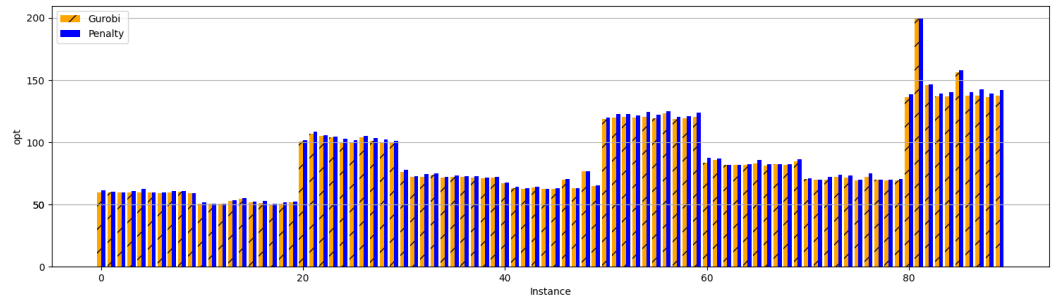
[2] https://www.gurobi.com/solutions/gurobi-optimizer/



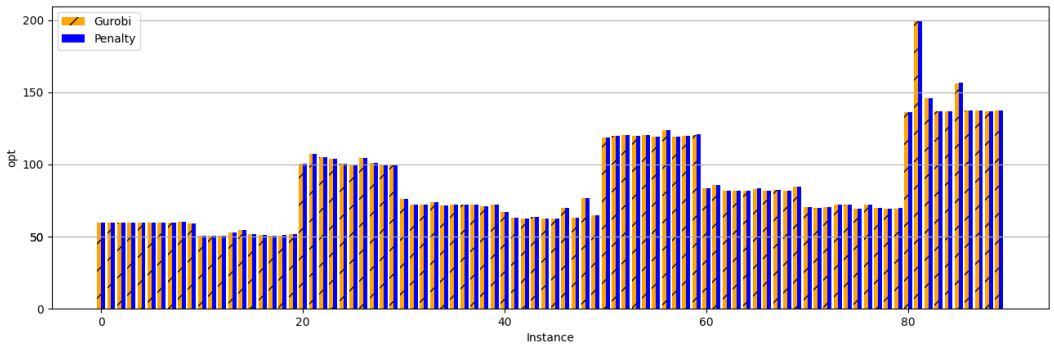

**(a)** Portfolio results by Gurobi and the penalty approach with a Huber-type $p^\rho$ and $\beta = 1.1$




**(b)** Portfolio results by Gurobi and the penalty approach with a smooth $p^\rho$ and $\beta = 5$.




**(c)** Portfolio results by Gurobi and the penalty approach with a smooth $p^\rho$ and $\beta = 2$.




**(d)** Portfolio results by Gurobi and the penalty approach with a smooth $p^\rho$ and $\beta = 1.1$.

**Figure 1:** Overview of the sparse portfolio tests.

Afterwards we used our exact penalty approach by computing a solution of (25) via Algencan

under Fortran, with the initial choice of $x^0$ as described above, and

$$y^0 := \sqrt{2\lambda_{\min}}, \quad \alpha_0 := \sqrt{2\lambda_{min}}/0.95, \quad \beta \in \{1.1, 2, 5\}.$$

Under Algencan, the subproblems in step two of Algorithm 6.1 were solved to an accuracy of $10^{-6}$ with specifically a tolerance of $10^{-8}$ in feasibility. The execution was halted once the solution $(x, y)$ provided by Algencan would fulfill the condition

$$x^T y \leq 10^{-6}.$$

With decreasing choices of $\beta$ we are able to recover the global solution given by Gurobi (compare plots in Figures 1b - 1d). In fact, the results are already very promising for the penalty updating factors $\beta = 10$ and, especially, $\beta = 5$, and for $\beta = 1.1$, the optimal function values computed by our exact penalty scheme coincides with the optimal function values provided by Gurobi for *all* 90 instances. Regarding the CPU-time for $\beta = 1.1$, the total time for the computation of the 30 instances with dimension 200 was around 4.1 seconds, for dimension 300 at 17.5 seconds, and for dimension 400 at 25 seconds.

In a second run, we replaced $p^\rho(y)$ via a piecewise Huber-type function in the sense that

$$p_i^\rho(y) = \xi \cdot \begin{cases} \varepsilon(y - \sqrt{2\rho} - \varepsilon) + \frac{1}{2}\varepsilon^2, & y > \rho + \varepsilon, \\ \frac{1}{2}(y_i - \sqrt{2\rho})^2, & y \in [\rho - \varepsilon, \rho + \varepsilon], \\ -\varepsilon(y - \sqrt{2\rho} + \varepsilon) + \frac{1}{2}\varepsilon^2, & y < \rho - \varepsilon, \end{cases} \qquad \xi = \frac{\rho}{\varepsilon\sqrt{2\rho} - \frac{1}{2}\varepsilon^2}$$

and set $\rho = 1$, $\varepsilon = 0.1$. The call to Algencan yields the results in Figure 1a.

As we can see, we did overall not recover the global solution in this case, however the computation was in fact shortened to 2.76 seconds for dimension 200, 7.95 seconds for dimension 300 and 12.73 seconds for dimension 400.

## 7.2 Sign-constrained Basis Pursuit

### 7.2.1 Preliminaries

In general, the aim is to find a sparse vector $x$ that approximately satisfies

$$Ax \approx b.$$

In problem settings as, for instance, mass spectrometry as described in [34], it is also necessary to introduce the sign constraints $x \geq 0$. We therefore arrive at the formulation

$$\min_x \ \|x\|_0 \quad \text{s.t.} \quad \|Ax - b\|_2^2 \leq \varepsilon, \ x \geq 0, \tag{27}$$

to which the penalty formulation with the choice of $p^\rho(y) = 1/2 \left\|y - \sqrt{2\rho}e\right\|_2^2$ is given by

$$\min_x \ p^\rho(y) + \alpha x^T y \quad \text{s.t.} \quad \|Ax - b\|_2^2 \leq \varepsilon, \ x \geq 0, \ y \geq 0. \tag{28}$$

Observe that (27) is a sparse optimization problem of the form $f(x) + \|x\|_0$ with $f \equiv 0$. These kind of problems are particularly challenging since a simple inspection shows that every feasible point is already a local minimum, which, of course, is also reflected by our stationarity

conditions. Nevertheless, we hope that an accumulation point $(x^*, y^*)$ of a sequence $(x^k, y^k)$ produced by the exact penalty approach is feasible and has a convincing sparsity pattern.

We will construct these problems by choosing a matrix $A$ and a suitable sparse vector $x^0 \geq 0$ such that

$$b = Ax^0 + r, \quad \varepsilon = \|r\|_2^2 \cdot (1 + \delta)$$

with a random vector $r$ and a small $\delta > 0$. By continuity, the interior of the set given by the constraints in (28) is nonempty. Hence, the Slater-CQ is fulfilled and the penalized formulation always admits Lagrangian multipliers. Furthermore, let $x^*$ be given such that $I_0(x^0) = I_0(x^*)$. Consider then the TNLP for problem (27) around a feasible point $x^*$:

$$\min_x \ 0 \quad \text{s.t.} \quad \|Ax - b\|_2^2 \leq \varepsilon, \ x_i = 0, \ i \in I_0(x^*).$$

Clearly, the point $x^0$ is also feasible for the above problem and, furthermore, strictly satisfies the inequality constraints. Hence, the Slater-CQ is also satisfied for the TNLP and as an inference $x^*$ is an S-stationary point.

### 7.2.2 Numerical Tests

We tested 200 instances in which we initialized $A$ as a random $\{0, ..., 99\}^{128\times512}$ matrix and chose an original signal $\overline{x}$, with $\overline{x}_i$ identically distributed on the interval $[0, 1]$. Afterwards, a support of size 16 was taken at random, so that

$$\|\overline{x}\|_0 = 16.$$

The vector $b := A\overline{x}$ was distorted by a Gaussian noise $r \in \mathbb{R}^{128}$ of mean 0 and variance 0.5. We chose $\varepsilon$ such that

$$\varepsilon > 1.1 \cdot \max_{i=1,...,200} (0.5 \cdot \|r_i\|_2)^2,$$

where $r_i$ denotes the error vector for instance $i$. Furthermore, $x^0$ was initialized as the zero vector, $\rho = 1$, $y^0 = e$, $\alpha^0 = 1.0$ and $\beta = 1.1$. Stopping parameters for Algencan where chosen as before, with the tolerance for approximate stationarity of $10^{-6}$ and for feasibility of $10^{-8}$. The corresponding numerical results are presented in Figure 2. Note, in particular, that the sparsity level generated by our method is, for all instances, at least as good as the initial guess $\bar{x}$, and even better for a number of test problems.

## 7.3 Logistic Regression

### 7.3.1 Problem Definition

Assume one is interested to train a decision-making algorithm based on probabilities, where we have $m$ data points $(z_i, t_i)$, with $z_i \in \mathbb{R}^n$, $t_i \in \{1, -1\}^n$, and are looking for a model $p$ with parameters $a = (a_1, ..., a_n)$ such that

$$p(a; z_i) \approx t_i \quad \forall i = 1, \ldots, m.$$

As a common approach, one chooses the sigmoid function

$$p(a; z_i) = \frac{1}{1 + \exp(-a^T z_i)}$$

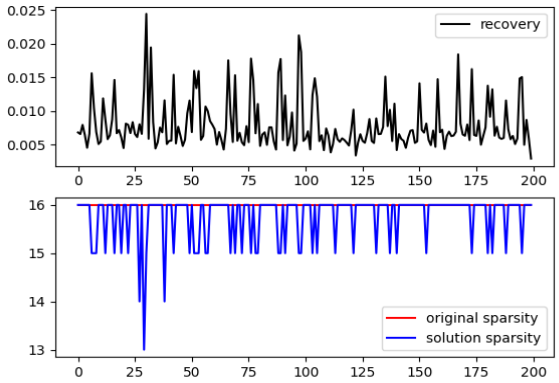



**Figure 2:** Measurement of the recovery $\|\overline{x} - x^s\|$ and comparison of the sparsity between the original vector $\overline{x}$ and sparsity of the solution $x^s$.

in order to find a suitable and sparse parameter vector $a$ by solving the optimization problem

$$\min_a \frac{1}{m} \sum_{i=1}^{m} \log(1 + \exp(-y_i z_i^T a)) + \rho \|a\|_0 \quad \text{s.t.} \quad -r \leq a \leq r$$

for some large $r$ to guarantee the solvability of the problem. Note that, here we are lacking the sign constraints on $a$. To deal with this problem, we separate $a$ into its positive and negative parts in the sense that

$$a = a^+ - a^-, \quad a^+ \geq 0, \ a^- \geq 0,$$

so that our newly found optimization problem is of the form

$$\min_{a^+,\ a^-} \ f(a^+ - a^-) + \rho \left\|(a^+, a^-)\right\|_0, \quad r \geq a^+ \geq 0, \ r \geq a^- \geq 0.$$

In general, this approach comes with some drawbacks as the split is not unique and increases the number of local minima. Consider for instance the problem

$$\min_x \ x^2 + \|x\|_0, \quad x \in \mathbb{R}.$$

Then, clearly, $x^* = 0$ is the only local and global minimum. If, however, we introduce the split, we obtain the formulation

$$\min_{x^+,\ x^-} \ (x^+ - x^-)^2 + \left\|(x^+, x^-)\right\|_0, \quad x^+ \geq 0, \ x^- \geq 0,$$

with $(x^+, x^-) = (0, 0)$ still being the unique global minimum, but with each $(x^+, x^-) = \lambda \cdot (1, 1)$ being a local minimum for each $\lambda \in \mathbb{R}$.

Unfortunately, this is not the only problem as we naturally increase the number of variables and, thus, also the required computational power. Nevertheless, the following subsection shows that this approach still works quite well in practice.

### 7.3.2 Numerical Tests

Since there were no constraints in place, computation where carried out via an implementation of a spectral gradient type method [8] under Python. We first tested our approach with the

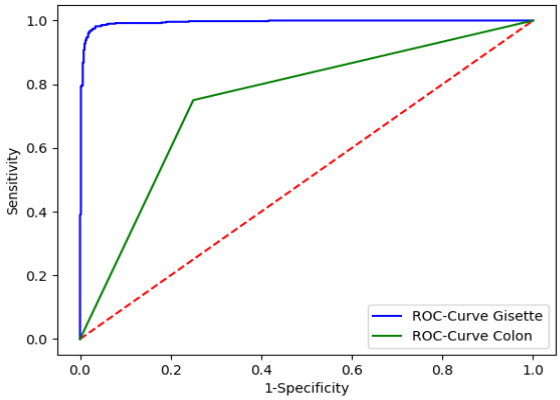



**Figure 3:** ROC-Curve for the gisette and the colon data set.

widely known colon-cancer data set[3] with 2000 features and 62 data samples. We initiated a training set by choosing 42 data samples, consisting of 14 positive and 28 negative labels. As starting parameters we chose the penalty function $p_i^\rho(y_i) = 1/2(y_i - \sqrt{2\rho})^2$ and set $(x^0, y^0) = (0, \sqrt{2\rho}e)$ as the initial guess, where $\rho = 0.1 \cdot \frac{1}{m}$. Furthermore, we set $\alpha_0 = 0.1$, $\beta = 10$. The spectral gradient step was executed for $10^4$ iterations or to a precision of $10^{-5}$, where we accepted the end result once complementarity between $x^k$ and $y^k$ was reached to a precision of $10^{-6}$. The accuracy measured to around 75% as integral of the ROC-curve, where $\|x\|_0 = 7$ from 2000 possible entries, whereas cpu time accrued to 2.35 seconds. In fact, our solution vector predicted 0 and 1 label to machine precision so that no matter a given threshold we would always correctly guess exactly 15 out of 20 possible instances in the validation set.

Second we chose as test example the gisette data set from the NIPS 2003 challenge[4]. The gisette data sets comes with specific training and validation data. Again, the initial guess was made with $(x^0, y^0) = (0, \sqrt{2\rho}e)$, where we used the same penalty function as before and set $\rho = 1/m$. The remaining parameters where chosen as $\alpha_0 = 1$, $\beta = 10$. The spectral gradient step was executed for $10^4$ iterations or to a precision of $10^{-2}$, where we accepted $10^{-2}$ as tolerance for the complementarity constraints. The accuracy measured to around 99.45 % as integral of the ROC-curve, where $\|x\|_0 = 551$ of 5000 possible entries, whereas cpu time accrued to 2.75 minutes. If we accept 0.5 as the threshold to which we predict a 1 or 0 if a value larger or less than 0.5 was observed, we would correctly guess in 97.2% of instances in the validation set.

## 7.4 Support Vector Machines

### 7.4.1 Problem Definition

The support vector machine problem may be stated as

$$\min_{c,\gamma} \frac{1}{2m} \|c\|_2^2 + \rho \|\max\{0, e - z \circ (Ac - \gamma)\}\|_0$$

with some matrix $A \in \mathbb{R}^{m \times n}$, and a $z \in \{-1, 1\}^m$ signaling that the $i$-th sample $a_i^T$ given as $i$-th row of $A$ belongs to the class $z_i$. By introducing an auxiliary variable $u$, we may rewrite

[3] https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html
[4] https://archive.ics.uci.edu/

| Dataset | features | train | test | acc-pnl | acc-libsvm | cpu-pnl | cpu-libsvm |
|---|---|---|---|---|---|---|---|
| `arcene` | 10000 | 100 | 100 | 82% | 83% | 10.5 sec | 0.4 sec |
| `jcnn` | 22 | 49990 | 91701 | 91.8 % | 92.1 % | 105.7 sec | 11.1 sec |
| `a9a` | 123 | 23373 | 8141 | 84.9% | 85% | 145sec | 15 sec |
| `binary` | 47236 | 20242 | 677399 | 96.3% | 96.3 % | 14.7 sec | 90 sec |
| `kddb` | 1129522 | 19264097 | 748401 | 94.4% | - | 161.8 sec | - |

**Table 1:** Accuracy and CPU-time for the SVM tests.

this problem into

$$\min_{c,\gamma,u} \frac{1}{2m} \|c\|_2^2 + \rho \|u\|_0, \quad u \geq 0,\ u \geq e - z \circ (Ac - \gamma). \tag{29}$$

The corresponding penalized problem is given by

$$\min_{c,\gamma,u} \frac{1}{2m} \|c\|_2^2 + p^\rho(y) + \alpha y^T u, \quad \text{s.t.} \quad u \geq 0,\ u \geq e - z \circ (Ac - \gamma). \tag{30}$$

These subproblems are then solved by an augmented Lagrangian approach.

#### 7.4.2 Numerical Tests

We applied our method to a few selected datasets from the source[5]. The choice of the penalty function was again $p_i^\rho(y_i) = 1/2(y_i - \sqrt{2\rho})^2$. The additional scaling by factor $1/m$ in the target function was introduced to avoid large values during the gradient method which occured with large sample-size. We met the matching choice of $\rho = 1/m,\ \alpha_0 = 1/m$ and set $\beta = 10$. The value of $\delta$ in 6.1 was reduced to $10^{-2}$. We compared our results (denoted by 'pnl') to the libsvm solver available as python module[6]. The table 1 suggests that our penalty approach is of particular interest once the size of features and training variables becomes exceedingly large. In the last case, the call to the `svm_train` function within the libsvm package did not yield any result.

### 7.5 Dictionary Learning

#### 7.5.1 Problem Definition

The dictionary learning problem can be understood as an extension to the basis pursuit denoising type of problem, where also the basis is searched for. Let $Z \in \mathbb{R}^{n \times m}$ be given. We look for $D \in \mathbb{R}^{l \times n},\ C \in \mathbb{R}^{l \times m}$ which minimize

$$\min_{D,C} \frac{1}{2} \left\|Z - D^T C\right\|_F^2 + \rho \|C\|_0, \quad \text{s.t.} \quad \left\|D_j^T\right\|_2^2 \leq 1 \quad \forall j = 1, ..., l,$$

where $D_j^T$ denote the rows in $D$, $\|\cdot\|_F$ is the Frobenius norm and we define

$$\|C\|_0 := \sum_{i,j=1}^{n,m} \|C_{i,j}\|_0 .$$

[5] https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html
[6] https://pypi.org/project/libsvm/

As with the logistic regression example seen before, there are no sign constraints with $C$. We again pass to the shift

$$C = C_+ - C_-, \quad C_+ \geq 0,\ C_- \geq 0.$$

Note that there are no constraints with respect to $C$, the variable we want to be sparse. Furthermore, the feasible set $X$ is, in particular, closed and convex and as such has a unique projection, which is, in this case, also easy to compute. Let

$$F(C_+, C_-, D) = \frac{1}{2} \left\| Z - D^T(C_+ - C_-) \right\|_F^2 .$$

We therefore require in step 2 of algorithm 6.1

$$\left\| P_X(D^{k+1} - \nabla_D F(C_+^{k+1}, C_-^{k+1}, D^{k+1})) - D^{k+1} \right\| \leq \varepsilon_k.$$

By passing to the limit every accumulation point is already stationary with respect to component $D$. It seems natural to simply apply a projected spectral gradient method to this type of problem. Notice that, in particular, the derivative with respect to $C$ and $D$ is given by

$$\nabla_{C_\pm} F(C_+, C_-, D) = \pm \left( DD^T C - DZ \right), \quad \nabla_D F(C_+, C_-, D) = CC^T D - CZ^T.$$

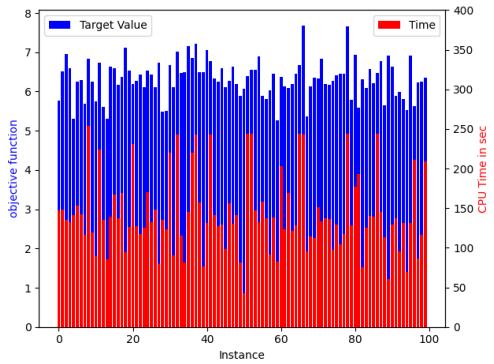
For the choice $(C_+, C_-, D) = (0, 0, 0)$ the above expressions both vanish and we clearly have an S-stationary point. We therefore initalized $(C_+^0, C_-^0, D^0)$ as a random matrix $\mathbb{R}^{l \times 2m+n}$ with entries taken from a standard normal distribution, projected onto $X$.

### 7.5.2 Numerical Tests

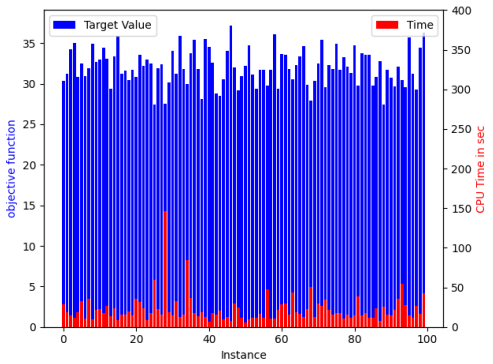
We conducted our numerical tests similar as in [24]. In 100 instances with $n = 10$, $l = 20$, $m = 30$ we generated $Z = C^T D$ from primary matrices $C$ and $D$, where $C$ only had three nonzero entries at random positions per column, where the values where taken from a standard normal distribution and where $D$ was chosen as a random standard normal matrix with normalized rows. We again chose the penalty function $p_i^\rho(y_i) = 1/2(y_i - \sqrt{2\rho})^2$ and set $\rho = 0.1$, $\alpha_0 = 0.1$, $\beta = 10$. The spectral gradient method was run for $10^4$ iterations with a tolerance of $10^{-3}$. Complementarity between $C_\pm$ and $Y_\pm$ was accepted to a tolerance of $10^{-6}$. We detail the achieved values as well as the required cpu time in the chart (4a). We measure the results against a proximal gradient type method applied to the same test set (4b) (note the different scaling of the axes regarding the objective function values). The proximal gradient method was mirrored from aformentioned source [24] using the averaging stepsize as detailed there. While computation time with the proximal gradient is lower compared to our exact penalty approach, we are able to significantly improve upon the reached target value.

The second test regarding dictionary-learning-type problems was done with the MNIST[7] data set. We have chosen the first 100 images and used and tried to find a sparse representation $Y_i \approx C_i^T D_i$ for each of the images $Y_i$. As $Y_i$ was represented by a 28 by 28 matrix, we let $C_i,\ D_i \in \mathbb{R}^{l \times 28}$ for $l = 6, 8, 10, 12, 14$ and survey the interesting characteristics in the Table 5a as an average over the 100 test runs. To get an idea for the quality of the achieved decomposition, we compare for $i = 1, ..., 18$ the original image $Y_i$ to the result $C_i^T D_i$ specifically for dimension $l = 10$ in figure (5b).

[7] Y. LeCun and C. Cortes. Mnist handwritten digit database. AT&T Labs [Online]. Available: https://yann.lecun.com/exdb/mnist, 2010




**(a)** Computation time and achieved target value via the exact penalty approach.

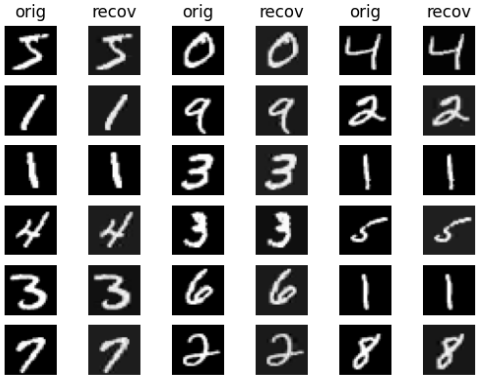



**(b)** Computation time and achieved target value via a proximal gradient type method.

**Figure 4:** A comparison of the penalty and proixmal gradient method applied to the Dictionary Learning problem

| dim | $f + .1\,\|\cdot\|_0$ | $\|\cdot\|_0$ | time |
|---|---|---|---|
| $l = 6$ | 1.2 | 23.3 | 16.7 |
| $l = 8$ | 0.6 | 21.8 | 16.3 |
| $l = 10$ | 0.3 | 19.9 | 14.3 |
| $l = 12$ | 0.2 | 17.6 | 12.1 |
| $l = 14$ | 0.09 | 16.3 | 10 |

**(a)** Comparison of the MNIST Dataset for 100 instances.




**(b)** Comparison of the original images (odd columns) to the recovered images (even columns)

**Figure 5:** The penalty method applied to MNIST in order to find sparse representation of images.

# 8 Final Remarks

This paper introduces a class of reformulations of the $\ell_0$-sparse optimization problem and develops suitable constraint qualifications as well as corresponding first- and second-order optimality conditions. The results are then used to apply an exact penalty-type method for the solution of the $\ell_0$-sparse optimization problem which is particularly useful if the contraints include nonnegativity conditions on the variables. Otherwise, one has to use a split of the free variables which might introduce additional local minima. Though the corresponding numerical results are still very promising, in this situation, it might be more favourable to apply another technique based on our reformulation which can be applied also in the case where there exist free variables. One natural possibility is the augmented Lagrangian method, and a closer look at this technique will therefore be part of our future research.

## Disclosure

There are no competing interests to report.

## References


[1] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. Silva. A relaxed constant positive linear dependence constraint qualification and applications. *Mathematical Programming*, 135(1-2):255–273, 2012.

[2] R. Andreani, J. M. Martínez, A. Ramos, and P. J. S. Silva. A cone-continuity constraint qualification and algorithmic consequences. *SIAM Journal on Optimization*, 26(1):96–110, 2016. doi:10.1137/15M1008488.

[3] R. Andreani, J. M. Martínez, A. Ramos, and P. J. S. Silva. Strict constraint qualifications and sequential optimality conditions for constrained optimization. *Mathematics of Operations Research*, 43(3):693–717, 2018. doi:10.1287/moor.2017.0879.

[4] A. Beck. *First-Order Methods in Optimization*. SIAM, 2017.

[5] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, MA, 1999.

[6] D. Bienstock. Computational study of a family of mixed-integer quadratic programming problems. *Mathematical programming*, 74:121–140, 1996.

[7] E. G. Birgin and J. M. Martínez. *Practical Augmented Lagrangian Methods for Constrained Optimization*. SIAM, Philadelphia, 2014. doi:10.1137/1.9781611973365.

[8] E. G. Birgin, J. M. Martínez, and M. Raydan. Nonmonotone spectral projected gradient methods on convex sets. *SIAM Journal on Optimization*, 10(4):1196–1211, 2000. doi:10.1137/s1052623497330963.

[9] O. P. Burdakov, C. Kanzow, and A. Schwartz. Mathematical programs with cardinality constraints: reformulation by complementarity-type conditions and a regularization method. *SIAM Journal on Optimization*, 26(1):397–425, 2016. doi:10.1137/140978077.

[10] M. Červinka, C. Kanzow, and A. Schwartz. Constraint qualifications and optimality conditions for optimization problems with cardinality constraints. *Mathematical Programming*, 160(1):353–377, 2016. doi:10.1007/s10107-016-0986-6.

[11] X. Chen, L. Guo, Z. Lu, and J. J. Ye. An augmented Lagrangian method for non-Lipschitz nonconvex programming. *SIAM Journal on Numerical Analysis*, 55(1):168–193, 2017. doi:10.1137/15M1052834.

[12] A. De Marchi, X. Jia, C. Kanzow, and P. Mehlitz. Constrained composite optimization and augmented lagrangian methods. *Mathematical Programming*, 201(1–2):863–896, Feb. 2023. doi:10.1007/s10107-022-01922-4.

[13] M. Durea and R. Strugariu. *An Introduction to Nonlinear Optimization Theory*. De Gruyter, Dec. 2014. doi:10.2478/9783110426045.

[14] J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360, 2001.

[15] M. Feng, J. E. Mitchell, J.-S. Pang, X. Shen, and A. Wächter. Complementarity formulations of $\ell_0$-norm optimization problems. *Pacific Journal of Optimization*, 14(2):273 – 305, 2018.

[16] J.-y. Gotoh, A. Takeda, and K. Tono. Dc formulations and algorithms for sparse optimization problems. *Mathematical Programming*, 169(1):141–176, July 2017. doi:10.1007/s10107-017-1181-0.

[17] C. Kanzow, P. Mehlitz, and D. Steck. Relaxation schemes for mathematical programmes with switching constraints. *Optimization Methods and Software*, 36(6):1223–1258, 2021. doi:10.1080/10556788.2019.1663425.

[18] C. Kanzow, A. Schwartz, and F. Weiß. The sparse(st) optimization problem: Reformulations, optimality, stationarity, and numerical results. *arXiv preprint arXiv:2210.09589*, 2022.

[19] H. Y. Le. Generalized subdifferentials of the rank function. *Optimization Letters*, 7(4):731–743, Feb. 2012. doi:10.1007/s11590-012-0456-x.

[20] H. A. Le Thi, T. P. Dinh, H. M. Le, and X. T. Vo. DC approximation approaches for sparse optimization. *European Journal of Operational Research*, 244(1):26–46, 2015.

[21] Y.-C. Liang and J. J. Ye. Optimality conditions and exact penalty for mathematical programs with switching constraints. *Journal of Optimization Theory and Applications*, 190(1):1–31, June 2021. doi:10.1007/s10957-021-01879-y.

[22] Z. Lu and Y. Zhang. Sparse approximation via penalty decomposition methods. *SIAM Journal on Optimization*, 23(4):2448–2478, Jan. 2013. doi:10.1137/100808071.

[23] Z.-Q. Luo, J.-S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, Cambridge, 1996. doi:10.1017/cbo9780511983658.

[24] A. D. Marchi. Proximal gradient methods beyond monotony. *Journal of Nonsmooth Analysis and Optimization*, Volume 4(Original research articles), June 2023. doi:10.46298/jnsao-2023-10290.

[25] P. Mehlitz. Asymptotic stationarity and regularity for nonsmooth optimization problems. *Journal of Nonsmooth Analysis and Optimization*, 1:6575, 2020. doi:10.46298/jnsao-2020-6575.

[26] P. Mehlitz. Stationarity conditions and constraint qualifications for mathematical programs with switching constraints. *Mathematical Programming*, 181(1):149–186, 2020. doi:10.1007/s10107-019-01380-5.

[27] B. S. Mordukhovich. *Variational Analysis and Applications*. Springer, Cham, 2018. doi:10.1007/978-3-319-92775-6.

[28] T. B. T. Nguyen, H. A. L. Thi, H. M. Le, and X. T. Vo. DC approximation approach for $\ell_0$-minimization in compressed sensing. In *Advanced Computational Methods for Knowledge Engineering: Proceedings of 3rd International Conference on Computer Science, Applied Mathematics and Applications-ICCSAMA 2015*, pages 37–48. Springer, 2015.

[29] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 1999.

[30] D. Ralph and S. J. Wright. Some properties of regularization and penalization schemes for mpecs. *Optimization Methods and Software*, 19(5):527–556, Oct. 2004. doi:10.1080/10556780410001709439.

[31] A. A. Ribeiro, M. Sachine, and E. H. Krulikovski. A comparative study of sequential optimality conditions for mathematical programs with cardinality constraints. *Journal of Optimization Theory and Applications*, 192(3):1067–1083, 2022.

[32] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317. Springer Science & Business Media, Berlin, 2009. doi:10.1007/978-3-642-02431-3.

[33] A. M. Tillmann, D. Bienstock, A. Lodi, and A. Schwartz. Cardinality minimization, constraints, and regularization: a survey. *arXiv preprint arXiv:2106.09606*, 2021.

[34] E. van den Berg and M. P. Friedlander. Sparse optimization with least-squares constraints. *SIAM Journal on Optimization*, 21(4):1201–1229, Oct. 2011. doi:10.1137/100785028.

[35] C.-H. Zhang. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2), Apr. 2010. doi:10.1214/09-aos729.

[36] S. Zhang and J. Xin. Minimization of transformed l 1 penalty: theory, difference of convex function algorithm, and robust application in compressed sensing. *Mathematical Programming*, 169(1):307–336, 2018.