A General Penalty-Method and a General Regularization-Method for Cardinality-Constrained Optimization Problems

Felix P. Broesamle[1*]

[1] Institute for Operations Research, Karlsruhe Institute of Technology, Karlsruhe, 76131, Germany .

Abstract

We consider cardinality-constrained optimization problems (CCOPs), which are general nonlinear programs with an additional constraint limiting the number of nonzero continuous variables. The continuous reformulation of CCOPs involves complementarity constraints, which pose significant theoretical and computational challenges. To address these difficulties, we propose and analyze two numerical solution approaches: a general penalty method and a general regularization method. Both approaches generate a sequence of easier to solve problems, and we show convergence of the corresponding KKT points against an M-stationary point under CC-MFCQ. Both methods rely on structural properties of the penalty and regularization functions, which we introduce and illustrate with examples. Finally, we present comprehensive numerical experiments to assess the practical performance of the proposed methods and to compare them with established approaches.

Keywords: Cardinality constraint, Penalty method, Regularization method, Relaxation, M-stationarity, Convergence

1 Introduction

In this paper, we consider general nonlinear optimization problems with an additional cardinality constraint, refered to as cardinality constrained optimization problems (CCOPs). The cardinality constraint restricts the number of nonzero variables in a solution. In other words, the cardinality of the support of feasible points is constrained.

[*] felix.broesamle@kit.edu

Cardinality constrained optimization problems (CCOPs) arise in a wide range of applications. Prominent examples include feature selection or sparse principal component analysis in machine learning [1, 2], image and signal processing, as well as portfolio optimization, in particular sparse portfolio selection [3]. For a comprehensive overview of applications and related problem classes, we refer to [4].

We now introduce notation used throughout the paper. Let $e \in \mathbb{R}^n$ denote the vector of all ones and $e_i \in \mathbb{R}^n$ the i-th canonical unit vector. For $x \in \mathbb{R}^n$, the ℓ_0-norm (0-norm) is defined as

$$\|x\|_0 \coloneqq |\text{supp}(x)|,$$

where

$$\text{supp}(x) \coloneqq \{i \in \{1, \dots, n\} : x_i \neq 0\}$$

denotes the support of x. A general cardinality constrained optimization problem is given by [5]

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0, \quad \|x\|_0 \leq S, \tag{CCOP}$$

where $S \in \mathbb{N}$ and $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ are assumed to be at least continuously differentiable. Throughout this work, we assume $\mathbb{N} \ni S < n$, since otherwise the cardinality constraint $\|x\|_0 \leq S$ is superfluent.

Early work on this class of problems includes [6], where cardinality constraints were studied in the context of sparse portfolio optimization. In recent years, CCOPs have attracted growing attention. This is due both to their relevance in applications such as feature selection in machine learning and to the introduction of a continuous reformulation in [5, 7], along with the development of a specially tailored optimality theory.

This optimality theory is closely related to that of *Mathematical Programs with Complementarity Constraints (MPCCs)* [8, 9] and *Mathematical Programs with Vanishing Constraints (MPVCs)* [10, 11]. Several numerical methods have been proposed for the continuous reformulation of CCOPs that build on established MPCC solution methods; see, for instance, [5, 12]. The regularization methods for the continuous reformulation presented in [5, 12] are respectively closely related to the regularization methods for MPCCs presented in [13, 14]. The specially tailored solution methods for the continuous reformulation often yield high-quality solutions. A comprehensive optimality theory for the continuous reformulation is now available; see, e.g., [15–17]. In addition to MPCC-based solution methods, specially tailored solution methods have also been developed; see, e.g., [18].

Finally, we note that a penalty method for solving the continuous reformulation was proposed in [19]. However, this approach was restricted by the additional sign restriction $x \geq 0$, which allowed the use of the ℓ_1-penalty term.

The structure of this paper is as follows. In Section 2, we introduce fundamental concepts in nonlinear optimization, and the continuous reformulation of CCOPs. Section

3 presents a penalty method for the continuous reformulation, inspired by penalty-based approaches developed for MPCCs [9, 20]. We characterize structural properties of the penalty term that allow convergence of the KKT points to an M-stationary point under CC-MFCQ. In Section 4, we propose a general regularization method, constructed using the same structural properties. Under CC-MFCQ, convergence of the KKT points to an M-stationary point is again established. Finally, in Section 5, we present comprehensive numerical experiments comparing the newly introduced solution methods with established approaches. The numerical experiments will show that the general penalty method introduced in Section 3 is highly robust and performs exceptionally well.

2 Preliminaries

In this section, we briefly review fundamental definitions and concepts from nonlinear programming and the continuous reformulation of CCOPs.

2.1 Nonlinear Programming

We consider a general nonlinear program, given by

$$\min_{x\in\mathbb{R}^n} \quad f(x) \quad \text{s.t.} \quad g_i(x) \le 0, \qquad \text{for all } i = 1,\dots,m, \qquad h_i(x) = 0, \qquad \text{for all } i = 1,\dots,p, \tag{NLP}$$

with all functions $f, g_i, h_i : \mathbb{R}^n \to \mathbb{R}$, at least continuously differentiable.

Definition 2.1 A point $x^* \in \mathbb{R}^n$ is called *Karush-Kuhn-Tucker point (KKT point)* of the optimization problem (NLP), if there exist $\lambda_1,\dots,\lambda_m,\mu_1,\dots,\mu_p \in \mathbb{R}$, such that the *Karush-Kuhn-Tucker conditions (KKT conditions)* hold:

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla g_i(x^*) + \sum_{i=1}^p \mu_i \nabla h_i(x^*) = 0,$$

$$g_i(x^*) \le 0, \quad \forall i = 1,\dots,m \quad \text{and} \quad h_i(x^*) = 0, \quad \forall i = 1,\dots,p,$$

$$\lambda_i \ge 0, \quad \lambda_i g_i(x^*) = 0 \quad \forall i = 1,\dots,m.$$

We call the corresponding multipliers $\lambda_1,\dots,\lambda_m,\mu_1,\dots,\mu_p \in \mathbb{R}$ to x^* *Lagrange multipliers of the KKT point* x^*.

Definition 2.2 ([21], Definition 2.1) A set of vectors $\{a_i : i \in I_1\} \cup \{b_i : i \in I_2\}$ is said to be *positive-linearly dependent* if there exist $\alpha \in \mathbb{R}^{|I_1|}, \beta \in \mathbb{R}^{|I_2|}, (\alpha,\beta) \ne 0$ with $\alpha \ge 0$ and

$$\sum_{i\in I_1} \alpha_i a_i + \sum_{i\in I_2} \beta_i b_i = 0.$$

Otherwise, we say that these vectors are *positive-linearly independent.*

Given a local minimum x^* of (NLP) sucht that certain conditions are satisfied at x^*, it is possible to show that x^* is also a KKT point. These conditions are called *constraint qualifications (CQ)*. The following *CQ* is in the following relevant.

Definition 2.3 Let x^* be feasible for the optimization problem (NLP). Then we say that x^* satisfies the *Mangasarian-Fromovitz Constraint Qualification (MFCQ)* if the following two conditions hold:

1. There exists $d \in \mathbb{R}^n$ such that

$$\nabla g_i(x^*)^T d < 0, \quad \forall\, i \in I(x^*), \qquad \text{and} \qquad \nabla h_i(x^*)^T d = 0, \quad \forall\, i = 1, \dots, p,$$

   where $I(x^*) \coloneqq \{i \in \{1, \dots, m\} : g_i(x^*) = 0\}$ denotes the *active set at* x^*.
2. The gradients $\nabla h_1(x^*), \dots, \nabla h_p(x^*)$ are linearly independent.

Definition 2.4 (NCP function ([22], Definition 1)) A function $\phi : \mathbb{R}^2 \to \mathbb{R}$ is called a *nonlinear complementarity problem function (NCP function)* if it satisfies the complementarity condition

$$\phi(a, b) = 0 \quad \Longleftrightarrow \quad a \geq 0, \quad ab = 0, \quad b \geq 0.$$

An overview of the properties and construction of NCP functions, together with an extensive collection of NCP functions, is provided in [22]. Moreover, NCP functions can be used to reformulate *nonlinear complementarity problems* as systems of equations [23].

Corollary 2.5 *The following functions are NCP functions:*

1. *The* Fischer–Burmeister function *[24], defined by* $\phi_{FB}(a, b) = \sqrt{a^2 + b^2} - a - b$.
2. *The* Kanzow–Kleinmichel function *[25], for* $\lambda \in (0, 4)$*, defined by* $\phi_{KKM(\lambda)}(a, b) = \sqrt{(a - b)^2 + \lambda ab} - a - b$. *For the special case of* $\lambda = 2$*, it holds* $\phi_{KKM(2)}(a, b) = \phi_{FB}(a, b)$.

2.2 A continuous reformulation of CCOP

In [5], the following mixed-integer-nonlinear formulation of (CCOP) was introduced

$$\begin{aligned}
\min_{x,y \in \mathbb{R}^n} \quad & f(x) \\
\text{s.t.} \quad & g(x) \leq 0, \qquad h(x) = 0, && \text{(CCOP-MINLP)} \\
& e^T y \geq n - S, \\
& x_i y_i = 0, && \text{for all } i = 1, \dots, n, \\
& y_i \in \{0, 1\}, && \text{for all } i = 1, \dots, n.
\end{aligned}$$

By standard relaxation of the binary variables, we obtain the continuous optimization problem

$$\begin{aligned}
\min_{x,y \in \mathbb{R}^n} \quad & f(x) \\
\text{s.t.} \quad & g(x) \leq 0, \qquad h(x) = 0, && \text{(CCOP-NLP)}
\end{aligned}$$

$$
\begin{aligned}
& e^T y \geq n - S, \\
& x_i y_i = 0, && \text{for all } i = 1, \ldots, n, \\
& 0 \leq y_i \leq 1, && \text{for all } i = 1, \ldots, n.
\end{aligned}
$$

The optimization problem (CCOP-NLP) is a continuous reformulation of (CCOP) [5]. Its theoretical properties, with regard to specially tailored stationarity concepts and constraint qualification, were first introduced in [5, 7] and have been further extended in subsequent works; see, e.g., [15–17].
One might consider treating (CCOP-NLP) as a nonlinear program and attempting to solve it with standard software. However, its feasible set is very complicated and violates most of the standard constraint qualifications typically required for standard nonlinear programming algorithms [5]. This difficulty mainly arises from the complementarity constraints $x_i y_i = 0$, for all $i = 1, \ldots, n$. Due to the complementarity

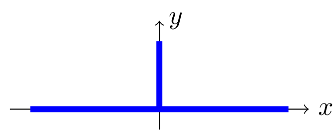


Fig. 1: $\{(x, y) \in \mathbb{R} \times \mathbb{R} : xy = 0, 0 \leq y \leq 1\}$

constraints, the continuous reformulation exhibits strong similarities to both classes of optimization problems, namely *Mathematical Programs with Complementarity Constraints (MPCCs)* [8, 9] and *Mathematical Programs with Vanishing Constraints (MPVCs)* [10, 11]. The generalized and specially tailored stationarity concepts and constraint qualifications for the continuous reformulation are closely related to those developed for MPCCs and MPVCs. However, (CCOP-NLP) can be viewed directly as an MPCC only under the additional assumption $x \geq 0$. A detailed comparison between the concepts introduced for MPCCs and those for (CCOP-NLP) under this assumption is provided in [7]. We introduce the following index sets:

$$
\begin{aligned}
I_g(x^*) &:= \{i \in \{1, \ldots, m\} : g_i(x^*) = 0\}, \\
I_0(x^*) &:= \{i \in \{1, \ldots, n\} : x_i^* = 0\}, \\
I_{\pm 0}(x^*, y^*) &:= \{i \in \{1, \ldots, n\} : x_i^* \neq 0, y_i^* = 0\}, \\
I_{00}(x^*, y^*) &:= \{i \in \{1, \ldots, n\} : x_i^* = 0, y_i^* = 0\}, \\
I_{0+}(x^*, y^*) &:= \{i \in \{1, \ldots, n\} : x_i^* = 0, y_i^* \in (0, 1)\}, \\
I_{01}(x^*, y^*) &:= \{i \in \{1, \ldots, n\} : x_i^* = 0, y_i^* = 1\}.
\end{aligned}
$$

In the presence of nonlinear constraints in (CCOP-NLP), the Guignard constraint qualfication (GCQ), the weakest standard constraint qualification, typically fails to hold [5]. This fundamental difficulty motivates the development of specially tailored stationarity concepts and constraint qualifications.

Definition 2.6 ([5], Definition 4.6) Let (x^*, y^*) be feasible for the continuous reformulation (CCOP-NLP). Then (x^*, y^*) is called the following:

a) *S-stationary* (S = Strong) if there exist multipliers $\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^p$, and $\gamma \in \mathbb{R}^n$ such that the following conditions hold:

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla h_i(x^*) + \sum_{i=1}^{n} \gamma_i e_i = 0,$$
$$\lambda_i \geq 0, \quad \lambda_i g_i(x^*) = 0, \quad \forall\, i = 1, \ldots, m,$$
$$\gamma_i = 0, \quad \forall\, i = 1, \ldots, n : y_i^* = 0.$$

b) *M-stationary* (M = Mordukovich) if there exist multipliers $\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^p$, and $\gamma \in \mathbb{R}^n$ such that the following conditions hold:

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla h_i(x^*) + \sum_{i=1}^{n} \gamma_i e_i = 0,$$
$$\lambda_i \geq 0, \quad \lambda_i g_i(x^*) = 0, \quad \forall\, i = 1, \ldots, m,$$
$$\gamma_i = 0, \quad \forall\, i = 1, \ldots, n : x_i^* \neq 0.$$

M-stationarity is a weaker condition than S-stationarity. For an M-stationary point, the multiplier $\gamma_i = 0$ is only enforced for i with $x_i^* \neq 0$. By feasibility of (x^*, y^*), it follows $y_i^* = 0$ for all i with $x_i^* \neq 0$. In contrast, S-stationarity requires $\gamma_i = 0$ for all i with $y_i^* = 0$, and therefore also for the so-called *biactive indices*, where $x_i^* = 0$ and $y_i^* = 0$.

Proposition 2.7 ([5], Proposition 4.8) *Let (x^*, y^*) be feasible for the continuous reformulation* (CCOP-NLP). *Then (x^*, y^*) is a KKT point of* (CCOP-NLP) *if and only if (x^*, y^*) is S-stationary for* (CCOP-NLP).

Since M-stationarity is a weaker condition than S-stationarity, it is also weaker than the standard KKT conditions. In fact, the M-stationarity conditions coincide exactly with the KKT conditions of the following *tightened nonlinear program* [5]:

$$\min_{x,y \in \mathbb{R}^n} \quad f(x) \qquad \text{s.t.} \quad g(x) \leq 0, \qquad h(x) = 0, \qquad \text{(TNLP}(x^*))$$
$$x_i = 0, \quad \text{for all } i \in I_0(x^*).$$

It follows that a local minimizer x^* of the original problem (CCOP) is also a local minimizer of (TNLP(x^*)), and thus an M-stationary point under suitable constraint qualifications [5].

Definition 2.8 (CC-CQ, [5]) Let (x^*, y^*) be feasible for (CCOP-NLP), and consider the corresponding tightened nonlinear program (TNLP(x^*)). We say that (x^*, y^*) satisfies the *CC-Constraint Qualification (Cardinality-Constrained Constraint Qualification, CC-CQ)* for the continuous reformulation (CCOP-NLP) if x^* satisfies the corresponding standard constraint qualification for (TNLP(x^*)).

Definition 2.9 ([7], Definition 3.11) Let (x^*, y^*) be feasible for the continuous reformulation (CCOP-NLP). Then (x^*, y^*) satisfies *CC-MFCQ* if the gradients

$$\{\nabla g_i(x^*) : i \in I_g(x^*)\} \cup \{\{\nabla h_i(x^*) : i \in \{1, \ldots, p\}\} \cup \{e_i : i \in I_0(x^*)\}\}$$

are positive-linearly independent.

Proposition 2.10 ([16], Proposition 2.3) *Let $(x^*, y^*) \in \mathbb{R}^n \times \mathbb{R}^n$ be feasible for* (CCOP-NLP). *If (x^*, y^*) is M-stationary, then there exists $z^* \in \mathbb{R}^n$ such that (x^*, z^*) is S-stationary.*

Numerically, this implies that any method which generates a sequence converging to an M-stationary point only, essentially gives an S-stationary point [16].

3 A general penalty method

The central idea is to address the challenging complementarity constraints

$$x_i y_i = 0, \quad \text{for all } i = 1, \ldots, n,$$

by omitting them from the constraints and penalizing their violation in the objective function via a suitable penalty term and a penalty parameter $\rho > 0$ with $\rho \uparrow \infty$. This approach leads to a sequence of relaxed penalty subproblems parametrized by ρ.
This approach has already been applied to MPCCs, e.g., in [20] and [9], and to (CCOP-NLP) under the assumption $x \geq 0$ in [19]. In this special case, the ℓ_1-norm can be used as the penalty term.

We consider the general case without a sign restriction and therefore introduce conditions that enable the construction of a suitable penalty term. The following conditions are motivated by those used in the construction of penalty terms for MPCCs, as proposed in [20].

Condition 3.1 *The function $\phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is at least continuously differentiable and satisfies the following conditions:*

A) *For all $x \in \mathbb{R}$ and $y \geq 0$:*

- *i) $\phi(x, y) = 0$ if and only if $xy = 0$.*
- *ii) $\phi(x, y) > 0$ if and only if $x \neq 0$ and $y > 0$.*

B) *Let $x \in \mathbb{R}$, and $y \geq 0$. Then it holds that:*

i)

$$\frac{\partial \phi(x,y)}{\partial x} = 0 \quad \text{if } y = 0, \quad \text{and} \quad \frac{\partial \phi(x,y)}{\partial y} = 0 \quad \text{if } x = 0,$$

ii)

$$\frac{\partial \phi(x,y)}{\partial x} \neq 0 \quad \text{and} \quad \frac{\partial \phi(x,y)}{\partial y} > 0 \quad \text{for all } (x,y) \in \mathbb{R} \times [0,1] \text{ with } x \neq 0, y > 0.$$

C) *Let $(x^k, y^k)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R} \times \mathbb{R}_{\geq 0}$ with $x^k \neq 0, y^k > 0$ for all $k \in \mathbb{N}$, and suppose that $(x^k, y^k) \to (x^*, y^*)$ as $k \to \infty$. If $x^* \neq 0, y^* = 0$, then*

$$\frac{\frac{\partial \phi(x^k,y^k)}{\partial x}}{\frac{\partial \phi(x^k,y^k)}{\partial y}} \overset{k\to\infty}{\longrightarrow} 0.$$

Conversely, if $x^* = 0, y^* > 0$, then

$$\frac{\frac{\partial \phi(x^k,y^k)}{\partial y}}{\frac{\partial \phi(x^k,y^k)}{\partial x}} \overset{k\to\infty}{\longrightarrow} 0.$$

From now on, we assume that $\phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is at least continuously differentiable and satisfies Condition 3.1.
The subproblems for the penalty method for (CCOP-NLP), parameterized by $\rho > 0$, are given by

$$\begin{aligned} \min_{x,y\in\mathbb{R}^n} \quad & f(x) + \rho\Phi(x,y) = f(x) + \rho \sum_{i=1}^{n} \phi(x_i, y_i) && \text{(CCOP-PEN}(\rho)\text{)} \\ \text{s.t.} \quad & g(x) \leq 0, \qquad h(x) = 0, \\ & e^T y \geq n - S, \\ & 0 \leq y_i \leq 1, && \text{for all } i = 1, \ldots, n. \end{aligned}$$

Condition 3.1 **A)** ensures that the penalty term $\Phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, defined by $\Phi(x,y) = \sum_{i=1}^n \phi(x_i, y_i)$, is strictly positive at feasible points of (CCOP-PEN(ρ)) that violate the complementarity constraints, and equals 0 whenever the complementarity constraints are satisfied. Hence, Condition 3.1 **A)** guarantees the fundamental property of a penalty term.

3.1 Penalty term

In this section, we present continuously differentiable functions that satisfy Condition 3.1.

Theorem 3.2 *The function* $\phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$*, defined by*

$$\phi(x, y) = x^2 y^2,$$

is continuously differentiable and satisfies Condition 3.1.

Proof We omit the proof, as the function is clearly continuously differentiable, and the properties of Condition 3.1 can be verified by a straightforward computation. □

The following observation motivates the construction of the function $\phi(x, y)$ in the subsequent theorem. Let $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be an NCP function. Then

$$g(|x|, y) = 0 \Longleftrightarrow y \geq 0, xy = 0,$$

so that the constraints $y_i \geq 0, x_i y_i = 0$ in (CCOP-NLP) can equivalently be replaced by $g(|x_i|, y_i) = 0$ for $i = 1, \dots, n$. Our goal is to construct a continuously differentiable function $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that $g(x, y) \geq 0$ for $x \in \mathbb{R}, y \geq 0$, and

$$g(x, y) = 0 \Longleftrightarrow xy = 0, x \in \mathbb{R}, y \geq 0, \tag{1}$$

which corresponds to Condition 3.1 **A)**.
For the Kanzow–Kleinmichel NCP function $\phi_{KKM(\lambda)}(x, y)$ with $\lambda \in (0, 4)$ from Corollary 2.5, it is known that $\phi_{KKM(\lambda)}(x, y)^2$ is continuously differentiable [25].
However, $\phi_{KKM(\lambda)}(|x|, y)^2$ is not continuously differentiable at points where $x = 0$ and $y \leq 0$. To address this, in the following theorem we define the function to be the constant 0-function for $y < 0$. This modification ensures that the resulting function is continuously differentiable and satisfies Condition 3.1.

Theorem 3.3 *The function* $\phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$*, based on the Kanzow–Kleinmichel NCP-function [25] with* $\lambda \in (0, 4)$*, defined by*

$$\phi(x, y) = \begin{cases} \left(-\sqrt{(|x| - y)^2 + \lambda |x| y} + |x| + y\right)^2, & \text{if } y \geq 0, \\ 0, & \text{if } y < 0, \end{cases}$$

is continuously differentiable and satisfies Condition 3.1.

Proof We omit the proof. The continuous differentiability of the function, as well as the properties of Condition 3.1 can be verified by a straightforward computation. □

3.2 Properties of the penalty subproblems

From now on, we assume that $\phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ in (CCOP-PEN(ρ)) is at least continuously differentiable and satisfies Condition 3.1. Additionally, we denote by Z the feasible set of (CCOP-NLP), and by Z_{PEN} the feasible set of (CCOP-PEN(ρ)) for any $\rho > 0$. Then, we have $Z = Z_{PEN} \cap \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n : x_i y_i = 0, \text{ for all } i = 1, \dots, n\}$.

The following result is inspired by [9, Theorem 5.1 (b)].

Theorem 3.4 *Let (x^*, y^*) be feasible for (CCOP-NLP) and satisfy CC-MFCQ for (CCOP-NLP). Then (x^*, y^*) is feasible for (CCOP-PEN(ρ)) and satisfies MFCQ for (CCOP-PEN(ρ)).*

Proof Let $(x^*, y^*) \in Z$. Since $Z \subseteq Z_{PEN}$, the point (x^*, y^*) is feasible for (CCOP-PEN(ρ)). Suppose that CC-MFCQ holds at (x^*, y^*). Then x^* satisfies MFCQ for the tightened nonlinear program (TNLP(x^*)), so there exists a $\tilde{d} \in \mathbb{R}^n$ such that

$$\nabla g_i(x^*)^T \tilde{d} < 0, \; i \in I_g(x^*), \tag{CC-MFCQ1}$$

$$\nabla h_i(x^*)^T \tilde{d} = 0, \; i = 1, \dots, p, \quad e_i^T \tilde{d}, \; i \in I_0(x^*), \tag{CC-MFCQ2}$$

and the set of gradients

$$\{\nabla h_1(x^*), \dots, \nabla h_p(x^*)\} \cup \{e_i : i \in I_0(x^*)\} \text{ is linearly independent.} \tag{CC-MFCQ3}$$

To establish MFCQ for (CCOP-PEN(ρ)) at (x^*, y^*), we seek a direction $d = (d_x, d_y) \in \mathbb{R}^n \times \mathbb{R}^n$ such that

$$\nabla g_i(x^*)^T d_x < 0, \; i \in I_g(x^*), \tag{PEN-MFCQ1}$$

$$-e^T d_y < 0, \text{ if } e^T y^* = n - S, \tag{PEN-MFCQ2}$$

$$-e_i^T d_y < 0, \; i \in I_{00}(x^*, y^*) \cup I_{\pm 0}(x^*, y^*), \quad e_i^T d_y < 0, \; i \in I_{01}(x^*, y^*), \tag{PEN-MFCQ3}$$

$$\nabla h_i(x^*)^T d_x = 0, \; i = 1, \dots, p, \tag{PEN-MFCQ4}$$

and the set of gradients

$$\{\nabla h_1(x^*), \dots, \nabla h_p(x^*)\} \text{ is linearly independent.} \tag{PEN-MFCQ5}$$

We directly obtain (PEN-MFCQ5) from (CC-MFCQ3). By setting $d_x = \tilde{d}$, we obtain (PEN-MFCQ1) and (PEN-MFCQ4) from (CC-MFCQ1) and (CC-MFCQ2), respectively.
If $-e^T y^* + n - S < 0$, the constraint on d_y corresponding to $e^T y = n - S$ is inactive. We define $d_y \in \mathbb{R}^n$ with an arbitrary $\sigma > 0$ by

$$(d_y)_i = \begin{cases} \sigma, & \text{if } y_i^* = 0, \\ -\sigma, & \text{if } y_i^* = 1, \\ 0, & \text{otherwise,} \end{cases} \qquad \text{for all } i = 1, \dots, n,$$

which ensures that (PEN-MFCQ3) is satisfied.
If $-e^T y^* + n - S = 0$, since $1 \le S < n$, it follows that $e^T y^* < n$. Consequently, there exists an index $j \notin I_{01}(x^*, y^*)$. We define $d_y \in \mathbb{R}^n$ with an arbitrary $\sigma > 0$ by

$$(d_y)_i = \begin{cases} n\sigma, & \text{if } y_i^* = 0, \\ -\sigma, & \text{if } y_i^* = 1, \\ n\sigma, & \text{otherwise,} \end{cases} \qquad \text{for all } i = 1, \dots, n.$$

This choice ensures that both (PEN-MFCQ2) and (PEN-MFCQ3) are satisfied. □

The following result is inspired by [9, Theorem 5.2].

Theorem 3.5 *Let (x^*, y^*) be feasible for (CCOP-NLP) and a KKT point of (CCOP-PEN(ρ)). Then (x^*, y^*) is S-stationary for (CCOP-NLP).*

Proof As (x^*, y^*) is feasible for (CCOP-NLP), we have $x_i^* y_i^* = 0$ for all $i = 1, \dots, n$. Let (x^*, y^*) be a KKT point. Then there exist corresponding Lagrange multipliers $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$, $\delta \in \mathbb{R}$, and $\nu \in \mathbb{R}^n$ satisfying

$$\nabla f(x^*) + \sum_{i=1}^{n} \rho \frac{\partial \phi(x_i^*, y_i^*)}{\partial x_i} e_i + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla h_i(x^*) = 0, \tag{PEN-KKT1}$$

$$-\delta e + \sum_{i=1}^{n} \rho \frac{\partial \phi(x_i^*, y_i^*)}{\partial y_i} e_i + \sum_{i=1}^{n} \nu_i e_i = 0, \tag{PEN-KKT2}$$

$$\lambda_i \begin{cases} \geq 0, & \text{if } g_i(x^*) = 0, \\ = 0, & \text{else,} \end{cases} \quad \text{for all } i = 1, \dots, n, \tag{PEN-KKT3}$$

$$\delta \begin{cases} \geq 0, & \text{if } e^T y^* = n - S, \\ = 0, & \text{else,} \end{cases} \tag{PEN-KKT4}$$

$$\nu_i \begin{cases} \leq 0, & \text{if } y_i^* = 0, \\ \geq 0, & \text{if } y_i^* = 1, \\ = 0, & \text{else,} \end{cases} \text{for all } i = 1, \dots, n. \tag{PEN-KKT5}$$

Here, ν_i denotes the Lagrange multiplier associated with the box-constraint $0 \leq y_i \leq 1$ for all $i = 1, \dots, n$. In order to show that (x^*, y^*) is S-stationary for (CCOP-NLP), we construct multipliers $\bar{\lambda} \in \mathbb{R}^m$, $\bar{\mu} \in \mathbb{R}^p$, and $\bar{\gamma} \in \mathbb{R}^n$ such that they satisfy

$$\nabla f(x^*) + \sum_{i=1}^{m} \bar{\lambda}_i \nabla g_i(x^*) + \sum_{i=1}^{p} \bar{\mu}_i \nabla h_i(x^*) + \sum_{i=1}^{n} \bar{\gamma}_i e_i = 0, \tag{S1}$$

$$\bar{\lambda}_i \begin{cases} \geq 0, & \text{if } g_i(x^*) = 0, \\ = 0, & \text{else,} \end{cases} \quad \text{for all } i = 1, \dots, m, \tag{S2}$$

$$\bar{\gamma}_i \begin{cases} = 0, & \text{if } y_i^* = 0, \\ \text{arbitrary}, & \text{else,} \end{cases} \quad \text{for all } i = 1, \dots, n. \tag{S3}$$

First, we set $\bar{\mu} = \mu$ and $\bar{\lambda} = \lambda$. Then (S2) follows directly from (PEN-KKT3). Next, we define $\bar{\gamma}_i = \rho \frac{\partial \phi(x_i^*, y_i^*)}{\partial x_i}$ for all $i = 1, \dots, n$, which yields (S1) from (PEN-KKT1). By Condition 3.1 **B.i)**, we have $\frac{\partial \phi(x_i^*, y_i^*)}{\partial x_i} = 0$, if $y_i^* = 0$. Therefore, $\bar{\gamma}_i = 0$ for all i with $y_i^* = 0$, and (S3) is satisfied. □

The following result is analogous to [19, Lemma 4.15].

Theorem 3.6 *Let (x^*, y^*) be a KKT point for (CCOP-PEN(ρ)), and let δ, the Lagrange multiplier corresponding to the constraint $e^T y \geq n - S$, satisfy $\delta = 0$. Then (x^*, y^*) is feasible for (CCOP-NLP).*

Proof By assumption, (x^*, y^*) is a KKT point for (CCOP-PEN(ρ)). Therefore, there exist Lagrange multipliers $\delta \in \mathbb{R}$, $\nu \in \mathbb{R}^n$ that satisfy

$$-\delta e + \sum_{i=1}^{n} \rho \frac{\partial \phi(x_i^*, y_i^*)}{\partial y_i} e_i + \sum_{i=1}^{n} \nu_i e_i = 0,$$

$$\delta \begin{cases} \geq 0, & \text{if } e^T y^* = n - S, \\ = 0, & \text{else}, \end{cases}$$

$$\nu_i \begin{cases} \leq 0, & \text{if } y_i^* = 0, \\ \geq 0, & \text{if } y_i^* = 1, \\ = 0, & \text{else}, \end{cases} \quad \text{for all } i = 1, \dots, n. \tag{2}$$

Here, ν_i denotes the Lagrange multiplier corresponding to the box-constraint $0 \leq y_i \leq 1$ for all $i = 1, \dots, n$.
By assumption, $\delta = 0$, which implies that, for all $i = 1, \dots, n$, we have

$$\rho \frac{\partial \phi(x_i^*, y_i^*)}{\partial y_i} + \nu_i = \delta = 0. \tag{3}$$

To show feasibility of (x^*, y^*) for (CCOP-NLP), it suffices to prove that $x_i^* y_i^* = 0$ for all $i = 1, \dots, n$. We proceed by contradiction.
Suppose there exists an index $j \in \{1, \dots, n\}$ such that $x_j^* \neq 0, y_j^* > 0$. Then, by Condition 3.1 **B.ii)**, we have $\frac{\partial \phi(x_j^*, y_j^*)}{\partial y_j} > 0$, and from (3) it follows that $\nu_j < 0$. However, (2) implies that $y_j^* = 0$, which is a contradiction. Hence, the claim follows. □

3.3 Convergence result

In the following, we consider a sequence of penalty subproblems (CCOP-PEN(ρ)) parameterized by $\rho > 0$, and study their behavior as $\rho \uparrow \infty$. The following result is motivated by the convergence theorems for the Scholtes-type regularization [12, Theorem 2] and penalty method for MPCCs [20, Theorem 2.1].

Theorem 3.7 *Let $(\rho^k)_{k \in \mathbb{N}}$ be a sequence with $\rho^k > 0$ for all $k \in \mathbb{N}$ and $\rho^k \uparrow \infty$ as $k \to \infty$. Let $(x^k, y^k)_{k \in \mathbb{N}}$ be a sequence of KKT points for (CCOP-PEN(ρ)) with $\rho = \rho^k$, and assume that $(x^k, y^k) \overset{k \to \infty}{\longrightarrow} (x^*, y^*)$, with (x^*, y^*) feasible for (CCOP-NLP). If the limit point (x^*, y^*) satisfies CC-MFCQ for (CCOP-NLP), then (x^*, y^*) is M-stationary for (CCOP-NLP).*

Proof By assumption, (x^k, y^k) is a KKT point for all $k \in \mathbb{N}$. Therefore, there exist Lagrange multipliers $(\lambda^k, \mu^k, \delta^k, \nu^k) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^n$ for all $k \in \mathbb{N}$ such that the conditions (PEN-KKT1) - (PEN-KKT5) are satisfied. We define for each $k \in \mathbb{N}$

$$\gamma_i^k := \rho^k \frac{\partial \phi(x_i^k, y_i^k)}{\partial x}, \qquad \text{for all } i = 1, \dots, n.$$

Together with (PEN-KKT1), we obtain for all $k \in \mathbb{N}$

$$\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^{n} \gamma_i^k e_i = 0. \tag{4}$$

We now show by contradiction the *boundedness* of the sequence $(\lambda^k, \mu^k, \gamma^k)_{k\in\mathbb{N}}$. Suppose, to the contrary, that $\|(\lambda^k, \mu^k, \gamma^k)\| \overset{k\to\infty}{\longrightarrow} \infty$. The normed sequence

$$\left(\frac{(\lambda^k, \mu^k, \gamma^k)}{\|(\lambda^k, \mu^k, \gamma^k)\|}\right)_{k\in\mathbb{N}}$$

is obviously bounded, and we can without loss of generality assume that the whole sequence converges

$$\lim_{k\to\infty} \frac{(\lambda^k, \mu^k, \gamma^k)}{\|(\lambda^k, \mu^k, \gamma^k)\|} = (\bar{\lambda}, \bar{\mu}, \bar{\gamma}) \neq 0.$$

By (PEN-KKT3), we have $\lambda_i^k \geq 0$ for all $i = 1, \dots, m$ and all $k \in \mathbb{N}$, which implies that $\bar{\lambda} \geq 0$. For all i such that $g_i(x^*) < 0$, continuity of g_i ensures that $g_i(x^k) < 0$ for sufficiently large k. Hence, by (PEN-KKT3), $\lambda_i^k = 0$ for all sufficiently large k, and thus $\bar{\lambda}_i = 0$. Consequently, it holds

$$\operatorname{supp}(\bar{\lambda}) \subseteq I_g(x^*). \tag{5}$$

We now show by contradiction that $\operatorname{supp}(\bar{\gamma}) \subseteq I_0(x^*)$. Suppose there exists an index $j \in \{1, \dots, n\}$ such that $\bar{\gamma}_j \neq 0$ and $x_j^* \neq 0$. Then, for sufficiently large k, we have $\gamma_j^k \neq 0$ and $x_j^k \neq 0$, which implies

$$\rho^k \frac{\partial \phi(x_j^k, y_j^k)}{\partial x} = \gamma_j^k \neq 0,$$

By Condition 3.1 **B)**, we have $y_j^k > 0$ for all sufficiently large k.

Since (x^*, y^*) is feasible for (CCOP-NLP), it follows that $y_j^* = 0$. Together, $y_j^k \overset{k\to\infty}{\longrightarrow} 0$ with $0 < y_j^k < 1$. From (PEN-KKT5), we then obtain $\nu_j^k = 0$ for sufficiently large k. Consequently, there exists a subsequence along which y_j^k is strictly monotone decreasing. Without loss of generality, we assume that y_j^k as strictly monotone decreasing along the entire sequence $(x^k, y^k)_{k\in\mathbb{N}}$.

From (PEN-KKT2), it follows that $\delta^k = \rho^k \frac{\partial \phi(x_j^k, y_j^k)}{\partial y} > 0$ for all sufficiently large $k \in \mathbb{N}$, under Condition 3.1 **B)**. Moreover, by (PEN-KKT4), it holds $e^T y^k = n - S$ for all sufficiently large k.

Since $y_j^k \downarrow 0$ as $k \to \infty$, maintaining the equality $e^T y^k = n - S$ requires the existence of at least one index m such that y_m^k is strictly monotone increasing along the sequence $(x^k, y^k)_{k\in\mathbb{N}}$. Hence, for sufficiently large k, we obtain $0 < y_m^k < 1$, which implies $y_m^* > 0, x_m^* = 0$ and $\nu_m^k = 0$.

From (PEN-KKT4), we have $\rho^k \frac{\partial \phi(x_m^k, y_m^k)}{\partial y} = \delta^k > 0$, which, by Condition 3.1 **B)**, implies $x_m^k \neq 0$ for sufficiently large k. Furthermore, Condition 3.1 **B)** ensures that $\gamma_m^k = \rho^k \frac{\partial \phi(x_m^k, y_m^k)}{\partial x} \neq 0$ for sufficiently large k.

Now, we have $(x_j^k, y_j^k) \overset{k\to\infty}{\longrightarrow} (x_j^*, y_j^*)$ with $x_j^* \neq 0, y_j^* = 0$, and $x_j^k \neq 0, y_j^k > 0$ for all $k \in \mathbb{N}$, as well as $(x_m^k, y_m^k) \overset{k\to\infty}{\longrightarrow} (x_m^*, y_m^*)$ with $x_m^* = 0, y_m^* > 0$, and $x_m^k \neq 0, y_m^k > 0$ for all $k \in \mathbb{N}$. Therefore, by applying Condition 3.1 **C)**, we obtain

$$\frac{\gamma_j^k}{\gamma_m^k} = \frac{\rho^k \frac{\partial \phi(x_j^k, y_j^k)}{\partial x}}{\rho^k \frac{\partial \phi(x_m^k, y_m^k)}{\partial x}} = \frac{\frac{\delta^k}{\frac{\partial \phi(x_j^k, y_j^k)}{\partial y}} \frac{\partial \phi(x_j^k, y_j^k)}{\partial x}}{\frac{\delta^k}{\frac{\partial \phi(x_m^k, y_m^k)}{\partial y}} \frac{\partial \phi(x_m^k, y_m^k)}{\partial x}} = \frac{\frac{\partial \phi(x_m^k, y_m^k)}{\partial y}}{\frac{\partial \phi(x_m^k, y_m^k)}{\partial x}} \cdot \frac{\frac{\partial \phi(x_j^k, y_j^k)}{\partial x}}{\frac{\partial \phi(x_j^k, y_j^k)}{\partial y}} \overset{k\to\infty}{\longrightarrow} 0.$$

This is a contradiction to

$$0 \neq |\bar{\gamma}_j| = \lim_{k\to\infty} \frac{|\gamma_j^k|}{\|(\lambda^k, \mu^k, \gamma^k)\|} \leq \lim_{k\to\infty} \frac{|\gamma_j^k|}{|\gamma_m^k|} = 0.$$

Therefore, we have

$$\operatorname{supp}(\bar{\gamma}) \subseteq I_0(x^*). \tag{6}$$

Dividing (PEN-KKT1) by $\|(\lambda^k, \mu^k, \gamma^k)\|$, using (5) and (6), and letting $k \to \infty$, we obtain from (4) for $k \to \infty$

$$\sum_{i\in I_g(x^*)} \bar{\lambda}_i \nabla g_i(x^*) + \sum_{i=1}^{p} \bar{\mu}_i \nabla h_i(x^*) + \sum_{i\in I_0(x^*)} \bar{\gamma}_i e_i = 0.$$

Since $\bar{\lambda} \geq 0$ and $(\bar{\lambda}, \bar{\mu}, \bar{\gamma}) \neq 0$, this leads to a contradiction with CC-MFCQ in (x^*, y^*). This would imply that the relevant gradients are positive linearly dependent, contradicting CC-MFCQ, which requires them to be positively linearly independent.
Therefore, the sequence $(\lambda^k, \mu^k, \gamma^k)_{k\in\mathbb{N}}$ is bounded. Without loss of generality, we assume that the sequence converges. Let $\lim_{k\to\infty}(\lambda^k, \mu^k, \gamma^k) = (\lambda, \mu, \gamma)$. Taking the limit in (4) as $k \to \infty$ yields

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla h_i(x^*) + \sum_{i=1}^{n} \gamma_i e_i = 0.$$

Analogously to above, we can show $\lambda \geq 0, \operatorname{supp}(\lambda) \subseteq I_g(x^*)$ and $\operatorname{supp}(\gamma) \subseteq I_0(x^*)$. Consequently, (x^*, y^*) is M-stationary for (CCOP-NLP) with multipliers (λ, μ, γ). □

Corollary 3.8 *Let $(\rho^k)_{k\in\mathbb{N}}$ be a sequence with $\rho^k > 0$ for all $k \in \mathbb{N}$ and $\rho^k \uparrow \infty$ for $k \to \infty$. Let $(x^k, y^k)_{k\in\mathbb{N}}$ be a sequence of KKT points for (CCOP-PEN(ρ)) with $\rho = \rho^k$, and assume that $(x^k, y^k) \overset{k\to\infty}{\longrightarrow} (x^*, y^*)$, with (x^*, y^*) feasible for (CCOP-NLP). If the limit point (x^*, y^*) satisfies CC-MFCQ for (CCOP-NLP), then (x^*, y^*) is M-stationary for (CCOP-NLP) and there exists a $z^* \in \mathbb{R}^n$ such that (x^*, z^*) is S-stationary for (CCOP-NLP).*

Proof By Theorem 3.7, this follows directly from Proposition 2.10. □

4 A general regularization method

Regularization approaches for the continuous reformulation (CCOP-NLP) are based on the same ideas as those developed for MPCCs.
The complementarity constraints $x_i y_i = 0, i = 1, \ldots, n$ in (CCOP-NLP) introduce substantial analytical and numerical challenges. To address these, the complementarity constraints are suitably replaced, resulting in regularized optimization problems with better analytical and numerical properties. The resulting regularized optimization problems depend on a parameter $t > 0$ and converge to the original optimization problem (CCOP-NLP) as $t \downarrow 0$.

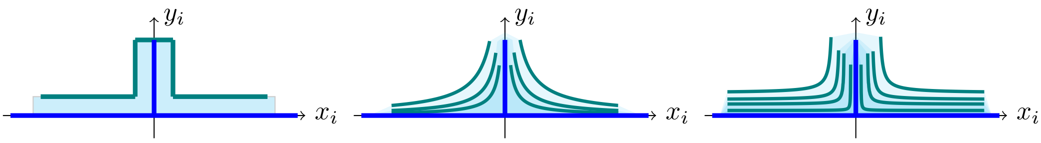
To relax the complementarity constraints $x_i y_i = 0, i = 1, \dots, n$, we consider a continuously differentiable function $\phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ that satisfies Condition 3.1. For a given parameter $t > 0$, the complementarity constraints are replaced by

$$\phi(x_i, y_i) \le t, \qquad \text{for all } i = 1, \dots, n.$$

This approach defines a general regularization method. The corresponding regularized optimization problems, parameterized by $t > 0$, are then given by

$$\begin{aligned} \min_{x,y \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & g(x) \le 0, \qquad h(x) = 0, \\ & -e^T y + n - S \le 0, \\ & \phi(x_i, y_i) \le t, && \text{for all } i = 1, \dots, n, \\ & 0 \le y_i \le 1, && \text{for all } i = 1, \dots, n. \end{aligned} \tag{REG(t)}$$

The following theorem is motivated by [12, Theorem 2].



(a) Kanzow–Schwartz (b) Scholtes-type (c) General

Fig. 2: Geometric interpretation of the regularization approaches: Kanzow–Schwartz regularization [5] (Figure 2a), Scholtes-type regularization [12] (Figure 2b), and general regularization (Figure 2c) where the function from Theorem 3.3 with $\lambda = 1$ is used

Theorem 4.1 *Let $(t^k)_{k\in\mathbb{N}}$ be a sequence with $t^k > 0$ for all $k \in \mathbb{N}$ and $t^k \downarrow 0$ as $k \to \infty$. Let $(x^k, y^k)_{k\in\mathbb{N}}$ be a sequence of KKT points of (REG(t)) with $t = t^k$, converging to (x^*, y^*). If CC-MFCQ holds at (x^*, y^*), then (x^*, y^*) is an M-stationary point of (CCOP-NLP).*

Proof First, observe that (x^*, y^*) is feasible for (CCOP-NLP). Since $(x^k, y^k)_{k\in\mathbb{N}}$ is a sequence of KKT points for (REG(t)) with $t = t^k$, there exist Lagrange multipliers $(\lambda^k, \mu^k, \tilde{\gamma}^k, \delta^k, \nu^k) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^n$ for all $k \in \mathbb{N}$ such that

$$\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla_i^k h_i(x^k) + \sum_{i=1}^n \tilde{\gamma}_i^k \frac{\partial \phi(x_i^k, y_i^k)}{\partial x} e_i = 0, \tag{REG-KKT1}$$

$$-\delta^k e + \sum_{i=1}^n \nu_i^k e_i + \sum_{i=1}^n \tilde{\gamma}_i{}^k \frac{\partial \phi(x_i^k, y_i^k)}{\partial y} e_i = 0, \tag{REG-KKT2}$$

$$\lambda_i^k \begin{cases} \geq 0, & \text{if } g_i(x^k) = 0, \\ = 0, & \text{else,} \end{cases} \quad \text{for all } i = 1, \dots, m, \tag{REG-KKT3}$$

$$\delta^k \begin{cases} \geq 0, & \text{if } e^T y^k = n - S, \\ = 0, & \text{else} \end{cases} \tag{REG-KKT4}$$

$$\tilde{\gamma}_i^k \begin{cases} \geq 0, & \text{if } \phi(x_i^k, y_i^k) = t^k, \\ = 0, & \text{else,} \end{cases} \quad \text{for all } i = 1, \dots, n, \tag{REG-KKT5}$$

$$\nu_i^k \begin{cases} \leq 0, & \text{if } y_i^k = 0, \\ \geq 0, & \text{if } y_i^k = 1, \\ = 0, & \text{else,} \end{cases} \quad \text{for all } i = 1, \dots, n. \tag{REG-KKT6}$$

Here, ν_i denotes the Lagrange multiplier corresponding to the box-constraint $0 \leq y_i \leq 1$ for $i = 1, \dots, n$. From (REG-KKT2), we then obtain, for all $i = 1, \dots, n$,

$$\delta^k = \nu_i^k + \tilde{\gamma}_i^k \frac{\partial \phi(x_i^k, y_i^k)}{\partial y}. \tag{7}$$

Suppose there exists an index $i \in \{1, \dots, n\}$ such that $\nu_i^k < 0$. Then, by (REG-KKT6), we have $y_i^k = 0$. Condition 3.1 **A)** implies $t^k > 0 = \phi(x_i^k, y_i^k)$ if and only if $x_i^k = 0$ or $y_i^k = 0$, which yields $\tilde{\gamma}_i^k = 0$.
Substituting into (REG-KKT2) gives $0 > \nu_i^k = \delta^k \geq 0$, which is a contradiction. Therefore, for all $i = 1, \dots, n$, it holds that $\nu_i^k \geq 0$.
For all $k \in \mathbb{N}$, we define

$$\gamma_i^k := \tilde{\gamma}_i^k \frac{\partial \phi(x_i^k, y_i^k)}{\partial x}, \qquad \text{for all } i = 1, \dots, n.$$

We prove the *boundedness* of the sequence $(\lambda^k, \mu^k, \gamma^k)_{k \in \mathbb{N}}$ by contradiction. Suppose that $\lim_{k \to \infty} \|(\lambda^k, \mu^k, \gamma^k)\| = \infty$. The normalized sequence

$$\left(\frac{(\lambda^k, \mu^k, \gamma^k)}{\|(\lambda^k, \mu^k, \gamma^k)\|} \right)_{k \in \mathbb{N}}$$

is bounded, and without loss of generality, we assume that it converges

$$0 \neq (\bar{\lambda}, \bar{\mu}, \bar{\gamma}) := \lim_{k \to \infty} \frac{(\lambda^k, \mu^k, \gamma^k)}{\|(\lambda^k, \mu^k, \gamma^k)\|}.$$

It holds that $\bar{\lambda} \geq 0$. For all i with $g_i(x^*) < 0$, continuity of g_i implies that $g_i(x^k) < 0$ for sufficiently large k, implying by (REG-KKT3) that $\lambda_i^k = 0$ and hence $\bar{\lambda}_i = 0$; consequently, $\text{supp}(\bar{\lambda}) \subseteq I_g(x^*)$.
We prove $\text{supp}(\bar{\gamma}) \subseteq I_0(x^*)$ by contradiction. Suppose there exists an index $j \in \{1, \dots, n\}$ such that $x_j^* \neq 0$ and $\bar{\gamma}_j \neq 0$. Feasiblity then implies $y_j^* = 0$.
Since $\bar{\gamma}_j \neq 0$, it follows that $\gamma_j^k \neq 0$ and hence $\tilde{\gamma}_j^k > 0$ for sufficiently large k. By (REG-KKT5) and Condition 3.1, this implies $x_j^k \neq 0$ and $y_j^k > 0$. Together with Condition 3.1 **B)**, this implies

$$\delta^k = \nu_j^k + \tilde{\gamma}_j^k \frac{\partial \phi(x_j^k, y_j^k)}{\partial y} > 0.$$

Therefore,

$$\delta^k = \nu_i^k + \tilde{\gamma}_i^k \frac{\partial \phi(x_i^k, y_i^k)}{\partial y} > 0,$$

must hold for all $i = 1, \dots, n$. Assuming $\delta^k > 0$ for sufficiently large k, (REG-KKT4) then implies

$$e^T y^k = n - S.$$

Since $y_j^k \to y_j^* = 0$ and $y_j^k > 0$ for all sufficiently large k, we may assume that y_j^k is strictly monotone decreasing along the entire sequence $(x^k, y^k)_{k\in\mathbb{N}}$. Moreover, because $e^T y^k = n - S$ holds for all sufficiently large k, the strict monotone decrease of y_j^k implies the existence of an index m such that y_m^k is strictly monotone increasing along the whole sequence $(x^k, y^k)_{k\in\mathbb{N}}$, exactly compensating the decrease of y_j^k to maintain $e^T y^k = n - S$. Therefore, for all sufficiently large k, we obtain

$$y_m^* > 0, x_m^* = 0 \text{ and } 0 < y_m^k < 1, \nu_m^k = 0, x_m^k \neq 0.$$

We thus conclude that $x_m^k \neq 0$ for sufficiently large k, since $\delta^k > 0$ and $\nu_m^k = 0$, by (REG-KKT2), (REG-KKT5) and Condition 3.1 **A)** and **B)**.

Moreover, we have $(x_j^k, y_j^k) \overset{k\to\infty}{\longrightarrow} (x_j^*, y_j^*)$ with $x_j^* \neq 0, y_j^* = 0$ and $x_j^k \neq 0, y_j^k > 0$ for all $k \in \mathbb{N}$, as well as $(x_m^k, y_m^k) \overset{k\to\infty}{\longrightarrow} (x_m^*, y_m^*)$ with $x_m^* = 0, y_m^* > 0$ and $x_m^k \neq 0, y_m^k > 0$ for all $k \in \mathbb{N}$.

Combining the above, and applying Condition 3.1 **C)**, we obtain

$$\frac{|\gamma_j^k|}{|\gamma_m^k|} = \frac{\left|\tilde{\gamma}_j^k \frac{\partial\phi(x_j^k,y_j^k)}{\partial x}\right|}{\left|\tilde{\gamma}_m^k \frac{\partial\phi(x_m^k,y_m^k)}{\partial x}\right|} \overset{\nu_j^k=\nu_m^k=0}{=} \frac{\left|\frac{\delta^k}{\frac{\partial\phi(x_j^k,y_j^k)}{\partial y}} \cdot \frac{\partial\phi(x_j^k,y_j^k)}{\partial x}\right|}{\left|\frac{\delta^k}{\frac{\partial\phi(x_m^k,y_m^k)}{\partial y}} \cdot \frac{\partial\phi(x_m^k,y_m^k)}{\partial x}\right|} = \left|\frac{\frac{\partial\phi(x_m^k,y_m^k)}{\partial y}}{\frac{\partial\phi(x_m^k,y_m^k)}{\partial x}}\right| \cdot \left|\frac{\frac{\partial\phi(x_j^k,y_j^k)}{\partial x}}{\frac{\partial\phi(x_j^k,y_j^k)}{\partial y}}\right| \overset{k\to\infty}{\longrightarrow} 0.$$

This leads to the contradiction

$$0 \neq |\bar{\gamma}_j| = \lim_{k\to\infty} \frac{|\gamma_j^k|}{\|(\lambda^k, \mu^k, \gamma^k)\|} \leq \lim_{k\to\infty} \frac{|\gamma_j^k|}{|\gamma_m^k|} = 0.$$

Hence, we obtain $\operatorname{supp}(\bar{\gamma}) \subseteq I_0(x^*)$.

To show the boundedness of the sequence $(\lambda^k, \mu^k, \gamma^k)_{k\in\mathbb{N}}$, we use the assumption that CC-MFCQ holds at (x^*, y^*). Dividing the first KKT condition (REG-KKT1) by $\|(\lambda^k, \mu^k, \gamma^k)\|$ and letting $k \to \infty$, it follows, together with the preceding arguments, that

$$\sum_{i\in I_g(x^*)} \bar{\lambda}_i \nabla g_i(x^*) + \sum_{i=1}^{p} \bar{\mu}_i \nabla h_i(x^*) + \sum_{i\in I_0(x^*)} \bar{\gamma}_i e_i = 0.$$

Together with $\bar{\lambda} \geq 0, \operatorname{supp}(\bar{\lambda}) \subseteq I_g(x^*), \operatorname{supp}(\bar{\gamma}) \subseteq I_0(x^*)$, and $(\bar{\lambda}, \bar{\mu}, \bar{\gamma}) \neq 0$, this contradicts CC-MFCQ in (x^*, y^*), since CC-MFCQ requires these gradients to be positive linearly independent. Consequently, the sequence of multipliers $(\lambda^k, \mu^k, \gamma^k)_{k\in\mathbb{N}}$ is bounded. Without loss of generality, we assume that the whole sequence $(\lambda^k, \mu^k, \gamma^k)_{k\in\mathbb{N}}$ converges

$$(\lambda^*, \mu^*, \gamma^*) \coloneqq \lim_{k\to\infty} (\lambda^k, \mu^k, \gamma^k).$$

Taking the limit as$k \to \infty$ in the first KKT condition (REG-KKT1), we obtain

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{i=1}^{p} \mu_i^* \nabla h_i(x^*) + \sum_{i=1}^{n} \gamma_i^* e_i = 0.$$

Hence, (x^*, y^*), together with the multipliers $\lambda^*, \mu^*, \gamma^*$, satisfies the first condition for M-stationarity. Analogous arguments as above yield $\lambda^* \geq 0$, $\operatorname{supp}(\lambda^*) \subseteq I_g(x^*)$, and $\operatorname{supp}(\gamma^*) \subseteq I_0(x^*)$. Therefore, (x^*, y^*) is M-stationary. □

Corollary 4.2 *Let $(t^k)_{k\in\mathbb{N}}$ be a sequence with $t^k > 0$ for all $k \in \mathbb{N}$ and $t^k \downarrow 0$ as $k \to \infty$. Let $(x^k, y^k)_{k\in\mathbb{N}}$ be a sequence of KKT points of (REG(t)) with $t = t^k$ converging to (x^*, y^*). If CC-MFCQ holds at (x^*, y^*), then (x^*, y^*) is M-stationary and there exists a $z^* \in \mathbb{R}^n$ such that (x^*, z^*) is S-stationary for (CCOP-NLP).*

Proof By Theorem 4.1, this follows directly from Proposition 2.10. □

The following result is analogous to [12, Theorem 3].

Theorem 4.3 *Let (x^*, y^*) be feasible for (CCOP-NLP) and CC-MFCQ hold there. Then there is a neighbourhood U of (x^*, y^*) such that for all $t > 0$ the standard MFCQ for (REG(t)) holds at every $(x, y) \in U$ feasible for (REG(t)).*

Proof We prove the statement by contradiction. Assume, to the contrary, that the statement does not hold. Then there exists a sequence $(x^k, y^k)_{k\in\mathbb{N}} \overset{k\to\infty}{\longrightarrow} (x^*, y^*)$ and $(t^k)_{k\in\mathbb{N}} > 0$ for all $k \in \mathbb{N}$, such that (x^k, y^k) is feasible for (REG(t)), with $t = t^k$, but in (x^k, y^k) MFCQ is violated for (REG(t)), with $t = t^k$. Consequently, there exist multipliers $(\lambda^k, \mu^k, \tilde{\gamma}^k, \delta^k, \nu^k) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^n$ for all $k \in \mathbb{N}$ such that

$$(\lambda^k, \mu^k, \tilde{\gamma}^k, \delta^k, \nu^k) \neq 0, \tag{8}$$

and the conditions,

$$\sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla_i^k h_i(x^k) + \sum_{i=1}^n \tilde{\gamma}_i^k \frac{\partial \phi(x_i^k, y_i^k)}{\partial x} e_i = 0, \tag{LOC-MFCQ1}$$

and (REG-KKT2) - (REG-KKT6) are satisfied. Here, ν_i denotes the multiplier corresponding to the box-constraint $0 \le y_i \le 1$, for all $i = 1, \dots, n$. Since MFCQ is violated for (REG(t)) with $t = t^k$, the relevant gradients are positve linearly dependent at (x^k, y^k).
As in the proof of Theorem 4.1, we can rule out the case $\nu_i^k < 0$, and thus assume $\nu_i^k \ge 0$ for all $i = 1, \dots, n$.
Define $\gamma_i^k := \tilde{\gamma}_i^k \dfrac{\partial \phi(x_i^k, y_i^k)}{\partial x}$ for all $i = 1, \dots, n$ and $k \in \mathbb{N}$. By (REG-KKT5), we have $\phi(x_i^k, y_i^k) = t^k$, and Condition 3.1 **A)** implies $y_i^k > 0, x_i^k \neq 0$ if $\tilde{\gamma}_i^k > 0$. In particular, for these indices i and $k \in \mathbb{N}$, Condition 3.1 **B)** ensures $\dfrac{\partial \phi(x_i^k, y_i^k)}{\partial x} \neq 0$. This yields

$$\operatorname{supp}(\gamma^k) = \operatorname{supp}(\tilde{\gamma}^k) \quad \forall k \in \mathbb{N}. \tag{9}$$

In addition, from (REG-KKT5) and Condition 3.1 **B)**, we also obtain

$$\tilde{\gamma}_i^k \frac{\partial \phi(x_i^k, y_i^k)}{\partial y} = \begin{cases} \gamma_i^k \cdot \dfrac{\frac{\partial \phi(x_i^k, y_i^k)}{\partial y}}{\frac{\partial \phi(x_i^k, y_i^k)}{\partial x}}, & \text{if } \phi(x_i^k, y_i^k) = t^k, \\ 0, & \text{else.} \end{cases} \tag{10}$$

Therefore, for all $k \in \mathbb{N}$, we can write (LOC-MFCQ1) and (REG-KKT2) as

$$\sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \gamma_i^k e_i = 0, \tag{11}$$

$$\delta^k = \begin{cases} \nu_i^k + \gamma_i^k \cdot \frac{\frac{\partial \phi(x_i^k, y_i^k)}{\partial y}}{\frac{\partial \phi(x_i^k, y_i^k)}{\partial x}}, \text{ if } \phi(x_i^k, y_i^k) = t^k, \\ \nu_i^k, \text{ else,} \end{cases} \quad \text{for all } i = 1, \dots, n. \tag{12}$$

For all $i = 1, \dots, n$ and all $k \in \mathbb{N}$, we have $\tilde{\gamma}_i^k \frac{\partial \phi(x_i^k, y_i^k)}{\partial y} \geq 0$, by (REG-KKT5) together with Condition 3.1 **A)** and **B)**.

We now show that this contradicts the assumption that CC-MFCQ holds at (x^*, y^*).

By assumption, $(\lambda^k, \mu^k, \tilde{\gamma}^k, \delta^k, \nu^k) \neq 0$ for all $k \in \mathbb{N}$. Without loss of generality, we can choose the multipliers such that $\|(\lambda^k, \mu^k, \gamma^k, \delta^k, \nu^k)\| = 1$ for all $k \in \mathbb{N}$ and the whole sequence converges

$$0 \neq (\lambda, \mu, \gamma, \delta, \nu) := \lim_{k \to \infty} (\lambda^k, \mu^k, \gamma^k, \delta^k, \nu^k). \tag{13}$$

We have $\lambda \geq 0$. For all i with $g_i(x^*) < 0$, continuity implies $g_i(x^k) < 0$ for sufficiently large k, and thus $\lambda_i^k = 0$. It follows that

$$\operatorname{supp}(\lambda) \subseteq I_g(x^*). \tag{14}$$

We proceed to prove $\operatorname{supp}(\gamma) \subseteq I_0(x^*)$ by contradiction.

Assume there exists an index $j \in \{1, \dots, n\}$ with $\gamma_j \neq 0, x_j^* \neq 0$. Feasibility then implies $y_j^* = 0$. Moreover, $\gamma_j^k \neq 0$ and hance $\tilde{\gamma}_j^k \neq 0$, with $\phi(x_j^k, y_j^k) = t^k$. By Condition 3.1 **A)**, this gives $x_j^k \neq 0, y_j^k > 0$, with $y_j^k \to 0$ as $k \to \infty$. Thus, for sufficiently large k, we have $y_j^k < 1$ and $\nu_j^k = 0$ by (REG-KKT6).

Using Condition 3.1 **C)** and (12), we obtain

$$\delta^k = \nu_j^k + \gamma_j^k \frac{\frac{\partial \phi(x_j^k, y_j^k)}{\partial y}}{\frac{\partial \phi(x_j^k, y_j^k)}{\partial x}} = \gamma_j^k \frac{1}{\frac{\frac{\partial \phi(x_j^k, y_j^k)}{\partial x}}{\frac{\partial \phi(x_j^k, y_j^k)}{\partial y}}} \stackrel{k \to \infty}{\longrightarrow} \infty.$$

Since $(\lambda^k, \mu^k, \gamma^k, \delta^k, \nu^k)_{k \in \mathbb{N}}$ converges, this yields a contradiction. Therefore, we conclude that

$$\operatorname{supp}(\gamma) \subseteq I_0(x^*). \tag{15}$$

It remains to show that $(\lambda, \mu, \gamma) \neq 0$. We proceed by contradiction and assume $(\lambda, \mu, \gamma) = 0$. Since $(\lambda, \mu, \gamma, \delta, \nu) \neq 0$, it follows that $(\delta, \nu) \neq 0$. From the arguments above, we have $\nu^k \geq 0$ and, by (12), $\delta^k \geq \max_{i=1,\dots,n} \nu_i^k$. Hence, $(\delta, \nu) \neq 0$ implies $\delta > 0$ and $\delta^k > 0$ for sufficiently large k. This is only possible if $e^T y^k = n - S$ for all sufficiently large k.

For all i with $y_i^* > 0$, we have $x_i^* = 0$, and without loss of generality, we assume that along the sequence $(x_i^k, y_i^k) \stackrel{k \to \infty}{\longrightarrow} (x_i^*, y_i^*)$, either $x_i^k = 0, y_i^k > 0$ or $x_i^k \neq 0, y_i^k > 0$ for all $k \in \mathbb{N}$.

If $x_i^k = 0, y_i^k > 0$ for all $k \in \mathbb{N}$, then, by (REG-KKT2) and Condition 3.1 **B)**, it follows that

$$0 < \delta^* = \lim_{k \to \infty} \nu_i^k + \tilde{\gamma}_i^k \frac{\partial \phi(x_i^k, y_i^k)}{\partial y} = \nu_i,$$

and, if $x_i^k \neq 0, y_i^k > 0$ for all $k \in \mathbb{N}$, then $\gamma = 0$, which, together with Condition 3.1 **C)** and (12), yields

$$0 < \delta^* = \lim_{k \to \infty} \nu_i^k + \gamma_i^k \frac{\frac{\partial \phi(x_i^k, y_i^k)}{\partial y}}{\frac{\partial \phi(x_i^k, y_i^k)}{\partial x}} = \nu_i.$$

Hence, for all sufficiently large k and i with $y_i^k > 0$, (REG-KKT6) implies $y_i^k = 1$, and consequently $y_i^* = 1$. However, since $e^T y^k = n - S < n$ for all sufficiently large k, there must exist at least one index m such that $y_m^k = 0$ for all sufficiently large k. By Condition 3.1 **A)**, it follows that $\phi(x_m^k, y_m^k) = 0 < t_k$ and $\nu_m^k = 0$. This, however, implies $\delta^k = 0$, a contradiction. Therefore, the assumption $(\lambda, \mu, \gamma) = 0$ is false, and we conclude that

$$(\lambda, \mu, \gamma) \neq 0.$$

Alltogether, taking the limit $k \to \infty$ and using (14) and (15), we obtain

$$\sum_{i \in I_g(x^*)} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla h_i(x^*) + \sum_{i \in I_0(x^*)} \gamma_i e_i = 0.$$

Since $(\lambda, \mu, \gamma) \neq 0$ and $\lambda \geq 0$, the corresponding gradients are positive linearly dependent, which contradicts CC-MFCQ. The statement therefore follows. □

The following global convergence result is analogous to [18, Theorem 4.1].

Theorem 4.4 *Let $(t^k)_{k\in\mathbb{N}}$ be a sequence with $t^k > 0$ for all $k \in \mathbb{N}$ and $t^k \downarrow 0$ as $k \to \infty$. Suppose that (x^k, y^k) is a globally optimal solution of (REG(t)) for $t = t^k$ and (x^*, y^*) is an accumulation point of the sequence $(x^k, y^k)_{k\in\mathbb{N}}$ as $k \to \infty$. Then, (x^*, y^*) is a globally optimal solution of (CCOP-NLP).*

Proof Denote by $Z(t^k)$ the feasible set of the regularized optimization problem (REG(t)) with $t = t^k$ and by Z the feasible set of (CCOP-NLP). For all $t \geq 0$, it holds that $Z \subseteq Z(t)$, and for $0 \leq t_1 \leq t_2$ we have $Z(t_1) \subseteq Z(t_2)$.
Taking a subsequence if necessary, we assume that $\lim_{k\to\infty} (x^k, y^k) = (x^*, y^*)$. We observe that (x^*, y^*) is feasible for (CCOP-NLP). Let (x^k, y^k) be a globally optimal solution of (REG(t)) with $t = t^k$ for all $k \in \mathbb{N}$. Since $Z \subseteq Z(t^k)$ for all $k \in \mathbb{N}$, it follows that

$$f(x^k) \leq f(x), \quad \forall (x, y) \in Z.$$

Letting $k \to \infty$ and using the continuity of f, we obtain

$$f(x^*) \leq f(x), \quad \forall (x, y) \in Z.$$

Hence, (x^*, y^*) is a globally optimal solution of (CCOP-NLP). □

5 Numerical experiments

In this section, we conduct an extensive numerical study to evaluate the performance of the proposed solution methods and compare them with established approaches.
All experiments were performed on a computer equipped with an Apple M1 chip, 8 GB of RAM, and an 8-core CPU (3.2 GHz). The numerical experiments were implemented in `Python 3.11.1`.
We use the following notation to denote the solution methods:

1. `RELAX`: Directly solves the continuous reformulation (CCOP-NLP).
2. `Gurobi`: Solves a mixed-integer (nonlinear) reformulation of the corresponding cardinality-constrained optimization problem.

3. `SCHOL-REG`: Applies the Scholtes-type regularization proposed in [12] and solves a sequence of the corresponding regularized optimization problems.
4. `KS-REG`: Applies the Kanzow–Schwartz regularization proposed in [14] and solves a sequence of the corresponding regularized optimization problems.
5. `KKM-PEN(`λ`)`: Applies the proposed penalty method and solves a sequence of optimization problems (CCOP-PEN(ρ)) with penalty term $\Phi(x, y) = \sum_{i=1}^{n} \phi(x, y)$, where $\phi(x, y) = \begin{cases} \left(-\sqrt{(|x| - y)^2 + \lambda |x| y} + |x| + y\right)^2, & \text{if } y \geq 0, \\ 0, & \text{if } y < 0, \end{cases}$ as introduced in Theorem 3.3. We consider $\lambda \in \{1, 0.1, 0.01\}$.
6. `QUAD-PEN`: Applies the proposed penalty method and solves a sequence of optimization problems (CCOP-PEN(ρ)), with penalty term $\Phi(x, y) = \sum_{i=1}^{n} x_i^2 y_i^2$.

To solve the respective mixed-integer programs, we use the commercial solver `Gurobi 12.0.3` [26], accessed via `gurobipy 12.0.3` and set the `Gurobi`-parameter `MIPFocus` $= 3$.
For solving each optimization problem in the penalty and regularization approaches, we use `IPOPT 3.14.16` via `Pyomo 6.8.2`. Thereby, `MUMPS 5.6.2` is used to solve the single linear equation systems. See [27, 28] for `IPOPT`, [29, 30] for `Pyomo`, and [31, 32] for `MUMPS`. We select the `IPOPT` parameter `tol` $= 10^{-9}$ and `contr_viol_tol` $= 10^{-10}$. For the regularization methods, the initial value is chosen as $t^0 = 1$, and decreased in each step according to $t^{k+1} = 0.1 \cdot t^k$. To solve the $(k + 1)$-th regularized optimization problem, we initialize the solver with the solution obtained from the k-th subproblem. The choice of the initial starting point for $t^0 = 1$ is specified in each experiment. If `IPOPT` terminates with an error, we proceed by reducing t^k according to the update rule, and reinitialize the solver with the last successfully computed solution. The regularization method is terminated when either the maximum violation of the complementarity constraint is below 10^{-5}, i.e., $\max_{i=1,\ldots,n}\{|x_i| y_i\} \leq 10^{-5}$, or after solving the regularized optimization problem with $t = 10^{-9}$.
For the penalty method, the penalty parameter is updated according to $\rho^{k+1} = 2\rho^k$. The initial penalty parameter ρ^0 and starting point are specified in each experiment. As in the regularization method, the solution of the previous subproblem is used as the starting point for the next one. The procedure is terminated once the maximum violation of the complementarity constraint satisfies $\max_{i=1,\ldots,n}\{|x_i| y_i\} \leq 10^{-5}$. If `IPOPT` reports an error while solving a subproblem, we simply increase the penalty parameter according to the update rule and reinitialize the solver with the last successfully computed solution. With this strategy, the penalty method terminated successfully in all experiments.
In all experiments, we initialize the numerical methods with the starting point (x^0, y^0), where we set $y^0 = e$.
To solve the continuous relaxation (CCOP-NLP) directly, we use `IPOPT` with the same settings and the same initial point as in the regularization methods and penalty methods.
The newly introduced general regularization method is excluded from the numerical experiments. In the final experiment in section 5.3, it fails to achieve the desired

maximum complementarity constraint violation. We note, however, the general regularization method performs very well in the numerical experiments presented in sections 5.1 and 5.2.

5.1 Sensitivity starting point

We consider the following cardinality-constrained optimization problem

$$\begin{aligned} \min_{x_1,x_2\in\mathbb{R}} \quad & f(x_1,x_2) = 6.85x_1 + e^{x_1+1} + 0.7(0.5x_1+2)^2 - 8.25x_2 + (x_2-1)^2 \qquad \text{(EXP1)} \\ \text{s.t.} \quad & (x_1-1)^2 + x_2^2 \le 10, \qquad \|x\|_0 \le 1. \end{aligned}$$

The local minima of (EXP1) are given by $z^1 = (-\sqrt{10}+1, 0) \approx (-2.16228, 0)$, $f(z^1) \approx -12.9078$, and $z^2 = (0, 3.0)$, $f(z^2) \approx -15.2317$, where z^2 is the unique global minimum. The optimization problem (EXP1) is constructed such that after removing the cardinality constraint, the global minimum lies between the two local minima of (EXP1). For the choice of the starting point (x^0, y^0), $x^0, y^0 \in \mathbb{R}^2$, we discretize the set $[2.625, 4.875] \times [-3.75, 3.75]$, by considering $L = \{(x_1, x_2) : x_1 = -2.625 + 0.375z_1, x_2 = -3.75 + 0.375z_2, z_1, z_2 \in \{0, 1, \dots, 20\}\}$, such that we obtain $|L| = 441$ starting points, by choosing (x^0, y^0) with $x^0 \in L$ and $y^0 = (1, 1)$.
We choose as the initial penalty parameter $\rho^0 = 2$. Table 1 reports the result of the experiment by indicating for each numerical solution method, how often the x-component of the computed solutions converges to z^1, z^2 or $(0, 0)$.

Table 1: Frequencies of finding local (global) minimizers by starting point for (EXP1)

Method	z^2 (global minimum)	z^1	$(0,0)$
`RELAX`	**122**	146	173
`SCHOL-REG`	**441**	0	0
`KS-REG`	**358**	83	0
`KKM-PEN(1)`	**441**	0	0
`KKM-PEN(0.1)`	**351**	90	0
`KKM-PEN(0.01)`	**341**	100	0
`QUAD-PEN`	**441**	0	0

First, we observe that all penalty and regularization methods outperform `RELAX`, and that the solution methods exhibit distinct local convergence behavior, reflecting their sensitivity to the choice of starting point as well as to the regularization and penalty parameters.

5.2 N-dimensional Rosenbrock function

We consider the following n-dimensional version of the well-known *Rosenbrock function (Banana function or Rosenbrock's banana function)* $f : \mathbb{R}^n \to \mathbb{R}$, given by $f(x) =$

$\sum_{i=1}^{n-1} \left(100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2\right)$. Based on this function, we consider the following cardinality-constrained optimization problem

$$\begin{aligned}
\min_{x \in \mathbb{R}^n} \quad & f(x) = \sum_{i=1}^{n-1} \left(100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2\right) && \text{(CC-RB)} \\
\text{s.t.} \quad & -10 \le x_i \le 10, && \text{for all } i = 1, \dots, n, \\
& \|x\|_0 \le S.
\end{aligned}$$

We consider $n = 10, 20, 30, 40$. For each n, we vary the sparsity parameter $S = 1, \dots, n-1$, resulting in a total of 96 test instances. To check the quality of the obtained solutions, we solve the following mixed-integer-nonlinear optimization problem.

$$\begin{aligned}
\min_{x,y,z \in \mathbb{R}^n} \quad & f(x) = \sum_{i=1}^{n-1} \left(100(x_{i+1} - z_i)^2 + (1 - x_i)^2\right) && \text{(CC-RB-MINLP)} \\
\text{s.t.} \quad & -10y_i \le x_i \le 10y_i, && \text{for all } i = 1, \dots, n, \\
& \sum_{i=1}^{n} y_i \le S, \\
& z_i = x_i^2, && \text{for all } i = 1, \dots, n, \\
& -10 \le x_i \le 10, \quad y_i \in \{0, 1\}, && \text{for all } i = 1, \dots, n.
\end{aligned}$$

We choose as the starting point for the numerical methods $(x^0, y^0) = (e, e)$ and as the starting penalty parameter $\rho^0 = 2$. The tables 2, 3, 4 and 5 summarize the results for $n = 10, 20, 30, 40$. The column `Global` reports the number of instances in which the x-component of the generated sequence converges to the global solution obtained by solving (CC-RB-MINLP) using `Gurobi` (without a `TimeLimit`). The column T shows the average computation time in seconds, defined as the time spent solving optimization problems. The column $\overline{mv}$ provides the average maximal absolute violation of the complementarity constriants across all test instances. Let $(x^*, y^*) \in \mathbb{R}^n \times \mathbb{R}^n$ denote the solution computed by a method for a single instance. The maximum absolute violation of the complementarity constraint is then defined as $\max\{|x_i^* y_i^*| : i = 1, \dots, n\}$.

Once again, all regularization and penalty methods outperform `RELAX`. Particularly noteworthy is that the methods `QUAD-PEN`, `KKM-PEN(1)` and `KKM-PEN(0.1)` successfully computed solutions for all instances, with their x-components converging to the global minimizer. The methods `SCHOL-REG` and `KKM-PEN(0.01)` also performed well, failing only once and twice, respectively, to find the global optimal solution. Moreover, the computational effort required by `Gurobi` illustrates the combinatorial explosion inherent in solving large-scale MINLPs to global optimality.

Table 2: Comparison for $n = 10$

METHOD	Global	$\overline{mv}$	T
Gurobi	9 / 9	-	0.0524
RELAX	0 / 9	$2.59 \cdot 10^{-32}$	0.0213
SCHOL-REG	8 / 9	$1.00 \cdot 10^{-6}$	0.1598
KS-REG	7 / 9	$4.89 \cdot 10^{-6}$	0.1457
KKM-PEN(1)	9 / 9	$8.23 \cdot 10^{-6}$	0.3010
KKM-PEN(0.1)	9 / 9	$6.27 \cdot 10^{-6}$	0.3033
KKM-PEN(0.01)	9 / 9	$7.63 \cdot 10^{-6}$	0.2844
QUAD-PEN	9 / 9	$9.26 \cdot 10^{-6}$	0.4164

Table 3: Comparison for $n = 20$

METHOD	Global	$\overline{mv}$	T
Gurobi	19 / 19	-	0.4124
RELAX	1/ 19	$4.96 \cdot 10^{-34}$	0.0205
SCHOL-REG	19 / 19	$9.53 \cdot 10^{-7}$	0.1698
KS-REG	17 / 19	$5.21 \cdot 10^{-6}$	0.1722
KKM-PEN(1)	19 / 19	$8.56 \cdot 10^{-6}$	0.3193
KKM-PEN(0.1)	19 / 19	$5.73 \cdot 10^{-6}$	0.3126
KKM-PEN(0.01)	19 / 19	$6.25 \cdot 10^{-6}$	0.3177
QUAD-PEN	19 / 19	$9.63 \cdot 10^{-6}$	0.3532

Table 4: Comparison for $n = 30$

METHOD	Global	$\overline{mv}$	T
Gurobi	29 / 29	-	6.6561
RELAX	1 / 29	$1.24 \cdot 10^{-34}$	0.0251
SCHOL-REG	29 / 29	$1.0 \cdot 10^{-6}$	0.1953
KS-REG	23 / 29	$4.39 \cdot 10^{-6}$	0.2460
KKM-PEN(1)	29 / 29	$8.66 \cdot 10^{-6}$	0.3410
KKM-PEN(0.1)	29 / 29	$5.56 \cdot 10^{-6}$	0.3384
KKM-PEN(0.01)	29 / 29	$5.82 \cdot 10^{-6}$	0.3382
QUAD-PEN	29 / 29	$9.75 \cdot 10^{-6}$	0.3765

5.3 Portfolio optimization

We consider cardinality-constrained optimization problems of the form

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & x^T Q x && \text{(P-OPT-EXP)} \\ \text{s.t.} \quad & \mu^T x \geq \rho, \quad e^T x \leq 1, \qquad 0 \leq x_i \leq u_i, \quad \forall i = 1, \ldots, n, \\ & \|x\|_0 \leq S. \end{aligned}$$

Table 5: Comparison for $n = 40$

METHOD	Global	$\overline{mv}$	T
Gurobi	39 / 39	-	114.5688
RELAX	1 / 39	$2.45 \cdot 10^{-31}$	0.0244
SCHOL-REG	39 / 39	$9.54 \cdot 10^{-7}$	0.2087
KS-REG	26 / 39	$5.91 \cdot 10^{-6}$	0.3124
KKM-PEN(1)	39 / 39	$8.71 \cdot 10^{-6}$	0.4012
KKM-PEN(0.1)	39 / 39	$5.48 \cdot 10^{-6}$	0.3729
KKM-PEN(0.01)	37 / 39	$5.85 \cdot 10^{-6}$	0.3655
QUAD-PEN	39 / 39	$9.8 \cdot 10^{-6}$	0.3999

This is a *sparse portfolio optimization problem*, where Q denotes the covariance matrix and μ the expected return of n considered assets. The constraint $e^T x \leq 1$ represents the budget limitiation, while $x_i \geq 0$ ensures that there is no short-selling and $x_i \leq u_i$ imposes upper bounds on the individual asset weights. The constraint $\mu^T x \geq \rho$ guarantees a minimum expected return level, while the objective minimizes risk as measured by the portfolio variance. The cardinality-constraint restricts the portfolio to contain at most S assets, i.e., the portfolio is in a certain sense sparse. For general portfolio optimization problems, we refer to [33] and for sparse portfolio optimization, see [3].

To generate test instances, we use the same randomly generated data sets for Q, μ, ρ, and u, as in [34], which are available at [35]. These data sets and test instances were also used in [5] and [18] to evaluate the proposed regularization methods for (CCOP-NLP). In this experiment, the initial penalty parameter is set to $\rho^0 = 100$. The starting point for the numerical methods is chosen as (x^0, y^0) with $x^0 = u$.

In this experiment, the method `RELAX` did not always terminate. In such cases, the number of test instances for which termination failed is reported in the tables 6, 7, 8, 9, 10, 11, 12, 13, 14 and 15 under `Failed`. For the regularization methods, we report an instance as `Failed` if the algorithm does not ensure that the maximum absolute complementarity constraint violation falls below 10^{-5}.

We use only the data sets for problem dimensions $n = 200$ and $n = 300$, with 30 test examples per dimension. In addition, we consider the cardinality parameters $S = 5, 10, 20, 30$ and 50, resulting in a total of 300 test instances.

To evaluate the quality of the computed points, we report under α the number of test instances for which the computed point $z^i_{\texttt{METHOD}}$ satisfies

$$f(z^i_{\texttt{METHOD}}) \leq (1 + \alpha) f(z^i_{\texttt{Gurobi}}), \tag{16}$$

where $z^i_{\texttt{METHOD}}$ denotes the computed point by the respective `METHOD` for the i-th test instance, and $z^i_{\texttt{Gurobi}}$ is the corresponding solution obtained using `Gurobi`. f denotes the objective function of (P-OPT-EXP). Under `MIPGap` we report for `Gurobi` the average `MIPGap`. In this experiment, the `Gurobi` parameter `TimeLimit` was set to 600 seconds. The columns $\overline{mv}$ and T report the same as in the previous experiment. We consider $\alpha \in \{0.01, 0.02, 0.05, 0.1, 025, 0.5, 1.0\}$.

Since (P-OPT-EXP) includes the sign restriction $x \geq 0$, we effectively apply the MPCC regularizations [13] and [14] as the respective regularization methods SCHOL-REG and KS-REG.

Table 6: Comparison of solution methods for $S = 5$ and $n = 200$

METHOD	$\alpha = 0.01$	0.02	0.05	0.1	0.25	0.5	1.0	MIPGap / $\overline{mv}$	T	Failed
Gurobi	-	-	-	-	-	-	-	$3.69 \cdot 10^{-5}$	1.43	0
RELAX	0	0	0	0	0	0	3	$2.10 \cdot 10^{-49}$	0.65	0
SCHOL-REG	0	0	0	7	21	**30**	-	$9.70 \cdot 10^{-7}$	7.83	0
KS-REG	0	0	0	0	3	11	24	$9.05 \cdot 10^{-6}$	5.45	0
KKM-PEN(1)	4	6	11	24	27	29	**30**	$5.77 \cdot 10^{-6}$	11.71	0
KKM-PEN(0.1)	6	10	21	28	**30**	-	-	$7.61 \cdot 10^{-6}$	10.82	0
KKM-PEN(0.01)	6	11	23	29	**30**	-	-	$7.42 \cdot 10^{-6}$	9.35	0
QUAD-PEN	0	0	0	11	28	**30**	-	$7.70 \cdot 10^{-6}$	16.44	0

Table 7: Comparison of solution methods for $S = 10$ and $n = 200$

METHOD	$\alpha = 0.01$	0.02	0.05	0.1	0.25	0.5	1.0	MIPGap / $\overline{mv}$	T	Failed
Gurobi	-	-	-	-	-	-	-	$5.26 \cdot 10^{-5}$	4.38	-
RELAX	0	0	0	0	0	0	0	$1.16 \cdot 10^{-49}$	0.63	0
SCHOL-REG	0	0	0	7	25	28	29	$1.21 \cdot 10^{-6}$	17.86	0
KS-REG	0	0	0	0	1	20	26	$7.45 \cdot 10^{-6}$	9.17	0
KKM-PEN(1)	0	0	3	7	12	23	29	$6.83 \cdot 10^{-6}$	17.14	0
KKM-PEN(0.1)	4	10	18	22	22	27	29	$6.83 \cdot 10^{-6}$	13.86	0
KKM-PEN(0.01)	8	19	24	27	27	27	27	$7.35 \cdot 10^{-6}$	10.86	0
QUAD-PEN	0	0	2	10	28	**30**	-	$7.64 \cdot 10^{-6}$	11.82	0

Table 8: Comparison of solution methods for $S = 20$ and $n = 200$

METHOD	$\alpha = 0.01$	0.02	0.05	0.1	0.25	0.5	1.0	MIPGap / $\overline{mv}$	T	Failed
Gurobi	-	-	-	-	-	-	-	$1.06 \cdot 10^{-3}$	172.79	-
RELAX	0	0	0	0	0	0	0	$1.63 \cdot 10^{-49}$	0.63	2
SCHOL-REG	0	0	5	20	25	28	28	$1.67 \cdot 10^{-6}$	20.40	0
KS-REG	0	0	0	1	11	21	25	$4.17 \cdot 10^{-6}$	14.02	0
KKM-PEN(1)	0	0	0	0	12	21	26	$7.28 \cdot 10^{-6}$	29.24	0
KKM-PEN(0.1)	0	0	3	4	7	18	25	$7.41 \cdot 10^{-6}$	22.13	0
KKM-PEN(0.01)	5	8	13	15	18	20	25	$7.87 \cdot 10^{-6}$	15.96	0
QUAD-PEN	0	0	7	27	**30**	-	-	$7.69 \cdot 10^{-6}$	19.26	0

Table 9: Comparison of solution methods for $S = 30$ and $n = 200$

METHOD	$\alpha = 0.01$	0.02	0.05	0.1	0.25	0.5	1.0	MIPGap / $\overline{mv}$	T	Failed
Gurobi	-	-	-	-	-	-	-	$3.8 \cdot 10^{-3}$	395.15	-
RELAX	0	0	0	0	0	0		0 $3.0 \cdot 10^{-50}$	0.643	1
SCHOL-REG	0	2	16	22	25	26	29	$8.11 \cdot 10^{-6}$	20.46	1
KS-REG	0	0	0	1	5	14	16	$4.7 \cdot 10^{-4}$	11.77	5
KKM-PEN(1)	0	0	0	0	5	16	22	$6.79 \cdot 10^{-6}$	36.98	0
KKM-PEN(0.1)	0	0	0	0	7	17	23	$7.58 \cdot 10^{-6}$	29.52	0
KKM-PEN(0.01)	0	2	4	4	9	19	24	$7.53 \cdot 10^{-6}$	21.20	0
QUAD-PEN	0	5	28	**30**	-	-	-	$7.04 \cdot 10^{-6}$	14.88	0

Table 10: Comparison of solution methods for $S = 50$ and $n = 200$

METHOD	$\alpha = 0.01$	0.02	0.05	0.1	0.25	0.5	1.0	MIPGap / $\overline{mv}$	T	Failed
Gurobi	-	-	-	-	-	-	-	$5.63 \cdot 10^{-3}$	697.59	-
RELAX	0	0	0	0	0	0	0	$2.47 \cdot 10^{-44}$	0.68	1
SCHOL-REG	7	16	17	20	26	28	29	$5.0 \cdot 10^{-6}$	22.86	0
KS-REG	0	0	1	4	6	9	14	$5.8 \cdot 10^{-4}$	14.94	6
KKM-PEN(1)	0	0	0	0	3	7	18	$6.73 \cdot 10^{-6}$	50.23	0
KKM-PEN(0.1)	0	0	0	0	2	8	14	$7.36 \cdot 10^{-6}$	44.02	0
KKM-PEN(0.01)	0	0	0	3	6	11	18	$7.22 \cdot 10^{-6}$	37.62	0
QUAD-PEN	13	28	**30**	-	-	-	-	$7.05 \cdot 10^{-6}$	16.10	0

Table 11: Comparison of solution methods for $S = 5$ and $n = 300$

METHOD	$\alpha = 0.01$	0.02	0.05	0.1	0.25	0.5	1.0	MIPGap / $\overline{mv}$	T	Failed
Gurobi	-	-	-	-	-	-	-	$3.22 \cdot 10^{-5}$	3.39	-
RELAX	0	0	0	0	0	0	4	$3.92 \cdot 10^{-50}$	1.38	1
SCHOL-REG	0	1	1	7	22	29	**30**	$9.85 \cdot 10^{-7}$	18.97	0
KS-REG	0	0	0	1	1	7	17	$9.05 \cdot 10^{-6}$	13.76	0
KKM-PEN(1)	1	2	8	20	25	26	28	$6.83 \cdot 10^{-6}$	34.06	0
KKM-PEN(0.1)	4	7	18	27	**30**	-	-	$6.95 \cdot 10^{-6}$	29.72	0
KKM-PEN(0.01)	10	13	24	29	**30**	-	-	$7.16 \cdot 10^{-6}$	31.01	0
QUAD-PEN	0	0	0	7	26	**30**	-	$7.30 \cdot 10^{-6}$	31.59	0

Overall, the method QUAD-PEN achieved the best performance, producing the best solutions with relatively low computation times. The performance of QUAD-PEN is quite remarkable, as it computed for each of the 300 test instances, a point satisfying (16) with at least $\alpha = \frac{3}{2}$. Across all 300 test instances, the respective MIPGap of the solutions computed by Gurobi did not exceed ≈ 0.03595. The method SCHOL-REG also delivered strong results. However, we emphasize the robust and consistent very

Table 12: Comparison of solution methods for $S = 10$ and $n = 300$

METHOD	$\alpha = 0.01$	0.02	0.05	0.1	0.25	0.5	1.0	MIPGap / $\overline{mv}$	T	Failed
Gurobi	-	-	-	-	-	-	-	$6.03 \cdot 10^{-5}$	10.07	-
RELAX	0	0	0	0	0	0	0	$2.54 \cdot 10^{-14}$	1.49	0
SCHOL-REG	0	0	0	6	27	**30**	-	$8.06 \cdot 10^{-7}$	40.47	0
KS-REG	0	0	0	0	1	12	23	$7.46 \cdot 10^{-6}$	24.23	0
KKM-PEN(1)	0	0	3	9	11	22	29	$6.55 \cdot 10^{-6}$	48.80	0
KKM-PEN(0.1)	5	7	16	21	24	25	29	$7.17 \cdot 10^{-6}$	39.27	0
KKM-PEN(0.01)	6	11	19	27	29	**30**	-	$7.17 \cdot 10^{-6}$	28.36	0
QUAD-PEN	0	0	0	7	28	**30**	-	$7.30 \cdot 10^{-6}$	39.29	0

Table 13: Comparison of solution methods for $S = 20$ and $n = 300$

METHOD	$\alpha = 0.01$	0.02	0.05	0.1	0.25	0.5	1.0	MIPGap / $\overline{mv}$	T	Failed
Gurobi	-	-	-	-	-	-	-	$1.56 \cdot 10^{-3}$	313.20	-
RELAX	0	0	0	0	0	0	0	$6.66 \cdot 10^{-47}$	1.58	0
SCHOL-REG	0	0	2	16	27	**30**	-	$1.31 \cdot 10^{-6}$	52.56	0
KS-REG	0	0	0	0	5	17	20	$1.77 \cdot 10^{-4}$	31.27	1
KKM-PEN(1)	0	0	0	0	5	20	29	$7.11 \cdot 10^{-6}$	80.47	0
KKM-PEN(0.1)	0	0	1	4	8	16	25	$7.38 \cdot 10^{-6}$	62.67	0
KKM-PEN(0.01)	1	4	13	14	15	21	22	$6.88 \cdot 10^{-6}$	43.35	0
QUAD-PEN	0	0	5	19	**30**	-	-	$7.37 \cdot 10^{-6}$	47.58	0

Table 14: Comparison of solution methods for $S = 30$ and $n = 300$

METHOD	$\alpha = 0.01$	0.02	0.05	0.1	0.25	0.5	1.0	MIPGap / $\overline{mv}$	T	Failed
Gurobi	-	-	-	-	-	-	-	$9.87 \cdot 10^{-3}$	629.8	-
RELAX	0	0	0	0	0	0	0	$3.92 \cdot 10^{-50}$	1.60	0
SCHOL-REG	1	1	8	18	23	26	28	$6.43 \cdot 10^{-6}$	62.73	1
KS-REG	0	0	0	1	4	11	15	$6.76 \cdot 10^{-4}$	32.31	6
KKM-PEN(1)	0	0	0	0	5	12	21	$7.48 \cdot 10^{-6}$	105.98	0
KKM-PEN(0.1)	0	0	0	0	6	12	22	$6.97 \cdot 10^{-6}$	81.61	0
KKM-PEN(0.01)	0	1	5	6	6	15	20	$7.02 \cdot 10^{-6}$	65.26	0
QUAD-PEN	1	2	13	29	**30**	-	-	$7.43 \cdot 10^{-6}$	51.36	0

good performance of QUAD-PEN, which reliably outperformed SCHOL-REG. Additionally, also the generally robust regularization method SCHOLTES-REG failed to compute points which satisfy the maximal absolute complementarity constraint violation by 10^{-5} for some instances. This issue does not arise in the penalty methods. However, the KKM-PEN(λ) methods exhibit the following behaviour: up to a certain threshold of maximal absolute complementarity constraint violation and the corresponding penalty

Table 15: Comparison of solution methods for $S = 50$ and $n = 300$

METHOD	$\alpha = 0.01$	0.02	0.05	0.1	0.25	0.5	1.0	MIPGap / $\overline{mv}$	T	Failed
Gurobi	-	-	-	-	-	-	-	$1.45 \cdot 10^{-2}$	602.07	-
RELAX	0	0	0	0	0	0	0	$3.86 \cdot 10^{-50}$	1.63	2
SCHOL-REG	3	7	14	16	20	22	26	$7.95 \cdot 10^{-6}$	72.42	1
KS-REG	0	0	1	1	5	7	8	$9.81 \cdot 10^{-4}$	34.37	9
KKM-PEN(1)	1	1	1	1	1	4	9	$6.39 \cdot 10^{-6}$	136.58	0
KKM-PEN(0.1)	0	0	1	1	3	7	9	$7.62 \cdot 10^{-6}$	116.17	0
KKM-PEN(0.01)	0	0	0	1	3	11	18	$7.24 \cdot 10^{-6}$	112.50	0
QUAD-PEN	2	9	28	**30**	-	-	-	$7.03 \cdot 10^{-6}$	51.46	0

parameter, the methods compute high-quality solutions. That is, the cardinality constraint is fully utilized, leading to low objective values. Below this threshold, many entries of the solution vector become numerically negligible, i.e., their absolute value falls below 10^{-7}. As a result, effectively, the cardinality constraint is no longer fully exploited, resulting in worse objective values, since the cardinality constraint represents a critical resource.
Once again, we observe the combinatorial explosion when using Gurobi. At the same time, we note that Gurobi frequently computed good solutions in a short amount of time.

6 Conclusions

In this work, we introduced a general penalty method and a general regularization method for the continuous reformulation of cardinality-constrained optimization problems. Both approaches share the theoretical property that, under the CC-MFCQ, the generated sequence of KKT points converges to an M-stationary point, which, for the continuous reformulation of CCOPs, is essentially an S-stationary point. Furthermore, we analyzed the properties of the subproblems and showed that the general regularization method exhibits a property analogous to the Scholtes-type regularization, while the penalty method satisfies results analogous to those established for the MPCC penalty method. In the numerical experiments, we observed that the simple penalty term $\Phi(x, y) = \sum_{i=1}^{n} x_i^2 y_i^2$ achieves excellent performance, even surpassing the Scholtes-type regularization, which is known to be highly effective in practice.

Acknowledgement

The author expresses his sincere gratitude to Professor Wolfgang Achtziger (FAU Erlangen-Nürnberg) for his outstanding support, inspiring guidance, and many fruitful discussions.

References

[1] Bertsimas, D., King, A., Mazumder, R.: Best subset selection via a modern optimization lens. The Annals of Statistics **44**(2), 813–852 (2016) https://doi.org/10.1214/15-AOS1388

[2] Berk, L., Bertsimas, D.: Certifiably optimal sparse principal component analysis. Mathematical Programming Computation **11**(3), 381–420 (2019) https://doi.org/10.1007/s12532-018-0153-6

[3] Bertsimas, D., Cory-Wright, R.: A scalable algorithm for sparse portfolio selection. INFORMS Journal on Computing **34**(3), 1489–1511 (2022) https://doi.org/10.1287/IJOC.2021.1127

[4] Bienstock, D., Lodi, A., Schwartz, A., Tillmann, A.M.: Cardinality minimization, constraints, and regularization: A survey. SIAM Review **66**(3), 403–477 (2024) https://doi.org/10.1137/21M142770X

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

[6] Bienstock, D.: Computational study of a family of mixed-integer quadratic programming problems. Mathematical Programming **74**(2), 121–140 (1996) https://doi.org/10.1007/BF02592208

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

[8] Scheel, H., Scholtes, S.: Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. Mathematics of Operations Research **25**(1), 1–22 (2000)

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

[10] Achtziger, W., Kanzow, C.: Mathematical programs with vanishing constraints: optimality conditions and constraint qualifications. Mathematical Programming **114**(1), 69–99 (2008) https://doi.org/10.1007/s10107-006-0083-3

[11] Achtziger, W., Hoheisel, T., Kanzow, C.: A smoothing-regularization approach to mathematical programs with vanishing constraints. Computational Optimization and Applications **55**(3), 733–767 (2013) https://doi.org/10.1007/

s10589-013-9539-6

[12] Branda, M., Bucher, M., Červinka, M., Schwartz, A.: Convergence of a scholtes-type regularization method for cardinality-constrained optimization problems with an application in sparse robust portfolio optimization. Computational Optimization and Applications **70**(2), 503–530 (2018) https://doi.org/10.1007/s10589-018-9985-2

[13] Scholtes, S.: Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. SIAM Journal on Optimization **11**(4), 918–936 (2001) https://doi.org/10.1137/S1052623499361233

[14] Kanzow, C., Schwartz, A.: A new regularization method for mathematical programs with complementarity constraints with strong convergence properties. SIAM Journal on Optimization **23**(2), 770–798 (2013) https://doi.org/10.1137/100802487

[15] Bucher, M., Schwartz, A.: Second-order optimality conditions and improved convergence results for regularization methods for cardinality-constrained optimization problems. Journal of Optimization Theory and Applications **178**(2), 383–410 (2018) https://doi.org/10.1007/s10957-018-1320-7

[16] Kanzow, C., Raharja, A.B., Schwartz, A.: An augmented lagrangian method for cardinality-constrained optimization problems. Journal of Optimization Theory and Applications **189**(3), 793–813 (2021) https://doi.org/10.1007/s10957-021-01854-7

[17] Kanzow, C., Raharja, A.B., Schwartz, A.: Sequential optimality conditions for cardinality-constrained optimization problems with applications. Computational Optimization and Applications **80**(1), 185–211 (2021) https://doi.org/10.1007/s10589-021-00298-z

[18] Liang, Y.-C., Lin, G.-H.: Relaxed method for optimization problems with cardinality constraints. Journal of Global Optimization **88**(2), 359–375 (2024) https://doi.org/10.1007/s10898-023-01317-5

[19] Bucher, M.: Optimality conditions and numerical methods for a continuous reformulation of cardinality constrained optimization problems. PhD thesis, Technische Universität Darmstadt (2018). http://tuprints.ulb.tu-darmstadt.de/7740/

[20] Hu, X.M., Ralph, D.: Convergence of a penalty method for mathematical programming with complementarity constraints. Journal of Optimization Theory and Applications **123**(2), 365–390 (2004) https://doi.org/10.1007/s10957-004-5154-0

[21] Hoheisel, T., Kanzow, C., Schwartz, A.: Mathematical programs with vanishing

constraints: a new regularization approach with strong convergence properties. Optimization **61**(6), 619–636 (2012) https://doi.org/10.1080/02331934.2011.608164

[22] Galántai, A.: Properties and construction of ncp functions. Computational Optimization and Applications **52**(3), 805–824 (2012) https://doi.org/10.1007/s10589-011-9428-9

[23] Mangasarian, O.L.: Equivalence of the complementarity problem to a system of nonlinear equations. SIAM Journal on Applied Mathematics **31**(1), 89–92 (1976) https://doi.org/10.1137/0131009

[24] Fischer, A.: A special newton-type optimization method. Optimization **24**(3-4), 269–284 (1992) https://doi.org/10.1080/02331939208843795

[25] Kanzow, C., Kleinmichel, H.: A new class of semismooth newton-type methods for nonlinear complementarity problems. Computational Optimization and Applications **11**(3), 227–251 (1998) https://doi.org/10.1023/A:1026424918464

[26] Gurobi Optimization, LLC: Gurobi Optimizer Reference Manual (2025). https://www.gurobi.com

[27] Wächter, A.: An interior point algorithm for large-scale nonlinear optimization with applications in process engineering. PhD thesis, Carnegie Mellon University (2002)

[28] Wächter, A., Biegler, L.T.: On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Mathematical Programming **106**(1), 25–57 (2006) https://doi.org/10.1007/s10107-004-0559-y

[29] Bynum, M.L., Hackebeil, G.A., Hart, W.E., Laird, C.D., Nicholson, B.L., Siirola, J.D., Watson, J.-P., Woodruff, D.L.: Pyomo–optimization Modeling in Python vol. 67, 3rd edn. (2021). https://doi.org/10.1007/978-3-030-68928-5

[30] Hart, W.E., Watson, J.-P., Woodruff, D.L.: Pyomo: modeling and solving mathematical programs in python. Mathematical Programming Computation **3**(3), 219–260 (2011) https://doi.org/10.1007/s12532-011-0026-8

[31] Amestoy, P.R., Duff, I.S., L'Excellent, J.-Y., Koster, J.: A fully asynchronous multifrontal solver using distributed dynamic scheduling. SIAM Journal on Matrix Analysis and Applications **23**(1), 15–41 (2001) https://doi.org/10.1137/S0895479899358194

[32] Amestoy, P.R., Buttari, A., L'Excellent, J.-Y., Mary, T.: Performance and scalability of the block low-rank multifrontal factorization on multicore architectures. ACM Trans. Math. Softw. **45**, 2–1226 (2019) https://doi.org/10.1145/3242094

[33] Markowitz, H.: Portfolio selection. The Journal of Finance **7**(1), 77–91 (1952) https://doi.org/10.2307/2975974

[34] Antonio, F., Gentile, C.: Sdp diagonalizations and perspective cuts for a class of nonseparable miqp. Oper. Res. Lett. **35**, 181–185 (2007) https://doi.org/10.1016/j.orl.2006.03.008

[35] Antonio, F., Gentile, C.: Mean-Variance problem with minimum buy-in and cardinality constraints (2024). http://groups.di.unipi.it/optimize/Data/MV.html Accessed 2025-04-06