Journal of Optimization Theory and Applications (2026) 209:85
https://doi.org/10.1007/s10957-026-03029-8

A Lasry–Lions Envelope Approach for Mathematical Programs with Complementarity Constraints

Jia Wang[1] · Andreas Themelis[2] · Ivan Markovsky[3,4] · Panagiotis Patrinos[1]

Received: 7 July 2025 / Accepted: 14 May 2026 / Published online: 5 June 2026
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2026

Abstract

We propose a homotopy method for solving mathematical programs with complementarity constraints (CCs). The indicator function of the CCs is relaxed by the Lasry–Lions double envelope, an extension of the Moreau envelope that enjoys an additional smoothness property, making it amenable to fast optimization algorithms. The proposed algorithm mimics the behavior of homotopy methods for systems of nonlinear equations or penalty methods for constrained optimization: it solves a sequence of smooth subproblems that progressively approximate the original problem, using the solution of each subproblem as the starting point for the next one. In the limiting setting, we establish the convergence to Mordukhovich and Clarke stationary points. We also provide a worst-case complexity analysis for computing an approximate stationary point. Preliminary numerical results on a suite of benchmark problems demonstrate the effectiveness of the proposed approach.

Communicated by Sofia Giuffre.

✉ Jia Wang
jia.wang@esat.kuleuven.be

Andreas Themelis
andreas.themelis@ees.kyushu-u.ac.jp

Ivan Markovsky
imarkovsky@cimne.upc.edu

Panagiotis Patrinos
panos.patrinos@esat.kuleuven.be

1 Department of Electrical Engineering (ESAT-STADIUS), KU Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium

2 Faculty of Information Science and Electrical Engineering (ISEE), Kyushu University, 744 Motooka, 819-0395 Nishi-ku, Fukuoka, Japan

3 International Centre for Numerical Methods in Engineering (CIMNE), Gran Capità, 08034 Barcelona, Spain

4 Catalan Institution for Research and Advanced Studies (ICREA), Pg. Lluis Companys 23, 08010 Barcelona, Spain

Keywords Mathematical programs with complementarity constraints · Constraint qualifications · Smoothing method · Homotopy method · Worst-case complexity

Mathematics Subject Classification 65K05 · 90C06 · 90C25 · 90C30

1 Introduction

We consider the following constrained mathematical program with complementarity constraints (MPCC):

$$\begin{aligned} &\underset{x\in C}{\text{minimize}} \quad f(x) \\ &\text{subject to} \quad G(x) \geq 0,\ H(x) \geq 0,\ \langle G(x), H(x)\rangle = 0, \end{aligned} \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth function, and $G, H : \mathbb{R}^n \to \mathbb{R}^p$ are smooth mappings. The set $C \subseteq \mathbb{R}^n$ is nonempty, closed and convex. We focus on smooth objective and constraint functions, as all the essential difficulties are associated with the complementarity structure. Additionally, we assume that the solution set of problem (1.1) is nonempty.

MPCC is a class of important problems as they arise in many engineering and economics applications. Examples include multi-rigid-body contact problems in robotic control [39], obstacle avoidance in autonomous driving [19], and policy modeling for price and subsidy regulation [36]; see [10] for a broader overview of applications. In addition, many practical applications are formulated as bilevel optimization problems, including Stackelberg games that represent hierarchical market structures [8, 49]. In these problems, the feasible set is defined as the solution set of a parametric lower-level optimization problem. By replacing the lower-level problem with its Karush-Kuhn-Tucker (KKT) optimality conditions, the bilevel problem can be transformed into an MPCC.

Although MPCCs have broad applications, they are challenging both theoretically and numerically: it is well known that standard constraint qualifications (CQs) such as the Mangasarian-Fromovitz constraint qualification (MFCQ) and the linear independence constraint qualification (LICQ) are violated at all feasible points of the complementarity constraints [11, Proposition 2.15]. As a result, the set of Lagrange multipliers at a stationary point may be unbounded [23, 34]. Nevertheless, notable theoretical progress has been made, including the development of stationarity concepts such as strongly (S) [44], Mordukhovich (M) [38], Clarke (C) [44], alternative (A) [12, 13], and weakly (W) [44] stationary points. Meanwhile, a variety of MPCC-tailored CQs have been developed, such as MPCC-LICQ [28, Definition 3(a)], MPCC-MFCQ [28, Definition 3(b)], MPCC-constant rank CQ (CRCQ) [28, Definition 3(c)], MPCC-constant positive linear dependence CQ (CPLD) [28, Definition 3(d)], MPCC-Abadie CQ [13, Definition 3.1], MPCC-Guignard CQ [14, Definition 3.1], among others. A comprehensive overview of further MPCC-CQs can be found in [48], while an in-depth review of stationarity concepts and optimality conditions is provided in the PhD thesis [11].

On the algorithmic side, many numerical methods have been developed to solve MPCCs. One of the most widely used approaches for handling complementarity constraints is the regularization method, which relaxes the constraint $G_i(x)H_i(x) = 0$ to $G_i(x)H_i(x) \leq \tau$ for all $i \in \{1, \ldots, p\}$ and solves a sequence of regularized subproblems as $\tau \searrow 0$; see [45]. Other relaxation strategies for $G_i(x)H_i(x) = 0$ have also been proposed; see [20, 26–28]. An alternative way is the penalty method, which incorporates the orthogonality constraints to the objective via a penalty term, i.e., $f(x) + \rho\langle G(x), H(x)\rangle$, and solves a sequence of subproblems with increasing $\rho > 0$; see [2, 21, 41]. In addition, other types of MPCC-tailored algorithms have been introduced, including interior point [34, 40], active-set [22], and augmented Lagrangian methods [24], all of which have demonstrated good numerical performance. Furthermore, nonlinear programming (NLP) solvers have also been successfully applied to MPCCs; see [15, 23]. Most of methods in the literature analyze convergence in the limiting setting, showing that accumulation points satisfy one of the S/M/C-stationarity conditions under appropriate assumptions. However, the methods with complexity guarantees, which aim to find an approximate stationary point within a finite number of steps, have not yet been fully studied in the literature. Moreover, many existing methods often involve solving general nonlinear subproblems or rely on complex parameter tuning strategies, making them less attractive for practical implementation. Motivated by these observations, the goal of this paper is to develop an algorithm for MPCCs that is both easy to implement and equipped with complexity guarantees.

Recently, a quadratic penalty method for nonconvex optimization was proposed in [18], which features both simple subproblem structure and straightforward parameter tuning. Therefore, we leverage this framework to design our algorithm for MPCCs and establish its complexity guarantees. It is important to note, however, that the quadratic penalty for complementarity constraints corresponds to the Moreau envelope of their indicator function, which is nonsmooth. To address this nonsmoothness, we propose using a surrogate: the Lasry–Lions double envelope [3]. The subproblems derived from the Lasry–Lions double envelope are smooth, and thus amenable to efficient smooth optimization techniques, such as fast gradient-based solvers [17] or L-BFGS-B [50]. To the best of our knowledge, the use of the Lasry–Lions double envelope is new. The only closely related work is the recent paper [46], which applied this tool in nonsmooth nonconvex optimization for signal processing.

1.1 Contribution

The main contribution of this work is the development of a novel homotopy-based algorithm for MPCCs, leveraging the Lasry–Lions double envelope to construct a smooth approximation of the complementarity constraints, which pose particular challenges due to their inherent nonsmoothness, nonconvexity, and lack of standard CQs. Second, we introduce novel MPCC-tailored CQs that are weaker than the widely used MPCC-MFCQ, under which we establish the convergence to C/M-stationary points. Third, we propose a practical finite-iteration version of the algorithm that is easy to implement, and we provide a rigorous worst-case complexity analysis for this variant. Finally, we validate the effectiveness and robustness of the proposed method through

extensive numerical experiments on MPCC benchmark problems, and show that it is competitive with state-of-the-art NLP solvers and the classical regularization method [45] in both solution quality and computational efficiency. In particular, the proposed method is capable of efficiently handling large-scale and complex complementarity constraints, making it a promising tool for practical applications in engineering and economics.

1.2 Notation

The set of natural numbers is $\mathbb{N} := \{1, 2, \ldots\}$. The *extended real line* is denoted by $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$. The set of nonnegative and strictly positive reals are $\mathbb{R}_+ := [0, \infty)$ and $\mathbb{R}_{++} := (0, \infty)$. We denote the positive part of a real number by $[\cdot]_+ := \max\{0, \cdot\}$. For $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$, the *domain* is defined as $\operatorname{dom}\varphi := \{x \in \mathbb{R}^n \mid \varphi(x) < \infty\}$, the *conjugate* is given by $\varphi^*(y) := \sup_x\{\langle y, x\rangle - \varphi(x)\}$ and the *biconjugate* is defined as $\varphi^{**}(x) := \sup_y\{\langle y, x\rangle - \varphi^*(y)\}$. Given $\varepsilon > 0$, we denote $\varepsilon\text{-}\arg\min\varphi := \{x \in \mathbb{R}^n \mid \varphi(x) \leq \inf\varphi + \varepsilon\}$. The collection of subsequences of $\mathbb{N}$ is $\mathcal{N}_\infty^\sharp$. The *convex hull* of a set $E \subseteq \mathbb{R}^n$ is denoted by $\operatorname{conv} E$, and its *interior* by $\operatorname{int} E$. The notation D always stands for the *complementarity set* in this work, i.e.,

$$D := \{z \in \mathbb{R}^2_+ \mid z_1 z_2 = 0\} \subset \mathbb{R}^2. \tag{1.2}$$

We denote the ith complementarity constraint pair by $F_i := (G_i, H_i) : \mathbb{R}^n \to \mathbb{R}^2$, so that the complementarity constraints can be expressed by $F_i(x) \in D$ for all $i \in \{1, \ldots, p\}$. The *Jacobian* of F_i at x is denoted by $JF_i(x)$, i.e.,

$$JF_i(x) = \begin{pmatrix} \nabla G_i(x)^\top \\ \nabla H_i(x)^\top \end{pmatrix} \in \mathbb{R}^{2\times n}.$$

The *Euclidean distance* between a point z and the set D is denoted $\operatorname{dist}(z, D) := \min_{z'\in D}\|z' - z\|$ and the (possibly set-valued) *Euclidean projection* onto the set D is $\Pi_D(z) := \arg\min_{z'\in D}\|z' - z\|$. The closed ball with center at $x \in \mathbb{R}^n$ and radius ξ is $\mathrm{B}(x, \xi) := \{x' \in \mathbb{R}^n \mid \|x' - x\| \leq \xi\}$.

2 Preliminaries and Supporting Results

In this section, we begin by reviewing key concepts from variational and nonsmooth analysis, which form the mathematical foundation for our discussion. We then introduce several stationarity concepts commonly used in the MPCC literature. Finally, we review some MPCC-tailored CQs, and use them as the foundation to propose a novel and weaker variant. The notation and terminology are primarily based on the works of [7, 20, 24].

2.1 Fundamentals of Variational Analysis

The *epigraph* of $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ is the set $\operatorname{epi}\varphi := \{(x,\alpha) \in \mathbb{R}^{n+1} \mid \varphi(x) \le \alpha\}$, and φ is said to be *lower semicontinuous (lsc)* if $\operatorname{epi}\varphi$ is closed, and proper if $\operatorname{epi}\varphi$ is nonempty and $\varphi(x) > -\infty$ for all $x \in \mathbb{R}^n$. For a sequence of sets $\{C^\nu\}$ with $C^\nu \subseteq \mathbb{R}^n$ for all $\nu \in \mathbb{N}$, we define the *outer limit* as

$$\operatorname*{Lim\,sup}_{\nu\to\infty} C^\nu := \{x \mid \exists K \in \mathcal{N}_\infty^\sharp,\ \{x^\nu\} \to_K x : x^\nu \in C^\nu\ \ \forall \nu \in K\}$$

and the *inner limit* as

$$\operatorname*{Lim\,inf}_{\nu\to\infty} C^\nu := \{x \mid \exists k \in \mathbb{N},\ \{x^\nu\} \to x : x^\nu \in C^\nu\ \ \forall \nu \ge k\}.$$

We say that $\{C^\nu\}$ *set-converges* if its outer limit and inner limit are equal, i.e.,

$$\operatorname*{Lim}_{\nu\to\infty} C^\nu := \operatorname*{Lim\,sup}_{\nu\to\infty} C^\nu = \operatorname*{Lim\,inf}_{\nu\to\infty} C^\nu.$$

We say that a sequence $\{\varphi^\nu\}$ of functions $\varphi^\nu : \mathbb{R}^n \to \overline{\mathbb{R}}$ *epi-converges* to $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ if

$$\operatorname*{Lim}_{\nu\to\infty} \operatorname{epi}\varphi^\nu = \operatorname{epi}\varphi,$$

which is denoted by $\varphi^\nu \xrightarrow{e} \varphi$. For set-valued mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, the *graph* of S is the set $\operatorname{gph} S := \{(x,u) \in \mathbb{R}^{n+m} \mid u \in S(x)\}$. We say that a sequence $\{S^\nu\}$ of set-valued mappings $S^\nu : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *graphically converges* to $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ if

$$\operatorname*{Lim}_{\nu\to\infty} \operatorname{gph} S^\nu = \operatorname{gph} S,$$

which is denoted by $S^\nu \xrightarrow{g} S$. Considering the complementarity set D, its *indicator function* and Euclidean projection are respectively given by

$$\delta_D(z) = \begin{cases} 0 & \text{if } z \in D \\ +\infty & \text{otherwise} \end{cases} \quad \text{and} \quad \Pi_D(z) = \begin{cases} \Pi_{\mathbb{R}^2_+}(z) & \text{if } z \notin \mathbb{R}^2_{++} \\ (z_1, 0) & \text{if } z_1 > z_2 > 0 \\ (0, z_2) & \text{if } z_2 > z_1 > 0 \\ \{(z_1,0),(0,z_2)\} & \text{if } z_1 = z_2 > 0. \end{cases}$$

We next review some cone definitions for a set $E \subseteq \mathbb{R}^n$, which are important for characterizing different stationarity conditions. The *tangent cone* at a point $x \in E$ is

$$T_E(x) := \{d \in \mathbb{R}^n \mid \exists E \ni x^\nu \to x \text{ and } t^\nu \searrow 0 \text{ such that } \tfrac{x^\nu - x}{t^\nu} \to d\}.$$

The *regular normal cone* of E at x is defined as

$$\hat{N}_E(x) := \{d \in \mathbb{R}^n \mid \langle d, d' \rangle \le 0\ \ \forall d' \in T_E(x)\}.$$

The *limiting normal cone* of E at x is defined as

$$N_E(x) := \{d \in \mathbb{R}^n \mid \exists E \ni x^\nu \to x \text{ and } d^\nu \in \hat{N}_E(x^\nu) \text{ such that } d^\nu \to d\}.$$

As for a convex set C, we have $\hat{N}_C(x) = N_C(x)$ for all $x \in C$. Based on the above definitions, the regular normal cone and the limiting normal cone of the complementarity set D at $z \in D$ are explicitly given by

$$\hat{N}_D(z) = \begin{cases} \{0\} \times \mathbb{R} & \text{if } z_1 > 0 \\ \mathbb{R} \times \{0\} & \text{if } z_2 > 0 \\ \mathbb{R}^2_- & \text{if } z = (0,0) \end{cases} \quad \text{and} \quad N_D(z) = \begin{cases} \{0\} \times \mathbb{R} & \text{if } z_1 > 0 \\ \mathbb{R} \times \{0\} & \text{if } z_2 > 0 \\ \mathbb{R}^2_- \cup D & \text{if } z = (0,0). \end{cases}$$

As a set-valued mapping, note that the limiting normal cone is *outer semicontinuous*, in the sense that $\operatorname{Lim\,sup}_{z^\nu \to z} N_D(z^\nu) \subseteq N_D(z)$, and has better calculus properties compared to the regular normal cone. Furthermore, we introduce additional cone concepts as

$$N_D^{\mathrm{C}}(z) := \begin{cases} \hat{N}_D(z) & \text{if } z \neq (0,0) \\ N_D((0,0)) \cup \mathbb{R}^2_+ & \text{if } z = (0,0) \end{cases}$$

and

$$N_D^{\mathrm{W}}(z) := \begin{cases} \hat{N}_D(z) & \text{if } z \neq (0,0) \\ \mathbb{R}^2 & \text{if } z = (0,0). \end{cases}$$

We will refer to the cone N_D^{C} as the *Clarke limiting cone*, since it is related to the Clarke-stationarity in MPCCs. Note that N_D^{C} is different from the *Clarke normal cone* in [42, Exercise 6.38]. The cone N_D^{W} is referred to as the *weak limiting cone*, which will be used to formulate a generalized version of MPCC-MFCQ (see Definition 2.7).

2.2 Stationarity Concepts

To each of these cones, $\hat{N}_D$, N_D and N_D^{C}, there corresponds one of the following stationarity concepts for MPCC problems.

Definition 2.1 *(Stationary point)* Relative to (1.1), a feasible point $x \in \mathbb{R}^n$ is said to be *strongly stationary (S-stationary)* if there exist $y_1, \ldots, y_p \in \mathbb{R}^2$ such that

$$\nabla f(x) + \sum_{i=1}^{p} JF_i(x)^\top y_i + N_C(x) \ni 0$$
$$y_i \in \hat{N}_D(F_i(x))$$

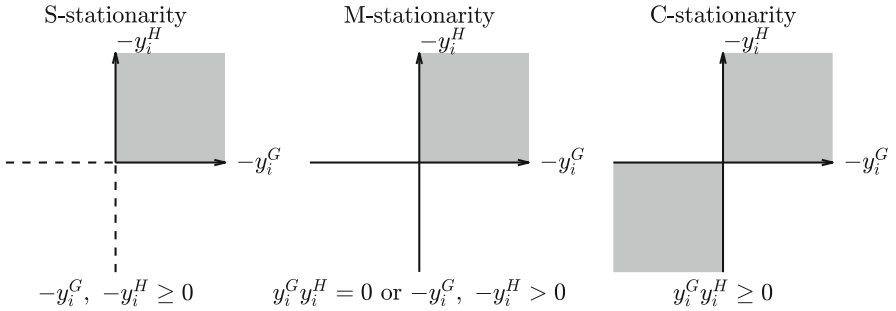


Fig. 1 Multipliers of the MPCC-Lagrangian associated with biactive constraints $G_i(x) = H_i(x) = 0$ for different stationary points

for all $i \in \{1, \ldots, p\}$, or equivalently, if it satisfies

$$-\nabla f(x) \in \sum_{i=1}^{p} JF_i(x)^\top \hat{N}_D(F_i(x)) + N_C(x).$$

If the above condition holds with $\hat{N}_D(F_i(x))$ being replaced by $N_D(F_i(x))$ for all $i \in \{1, \ldots, p\}$, then we say that x is *Mordukhovich stationary (M-stationary)*. Furthermore, if $\hat{N}_D(F_i(x))$ is instead replaced by $N_D^{\mathrm{C}}(F_i(x))$ for all $i \in \{1, \ldots, p\}$, then x is said to be *Clarke stationary (C-stationary)*.

In the literature, $-y_i$ typically serves as the multiplier associated with the *MPCC-Lagrangian* $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^{2p} \to \mathbb{R}$, given by

$$\mathcal{L}(x, \boldsymbol{y}) = f(x) + \sum_{i=1}^{p} y_i^G G_i(x) + \sum_{i=1}^{p} y_i^H H_i(x),$$

in which $\boldsymbol{y} = (y_1, \ldots, y_p) \in \mathbb{R}^{2p}$ and $y_i = (y_i^G, y_i^H) \in \mathbb{R}^2$; see [23]. With this notation, the stationarity concepts introduced in Definition 2.1 are consistent with the standard definitions in [20, Definition 2.3]. Each of these stationarity concepts treats differently the signs of the multipliers associated with *biactive* complementary constraint, that is, those satisfying $G_i(x) = H_i(x) = 0$; see Fig. 1 for an illustration. In contrast, when the solution satisfies either $G_i(x) > 0, H_i(x) = 0$ or $G_i(x) = 0, H_i(x) > 0$ for all $i \in \{1, \ldots, p\}$, the various stationarity conditions are equivalent. Note that the S-stationarity is equivalent to the standard KKT conditions of an MPCC, since there are no feasible descent directions [20]. On the other hand, for C/M-stationary points, feasible descent directions may exist. From this point onward, we will refer to y_i as the multiplier associated with the generalized equations in Definition 2.1. We now introduce the concept of asymptotic M-stationarity.

Definition 2.2 *(AM-stationary point [24, Definition 2.2])* Relative to (1.1), a feasible point $x \in \mathbb{R}^n$ is said to be *asymptotic Mordukhovich stationary (AM-stationary)* if

there exist sequences $\{(x^\nu, \boldsymbol{y}^\nu, \boldsymbol{z}^\nu)\}$ such that

$$\begin{aligned}&\nabla f(x^\nu) + \textstyle\sum_{i=1}^p JF_i(x^\nu)^\top y_i^\nu + N_C(x^\nu) \to 0\\ &(x^\nu, z_i^\nu) \to (x, F_i(x))\\ &y_i^\nu \in N_D(z_i^\nu)\end{aligned}$$

for all $i \in \{1, \ldots, p\}$, where $\boldsymbol{y}^\nu = (y_1^\nu, \ldots, y_p^\nu) \in \mathbb{R}^{2p}$ and $\boldsymbol{z}^\nu = (z_1^\nu, \ldots, z_p^\nu) \in \mathbb{R}^{2p}$.

Clearly, any point that is M-stationary is also AM-stationary, as it corresponds to the constant sequence $(x^\nu, y_i^\nu, z_i^\nu) = (x, y_i, F_i(x))$ for all $i \in \{1, \ldots, p\}$, and the converse holds as long as $\{\boldsymbol{y}^\nu\}$ is bounded [24]. Note that because of the continuity of ∇f and JF_i, there is no loss of generality in simplifying the above definition by taking $z_i^\nu = F_i(x^\nu)$. The extra variable z_i^ν is however convenient as it allows infeasible point x^ν, in the sense that $F_i(x^\nu) \in D$ need not hold.

2.3 Constraint Qualifications

To guarantee that a local minimizer of the MPCC (1.1) is stationary in the sense of Definition 2.1, special MPCC-tailored CQs are used; in the literature, MPCC-MFCQ is among the most widely used ones. For example, the methods proposed in [35, 45] are shown to converge to C-stationary points under this CQ; see also [20, Theorems 3.1 and 3.3]. Let us consider general constraints

$$\begin{aligned}&G_i(x) \geq 0,\ H_i(x) \geq 0,\ G_i(x)H_i(x) = 0,\ i \in \{1, \ldots, p\}\\ &c_i(x) = 0,\ i \in \mathcal{E}\\ &c_i(x) \leq 0,\ i \in \mathcal{I}\end{aligned} \tag{2.1}$$

and define some crucial index sets at a feasible point x of (2.1):

$$\begin{aligned}I_{\mathcal{I}}(x) &:= \{i \in \mathcal{I} \mid c_i(x) = 0\},\\ I_{0+}(x) &:= \{i \in \{1, \ldots, p\} \mid G_i(x) = 0,\ H_i(x) > 0\},\\ I_{+0}(x) &:= \{i \in \{1, \ldots, p\} \mid G_i(x) > 0,\ H_i(x) = 0\},\\ I_{00}(x) &:= \{i \in \{1, \ldots, p\} \mid G_i(x) = 0,\ H_i(x) = 0\}.\end{aligned} \tag{2.2}$$

For simplicity we may omit the dependence on x in the index sets above, if there is no confusion. We are now in a position to recall the conventional MPCC-MFCQ needed.

Definition 2.3 *(MPCC-MFCQ [28, Definition 3(b)])* Let $x \in C$ be feasible for (2.1). Then, the *MPCC-Mangasarian-Fromovitz constraint qualification (MPCC-MFCQ)* holds at x if whenever

$$\begin{aligned}&\alpha_i \in \mathbb{R}_+,\ \pi_i \in \mathbb{R},\ (y_i^G, y_i^H) \in \mathbb{R}^2\\ &\textstyle\sum_{i \in I_{\mathcal{I}}} \alpha_i \nabla c_i(x) + \sum_{i \in \mathcal{E}} \pi_i \nabla c_i(x)\\ &+ \textstyle\sum_{i \in I_{0+} \cup I_{00}} y_i^G \nabla G_i(x) + \sum_{i \in I_{+0} \cup I_{00}} y_i^H \nabla H_i(x) + N_C(x) \ni 0\end{aligned}$$

it holds that

$$\begin{aligned}&\alpha_i = 0,\ i \in I_{\mathcal{I}}\\ &\pi_i = 0,\ i \in \mathcal{E}\\ &y_i^G = 0,\ i \in I_{0+} \cup I_{00}\\ &y_i^H = 0,\ i \in I_{+0} \cup I_{00}.\end{aligned}$$

Note that the conventional MPCC-MFCQ is not formulated within the framework of general composite constraints of the form $\delta_E(h(x))$, where E can represent sets such as $\mathbb{R}_+$, $\{0\}$ or D, and h may correspond to constraint functions like c_i or F_i. To leverage the powerful tools of variational analysis, the constraints in (2.1) can be reformulated within the composite framework. In this framework, a widely used concept of CQ is the *basic constraint qualification (BCQ)*, as presented in [7, Theorem 5.5] and [42, Theorem 6.14], which is also called *no nonzero abnormal multiplier CQ* or *generalized MFCQ* in the literature [24]. Accordingly, for the feasible set defined in (2.1), we introduce the BCQ as follows:

Definition 2.4 *(BCQ [42, Theorem 6.14])* Let $x \in C$ be feasible for (2.1). Then, the *basic constraint qualification (BCQ)* holds at x if whenever

$$\begin{aligned}&\alpha_i \in N_{\mathbb{R}_-}(c_i(x)),\ \pi_i \in \mathbb{R},\ y_i \in N_D(F_i(x))\\ &\textstyle\sum_{i\in\mathcal{I}} \alpha_i \nabla c_i(x) + \sum_{i\in\mathcal{E}} \pi_i \nabla c_i(x) + \sum_{i=1}^p JF_i(x)^\top y_i + N_C(x) \ni 0\end{aligned}$$

it holds that

$$\begin{aligned}&\alpha_i = 0,\ i \in \mathcal{I}\\ &\pi_i = 0,\ i \in \mathcal{E}\\ &y_i = (0,0),\ i \in \{1,\dots,p\}.\end{aligned}$$

Both MPCC-MFCQ and BCQ handle different types of constraints separately, lacking a unified treatment. Motivated by this observation, one of the main goals of this paper is to introduce a new algorithm that handles all constraints in (2.1) in a unified manner. To this end, we begin with the following definition and example:

Definition 2.5 *(Well-behaved CC)* $F_i := (G_i, H_i) : \mathbb{R}^n \to \mathbb{R}^2$ is said to be a *well-behaved complementarity constraint (well-behaved CC)* if there exists $\xi \in \mathbb{R}_{++}$ such that $F_i(x) \notin \mathrm{B}(0,\xi) \cap \mathbb{R}^2_{++}$ for every $x \in \mathbb{R}^n$.

The appeal of well-behaved CCs in avoiding the diagonal in the positive orthant close to the origin will be apparent in Lemma 4.2.

Example 2.6 *(Equality and inequality constraints are well-behaved)* Consider equality and inequality constraints, and their CC reformulation:

$$\begin{aligned}&c_i(x) = 0,\ i \in \mathcal{E}\\ &c_i(x) \le 0,\ i \in \mathcal{I}\end{aligned} \quad \Leftrightarrow \quad F_i(x) = \begin{cases}(c_i(x), a), & i \in \mathcal{E}\\ (-c_i(x), 0), & i \in \mathcal{I}\end{cases} \text{ with } F_i(x) \in D,$$

where $a > 0$ is a given constant. Then, F_i is a well-behaved CC for all $i \in \mathcal{E} \cup \mathcal{I}$. Specifically, when $i \in \mathcal{E}$, any $\xi \in (0, a)$ is such that $F_i(x) \notin \mathrm{B}(0,\xi)$ for all $x \in \mathbb{R}^n$.

Therefore, F_i is a well-behaved CC according to Definition 2.5. When $i \in \mathcal{I}$, it is evident that $F_i(x)$ never lies in $\mathbb{R}^2_{++}$, which implies that F_i is well-behaved. It is worth noting that there are several alternative ways to reformulate the equality constraint. For example, one could use $F_i = (c_i(x), -c_i(x)) \in D$, or introduce two separate components: $F_{i^+} = (c_i(x), 0) \in D$ and $F_{i^-} = (-c_i(x), 0) \in D$. All of these formulations result in well-behaved CCs, as none of the corresponding $F_i(x)$ lies in $\mathbb{R}^2_{++}$. In the experiments presented in Section 6, we use the F_{i^+}, F_{i^-} reformulation, as it demonstrates better numerical performance compared to the alternatives.

The concept of well-behaved CCs is particularly useful when equality and inequality constraints are reformulated as CCs. Through this reformulation, optimization problems with constraints of the form (2.1) can be uniformly recast into the structure of (1.1). Building on the notion of well-behaved CCs, we now introduce a generalized version of MPCC-MFCQ :

Definition 2.7 *(MPCC-MFCQ$^+$)* Relative to (1.1), let $\mathcal{W} \subseteq \{1, \ldots, p\}$ be the index set of well-behaved CCs and $y_i = (y_i^G, y_i^H) \in \mathbb{R}^2$ for $i \in \{1, \ldots, p\}$. Then, *MPCC-MFCQ$^+$* holds at a feasible point x if whenever

$$\begin{aligned} &y_i \in N_D(F_i(x)), \ i \in \mathcal{W} \\ &y_i \in N_D^{\mathrm{W}}(F_i(x)), \ i \notin \mathcal{W} \\ &\textstyle\sum_{i=1}^p JF_i(x)^\top y_i + N_C(x) \ni 0 \end{aligned}$$

one has that

$$\begin{aligned} &\text{either } G_i \text{ is constant or } y_i^G = 0, \\ &\text{either } H_i \text{ is constant or } y_i^H = 0, \\ &i \in \{1, \ldots, p\}. \end{aligned}$$

The statement of constant G_i and H_i in MPCC-MFCQ$^+$ may seem unusual at first glance, but it arises naturally from the reformulation of equality and inequality constraints as CCs. Specifically, when such constraints are reformulated as CCs following Example 2.6, one may have $F_i = (G_i, H_i)$ for $i \in \mathcal{W}$, where $G_i(x) = c_i(x)$ and $H_i(x) = a \geq 0$. Since H_i is a constant mapping, it holds that

$$JF_i(x)^\top y_i = \left(\nabla c_i(x), 0\right) \begin{pmatrix} y_i^G \\ y_i^H \end{pmatrix}.$$

Therefore, the multiplier y_i^H corresponding to such H_i is excluded from the requirement of being zero, as it is always annihilated by the zero block in the Jacobian.

It is important to note that when $\mathcal{W} = \mathcal{E} \cup \mathcal{I}$, MPCC-MFCQ$^+$ is equivalent to MPCC-MFCQ . In contrast, when $\mathcal{W} \supset \mathcal{E} \cup \mathcal{I}$, MPCC-MFCQ$^+$ is weaker than MPCC-MFCQ , since the region of the multiplier y_i for $i \in \mathcal{W} \setminus (\mathcal{E} \cup \mathcal{I})$ is restricted from $N_D^{\mathrm{W}}(F_i(x))$ to $N_D(F_i(x))$. Similar to MPCC-MFCQ$^+$, by leveraging the concept of well-behaved CCs, we can modify BCQ accordingly. In this case, however, rather than obtaining a weaker CQ, we obtain an equivalent condition, but with a much simplified and condensed statement.

Definition 2.8 *(BCQ^+)* Relative to (1.1), let $y_i = (y_i^G, y_i^H) \in \mathbb{R}^2$ for $i \in \{1, \ldots, p\}$. Then, BCQ^+ holds at a feasible point x if whenever

$$\begin{aligned} &y_i \in N_D(F_i(x)),\ i \in \{1, \ldots, p\} \\ &\textstyle\sum_{i=1}^p JF_i(x)^\top y_i + N_C(x) \ni 0 \end{aligned}$$

one has that

$$\begin{aligned} &\text{either } G_i \text{ is constant or } y_i^G = 0, \\ &\text{either } H_i \text{ is constant or } y_i^H = 0,\ i \in \{1, \ldots, p\}. \end{aligned}$$

The equivalence of BCQ^+ and BCQ stems from the fact that they both involve only limiting normal cones. Building on BCQ^+, we now introduce our main MPCC-tailored CQ as follows, which will serve as a key tool in the convergence analysis of our proposed algorithm.

Definition 2.9 *(BCCQ)* Relative to (1.1), let $\mathcal{W} \subseteq \{1, \ldots, p\}$ be the index set of well-behaved CCs and $y_i = (y_i^G, y_i^H) \in \mathbb{R}^2$ for $i \in \{1, \ldots, p\}$. Then, the *basic Clarke constraint qualification (BCCQ)* holds at a feasible point x if whenever

$$\begin{aligned} &y_i \in N_D(F_i(x)),\ i \in \mathcal{W} \\ &y_i \in N_D^{\mathrm{C}}(F_i(x)),\ i \notin \mathcal{W} \\ &\textstyle\sum_{i=1}^p JF_i(x)^\top y_i + N_C(x) \ni 0 \end{aligned}$$

one has that

$$\begin{aligned} &\text{either } G_i \text{ is constant or } y_i^G = 0, \\ &\text{either } H_i \text{ is constant or } y_i^H = 0,\ i \in \{1, \ldots, p\}. \end{aligned}$$

Applying the methods from Example 2.6 to reformulate (2.1) entirely into CCs reveals the following chain of implications among these MPCC-tailored CQs:

$$\text{MPCC-MFCQ} \implies \text{MPCC-MFCQ}^+ \implies \text{BCCQ} \implies \text{BCQ}^+ \iff \text{BCQ}\ . \tag{2.3}$$

Beyond the previous discussion, this chain shows that BCCQ is implied by MPCC-MFCQ^+. Specifically, BCCQ restricts the region of the multiplier y_i from $N_D^{\mathrm{W}}(F_i(x))$ to $N_D^{\mathrm{C}}(F_i(x))$ for $i \notin \mathcal{W}$, while keeping the region of the multiplier y_i for $i \in \mathcal{W}$ unchanged. Consequently, if MPCC-MFCQ^+ holds at a feasible point x, then BCCQ is necessarily satisfied as well.

3 Lasry–Lions Double Envelope

We now introduce the fundamental concepts and key properties of the Lasry–Lions double envelope. This construction is derived from the Moreau envelope and serves as a tool for smoothing the given function.

Definition 3.1 *(Moreau envelope)* Given $\lambda > 0$, the *Moreau envelope* of a proper and lsc function $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ is $\mathrm{env}_\lambda \varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ given by

$$\mathrm{env}_\lambda \varphi(x) := \inf_{x' \in \mathbb{R}^n} \{\varphi(x') + \tfrac{1}{2\lambda}\|x' - x\|^2\}.$$

The set of minimizers is denoted by $\mathrm{prox}_{\lambda\varphi}(x)$. Furthermore, φ is *prox-bounded* if there exists $\lambda > 0$ such that $\mathrm{env}_\lambda \varphi(x) > -\infty$ for some $x \in \mathbb{R}^n$. The supremum of the set of all such λ is the threshold λ_φ.

It is important to note that when φ is lower bounded, such as in the case of an indicator function, we have $\lambda_\varphi = +\infty$. Building on Definition 3.1, we define the Lasry-Lions envelope as follows:

Definition 3.2 *(Lasry–Lions double envelope)* Given $\lambda > \mu > 0$, the *Lasry–Lions double envelope* of a proper and lsc function $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ is $\mathrm{env}_{\lambda,\mu} \varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ given by

$$\mathrm{env}_{\lambda,\mu} \varphi(x) := -\mathrm{env}_\mu(-\mathrm{env}_\lambda \varphi)(x) = \sup_{x' \in \mathbb{R}^n} \{\mathrm{env}_\lambda \varphi(x') - \tfrac{1}{2\mu}\|x' - x\|^2\}. \quad (3.1)$$

We denote the set of maximizers of (3.1) by $P_{\lambda,\mu}\varphi(x) := \mathrm{prox}_{-\mu\,\mathrm{env}_\lambda \varphi}(x)$. Furthermore, the *proximal hull* of φ is given by setting $\mu = \lambda$ in (3.1), and is denoted as $\mathrm{env}_{\lambda,\lambda} \varphi(x)$.

We summarize below some fundamental properties of the Lasry–Lions double envelope that are particularly valuable for the design and theoretical analysis of algorithms.

Proposition 3.3 (Basic properties of the Lasry–Lions double envelope) *Let $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper, lsc, and prox-bounded with threshold λ_φ. The following hold:*

(i) $\mathrm{env}_{\lambda,\mu} \varphi(x)$ *with* $0 < \mu < \lambda < \lambda_\varphi$ *is a smooth function whose gradient is* $\max\{\frac{1}{\mu}, \frac{1}{\lambda-\mu}\}$*-Lipschitz continuous.*

(ii) $\mathrm{env}_\lambda \varphi(x) \le \mathrm{env}_{\lambda,\mu} \varphi(x) \le \mathrm{env}_{\lambda-\mu} \varphi(x) \le \varphi(x)$ *for all* $0 < \mu < \lambda < \lambda_\varphi$.

(iii) $\inf \mathrm{env}_{\lambda,\mu} \varphi = \inf \mathrm{env}_\lambda \varphi = \inf \varphi$ *and* $\arg\min \mathrm{env}_{\lambda,\mu} \varphi = \arg\min \mathrm{env}_\lambda \varphi = \arg\min \varphi$ *for all* $0 < \mu < \lambda < \lambda_\varphi$.

(iv) Suppose that there exist $0 < \bar{\mu} < \bar{\lambda} < \lambda_\varphi$ *such that* $\mathrm{env}_{\bar{\lambda},\bar{\mu}} \varphi$ *is level bounded. Then, any sequence* $x \in \varepsilon\text{-}\arg\min \mathrm{env}_{\lambda,\mu} \varphi$ *with* $\mu < \lambda$, $\lambda \searrow 0$, $\mu \searrow 0$ *and* $\varepsilon \searrow 0$ *is bounded and has all its cluster points in* $\arg\min \varphi$.

(v) Suppose further that φ is bounded from below, let $\lambda > 0$ be given and $j := \frac{1}{2}\|\cdot\|^2$. Then, $\mathrm{env}_{\lambda,\mu} \varphi$ *epi-converges to* $\mathrm{env}_{\lambda,\lambda} \varphi$ *with $\mu > 0$ and $\mu \nearrow \lambda$, the proximal hull is given by* $\mathrm{env}_{\lambda,\lambda} \varphi = (\varphi + \lambda^{-1} j)^{**} - \lambda^{-1} j$ *with domain* $\mathrm{dom}\, \mathrm{env}_{\lambda,\lambda} \varphi = \mathrm{conv}(\mathrm{dom}\, \varphi)$.

Proof Proposition 3.3(i) follows from [3, Theorem 4.1], [42, Proposition 12.62] and [46, Proposition 2.3]; Propositions 3.3(ii) and 3.3(iii) are derived from [42, Example 1.46]; Proposition 3.3(iv) is based on [46, Theorem 3.1].

Regarding Proposition 3.3(v), since φ is bounded from below, Proposition 3.3(iii) and [42, Example 1.44] yield $\inf \mathrm{env}_{\lambda,\lambda} \varphi = \inf \mathrm{env}_\lambda \varphi = \inf \varphi > -\infty$. By

[42, Exercise 1.24 and Example 1.44], it follows that $\mathrm{env}_{\lambda,\lambda}\,\varphi$ is proper, lsc and prox-bounded with threshold $+\infty$. Furthermore, the Lasry–Lions double envelope satisfies $\mathrm{env}_{\lambda,\mu}\,\varphi = \mathrm{env}_{\lambda-\mu}(\mathrm{env}_{\lambda,\lambda}\,\varphi)$ [42, Example 1.46]. Since $\mathrm{env}_{\lambda,\lambda}\,\varphi$ is proper and lsc, [42, Theorem 1.25] implies pointwise convergence $\mathrm{env}_{\lambda-\mu}(\mathrm{env}_{\lambda,\lambda}\,\varphi)(x) \nearrow \mathrm{env}_{\lambda,\lambda}\,\varphi(x)$ for all x as $\lambda - \mu \searrow 0$. Finally, with the aid of [42, Proposition 7.4(d)], we conclude that $\mathrm{env}_{\lambda,\mu}\,\varphi \xrightarrow{\mathrm{e}} \mathrm{env}_{\lambda,\lambda}\,\varphi$ as $\mu \nearrow \lambda$. The domain $\mathrm{dom}\,\mathrm{env}_{\lambda,\lambda}\,\varphi = \mathrm{conv}(\mathrm{dom}\,\varphi)$ and formulation $\mathrm{env}_{\lambda,\lambda}\,\varphi = (\varphi + \lambda^{-1}j)^{**} - \lambda^{-1}j$ follow from [42, Example 11.26(c)]. □

Proposition 3.3(v) establishes the epi-convergence of the Lasry–Lions double envelope to the proximal hull of the original function. Furthermore, the expression $(\varphi + \lambda^{-1}j)^{**} - \lambda^{-1}j$ shows that the proximal hull $\mathrm{env}_{\lambda,\lambda}\,\varphi$ approaches the closed convex hull φ^{**} as $\lambda \to +\infty$. Therefore, this result provides an important insight for algorithm design: one can start with large values of λ and μ, such that $\lambda - \mu$ is small, yielding a nearly convex and smooth surrogate function, and then gradually decrease λ and μ over iterations to improve approximation accuracy. On the other hand, the following lemma establishes the epi-convergence of the Lasry–Lions double envelope to the original function, which is crucial for guaranteeing the convergence of the minima and minimizers of the surrogate problems to those of the original problem.

Lemma 3.4 (Epi-convergence to the original function) *Let $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper, lsc, and prox-bounded with threshold λ_φ. If λ and μ satisfy $\lambda \searrow 0$ and $\mu \searrow 0$ with $0 < \mu < \lambda < \lambda_\varphi$, then $\mathrm{env}_{\lambda,\mu}\,\varphi$ epi-converges to φ.*

Proof According to [42, Theorem 1.25], we have pointwise convergence $\mathrm{env}_{\lambda}\,\varphi(x) \nearrow \varphi(x)$ for all x as $\lambda \searrow 0$. Then, based on [42, Proposition 7.4(d)], we have $\mathrm{env}_{\lambda}\,\varphi \xrightarrow{\mathrm{e}} \varphi$ as $\lambda \searrow 0$. Similarly for $\mathrm{env}_{\lambda-\mu}\,\varphi \xrightarrow{\mathrm{e}} \varphi$ as $\lambda - \mu \searrow 0$. Finally, according to the sandwich property Proposition 3.3(ii), we conclude $\mathrm{env}_{\lambda,\mu}\,\varphi \xrightarrow{\mathrm{e}} \varphi$ from [42, Proposition 7.4(g)]. □

We now examine the envelopes of the indicator function associated with CCs, i.e., $\varphi = \delta_D$. The Moreau envelope of δ_D admits the following compact representation:

$$\mathrm{env}_{\lambda}\,\delta_D(z) = \tfrac{1}{2\lambda}\,\mathrm{dist}\,(z, D)^2 = \tfrac{1}{2\lambda}\left(\|z\|^2 - \max\{[z_1]_+, [z_2]_+\}^2\right). \tag{3.2}$$

We next present the explicit formulation of the Lasry–Lions double envelope of δ_D.

Theorem 3.5 (Lasry–Lions double envelope of δ_D) *Relative to the complementarity set D in (1.2), let $z \in \mathbb{R}^2$, $\lambda > \mu > 0$ and $\beta := \mu/\lambda \in (0, 1)$, the Lasry–Lions double envelope of the indicator function δ_D is given by*

$$\mathrm{env}_{\lambda,\mu}\,\delta_D(z) = \tfrac{1}{\lambda} r_\beta(z), \tag{3.3}$$

with

$$r_\beta(z) = \begin{cases} \frac{1}{2(1-\beta)}\|z\|^2 & \text{if } z \in \mathbb{R}^2_- \quad [\mathrm{O}_-] \\ \frac{1}{2\beta(2-\beta)}(z_1+z_2)^2 - \frac{1}{2\beta}\|z\|^2 & \text{if } (1-\beta)z_1 \le z_2 \le \frac{1}{1-\beta}z_1 \quad [\mathrm{T}_\beta] \\ \frac{1}{2(1-\beta)}\min\{z_1, z_2\}^2 & \text{otherwise}, \quad [\mathrm{H}^{\pm}_{\beta}] \end{cases} \tag{3.4}$$

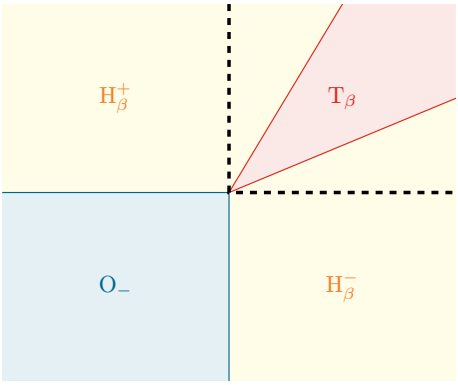


Fig. 2 Polygonal areas of the piecewise definition of $\mathrm{env}_{\lambda,\mu}\,\delta_D$. The thick dashed line corresponds to the complementarity constraint set D, while the red lines dividing the yellow $\mathrm{H}_\beta^\pm$ regions from the red T_β region are the graphs of $z_2 = \frac{1}{1-\beta} z_1$ and $z_2 = (1-\beta) z_1$ in the positive orthant, where $\beta = \mu/\lambda \in (0, 1)$. Apparently, as $\beta \searrow 0$ the T_β region collapses into a half-line, while as $\beta \nearrow 1$ it expands to the entire positive orthant

where the notations $\mathrm{O}_-, \mathrm{H}_\beta^+, \mathrm{H}_\beta^-, \mathrm{T}_\beta$ *refer to the regions as labeled in Fig. 2.*

Proof See Appendix A. □

In light of Theorem 3.5, the gradient of $\mathrm{env}_{\lambda,\mu}\,\delta_D(z)$ is characterized by

$$\nabla\,\mathrm{env}_{\lambda,\mu}\,\delta_D(z) = \tfrac{1}{\lambda} R_\beta(z),$$

where

$$R_\beta(z) = \begin{cases} \frac{1}{1-\beta} z & \text{if } z \in \mathbb{R}^2_- \quad [\mathrm{O}_-] \\ \frac{1}{\beta(2-\beta)} \binom{z_1+z_2}{z_1+z_2} - \frac{1}{\beta} z & \text{if } (1-\beta) z_1 \le z_2 \le \frac{1}{1-\beta} z_1 \quad [\mathrm{T}_\beta] \\ \frac{1}{1-\beta} \binom{z_1}{0} & \text{if } z_2 \ge \frac{1}{1-\beta} [z_1]_+ \quad [\mathrm{H}_\beta^+] \\ \frac{1}{1-\beta} \binom{0}{z_2} & \text{if } z_1 \ge \frac{1}{1-\beta} [z_2]_+. \quad [\mathrm{H}_\beta^-] \end{cases} \tag{3.5}$$

Proposition 3.3(iii) implies that the global minimum of r_β is $r_\beta^\star = 0$ and the minimizer set is D. Furthermore, from the representation (3.2) and Proposition 3.3(ii), it holds that

$$r_\beta(z) \ge \lambda\,\mathrm{env}_\lambda\,\delta_D(z) = \tfrac{1}{2}\,\mathrm{dist}\,(z, D)^2 \ge 0. \tag{3.6}$$

The distinction between the Moreau and Lasry–Lions double envelopes is illustrated in Fig. 3. The next proposition highlights two important properties of r_β: the global Lipschitz continuity of the gradient and the *Polyak-Łojasiewicz (PL)* property.

Proposition 3.6 *Fix* $\beta \in (0, 1)$*, and let functions* r_β *and* R_β *be as in* (3.4) *and* (3.5)*. The following hold:*

(i) R_β *is globally Lipschitz continuous with modulus* $L_\beta = \max\{\frac{1}{\beta}, \frac{1}{1-\beta}\}$.

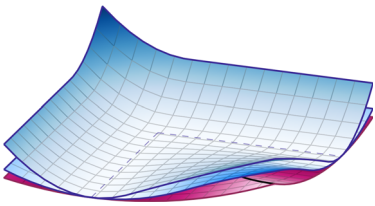


Fig. 3 Comparison between the Moreau and Lasry–Lions double envelopes. The red surface is the Moreau envelope $\mathrm{env}_\lambda\, \delta_D$ when $\lambda = 1$. It is nonsmooth on $\{z \in \mathbb{R}^2 \mid z_1 = z_2 > 0\}$. The middle surface is the Lasry–Lions envelope $\mathrm{env}_{\lambda,\beta\lambda}\, \delta_D$ when $\lambda = 1$ and $\beta = 0.5$, which is smooth everywhere and $\mathrm{env}_{\lambda,\beta\lambda}\, \delta_D(z) \geq \mathrm{env}_\lambda\, \delta_D(z)$. Since $\mathrm{env}_{\lambda,\beta\lambda}\, \delta_D$ provides an extra design freedom β compared to $\mathrm{env}_\lambda\, \delta_D$, by increasing β one obtains a more accurate approximation of $\delta_D(z)$, as in the top surface corresponding to $\lambda = 1$ and $\beta = 0.8$. Moreover, one has $\mathrm{env}_{\lambda,\beta\lambda}\, \delta_D(z) = \mathrm{env}_\lambda\, \delta_D(z) = \delta_D(z) = 0$ for $z \in D$ (dashed line)

(ii) r_β satisfies the PL inequality

$$\tfrac{1}{2}\left\|R_\beta(z)\right\|^2 \geq l_\beta(r_\beta(z) - r_\beta^\star) \quad \forall z \in \mathbb{R}^2$$

with $l_\beta = \min\{\frac{1}{1-\beta}, \frac{1-\beta}{\beta(2-\beta)}\}$ *and* $r_\beta^\star = 0$.

Proof By Proposition 3.3(i), the Lipschitz modulus of $\nabla\, \mathrm{env}_{\lambda,\mu}\, \delta_D$ is $L_{\nabla\, \mathrm{env}_{\lambda,\mu}\, \delta_D} = \max\{\frac{1}{\mu}, \frac{1}{\lambda-\mu}\}$. Moreover, we have $R_\beta(z) = \lambda \nabla\, \mathrm{env}_{\lambda,\mu}\, \delta_D(z)$. Therefore, the Lipschitz modulus of R_β is given by

$$L_\beta = \lambda L_{\nabla\, \mathrm{env}_{\lambda,\mu}\, \delta_D} = \max\{\tfrac{1}{\beta}, \tfrac{1}{1-\beta}\}.$$

On the other hand, it is straightforward to verify that the PL inequality holds with $l_\beta = \frac{1}{1-\beta}$ when $z \in \mathrm{O}_- \cup \mathrm{H}_\beta^{\pm}$, recalling that these regions are defined in Fig. 2. Let us consider the remaining case where $z \in \mathrm{T}_\beta$. Then, we obtain

$$\begin{aligned}
&\tfrac{1}{2}\left\|R_\beta(z)\right\|^2 - l_\beta r_\beta(z) \\
&= \tfrac{1}{2}\left(\tfrac{(1-\beta)^2+1}{\beta^2(2-\beta)^2} + \tfrac{1-\beta}{\beta(2-\beta)} l_\beta\right)\|z\|^2 - \left(\tfrac{1-\beta}{\beta^2(2-\beta)^2} + \tfrac{1}{\beta(2-\beta)} l_\beta\right) z_1 z_2.
\end{aligned}$$

By letting $l_\beta = \frac{1-\beta}{\beta(2-\beta)}$, we further have

$$\tfrac{1}{2}\left\|R_\beta(z)\right\|^2 - l_\beta r_\beta(z) = \tfrac{(1-\beta)^2}{\beta^2(2-\beta)^2}\left(z_1 - z_2\right)^2 + \tfrac{1}{2\beta^2(2-\beta)^2}\|z\|^2 \geq 0,$$

yielding the claimed inequality. □

The Lasry–Lions double envelopes in Definition 3.2 were originally introduced by [32] as approximations that can exhibit better behavior than the Moreau envelopes; for further details, see also [3, 46]. Indeed, the properties established in this section make r_β particularly well suited for smoothing the highly combinatorial structure of CCs; see Fig. 4 below for an illustrative example. Therefore, it plays a central role in the design and analysis of the method proposed in Sections 4 and 5.

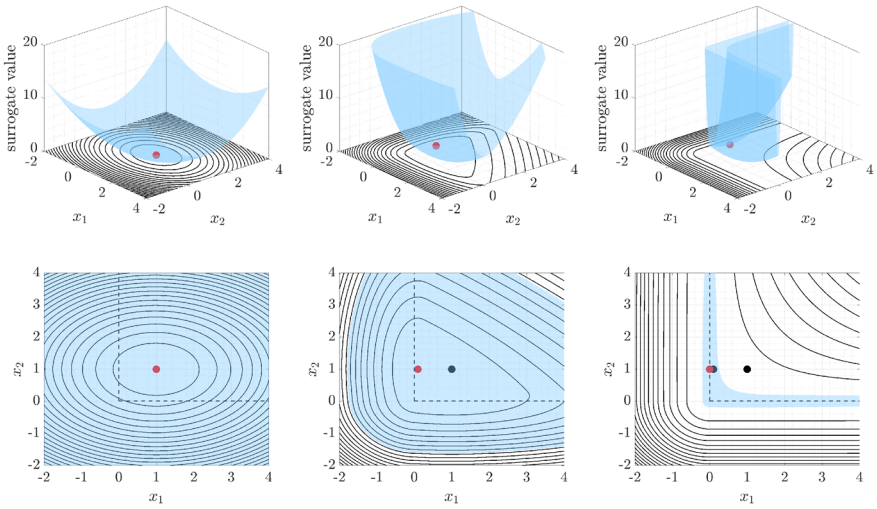


Fig. 4 Solving the *kth3* (4.3) via the LL-based homotopy approach. The red dot indicates the current subproblem's solution, while the black dot indicates the solution from the previous subproblem. The black curves are contours of $s_{\lambda,\beta}$. The dashed line represents the feasible set. The bottom row of subfigures provides a top-down view of the subfigures on top

4 A Homotopy Approach for MPCC

Rather than addressing the original problem (1.1) directly, we adopt an alternative strategy by solving a sequence of smoothed subproblems. They are formally defined as follows, given two parameters $\lambda > \mu > 0$:

$$\underset{x\in C}{\text{minimize}} \quad f(x) + \sum_{i=1}^{p} \text{env}_{\lambda,\mu}\, \delta_D(F_i(x)). \tag{4.1}$$

The stationary points x of (4.1) are characterized by

$$\nabla f(x) + \frac{1}{\lambda} \sum_{i=1}^{p} JF_i(x)^\top R_\beta(F_i(x)) + N_C(x) \ni 0, \tag{4.2}$$

where we remind that $\beta := \mu/\lambda \in (0, 1)$ and $R_\beta : \mathbb{R}^2 \to \mathbb{R}^2$ is the Lipschitz mapping defined in (3.5). An approximate stationary point at iteration ν, computed with parameters λ^ν and β^ν, is denoted by x^ν. This point is then used to initialize the next smoothed subproblem with updated parameters $\lambda^{\nu+1}$ and $\beta^{\nu+1}$.

Example 4.1 (kth3) The basic idea of our Lasry–Lions (LL)-based homotopy approach is illustrated by using the simple example *kth3* from the MacMPEC collection [33].

Consider the MPCC:

$$\begin{aligned} &\underset{x\in\mathbb{R}^2}{\text{minimize}} \;\; 0.5(x_1-1)^2+(x_2-1)^2 \\ &\text{subject to} \;\; 0\le x_1 \perp x_2 \ge 0, \end{aligned} \tag{4.3}$$

whose global minimizer is $(0, 1)$. The associated subproblem is

$$\underset{x\in\mathbb{R}^2}{\text{minimize}}\; s_{\lambda,\beta}(x) = 0.5(x_1-1)^2+(x_2-1)^2+\tfrac{1}{\lambda}r_\beta(x).$$

Let $\beta = 0.9$ be fixed and start with $\lambda^1 = 100$. As shown in the left column of Fig. 4, and in light of Proposition 3.3(v), the resulting smoothed subproblem is nearly convex and well-conditioned. This allows it to be efficiently solved by first-order methods, yielding the solution $x^1 = (1, 1)$. Next, set $\lambda^2 = 1$. The corresponding subproblem, depicted in the middle column of Fig. 4, yields the solution $x^2 = (0.1, 1)$. Finally, with $\lambda^3 = 0.01$, the subproblem grows ill-conditioned, as shown in the right column of Fig. 4. While this improves the approximation of the feasible set, it would typically hinder convergence due to numerical instability. However, thanks to warm-starting from the previous solution x^2, the algorithm avoids the local minimizer $(1, 0)$ and successfully recovers the global minimizer $x^3 = (0, 1)$.

Building on the above notation, we introduce

$$s(x) := f(x)+\sum_{i=1}^{p}\delta_D(F_i(x)) \quad\text{and}\quad s_{\lambda,\beta}(x) := f(x)+\tfrac{1}{\lambda}\sum_{i=1}^{p} r_\beta(F_i(x)). \tag{4.4}$$

Then, according to Lemma 3.4 and [7, Lemma 5.1 and Proposition 5.2], given two sequences $\{\lambda^\nu\}\subset\mathbb{R}_{++}$ and $\{\beta^\nu\}\subset(0,1)$ with $\lambda^\nu\searrow 0$ and $\limsup_{\nu\in\mathbb{N}}\beta^\nu<1$, we have

$$s_{\lambda^\nu,\beta^\nu}\xrightarrow{\text{e}} s. \tag{4.5}$$

As shown in (4.2), $y_i^\nu=\frac{1}{\lambda^\nu}R_{\beta^\nu}(H_i(x^\nu))$ serves as the multiplier, thus the goal is to show that the sequence $\{y_i^\nu\}$ is bounded and its limit points belong to either the cone $N_D^{\mathrm{C}}(F_i(x))$ or $N_D(F_i(x))$ with $\{x^\nu\}\to x$. To this end, we first discuss the properties of $\{y_i^\nu\}$.

Lemma 4.2 *Given two sequences $\{\lambda^\nu\}\subset\mathbb{R}_{++}$ and $\{\beta^\nu\}\subset(0,1)$ with $\lambda^\nu\searrow 0$ and $\limsup_{\nu\in\mathbb{N}}\beta^\nu<1$, let $\{y^\nu\}$ be a sequence such that $y^\nu=\frac{1}{\lambda^\nu}R_{\beta^\nu}(z^\nu)$ for all $\nu\in\mathbb{N}$ for some $\{z^\nu\}\to z\in D$, and let y be any cluster point of the sequence $\{y^\nu\}$ or (provided that $y^\nu\neq 0$ for all ν) of $\{\frac{y^\nu}{\|y^\nu\|}\}$. Then, it holds that $y\in N_D^{\mathrm{C}}(z)$. Furthermore, a sufficient condition for the inclusion $y\in N_D(z)$ to hold (as opposed to $y\in N_D^{\mathrm{C}}(z)$) is the existence of $\xi\in\mathbb{R}_{++}$ and $k\in\mathbb{N}$ such that $z^\nu\notin \mathrm{B}(0,\xi)\cap\mathbb{R}^2_{++}$ holds for all $\nu\ge k$.*

Proof Let O_-, $\mathrm{H}_\beta^\pm$ and T_β denote the regions of $\mathbb{R}^2$ as in Fig. 2, let y be a limit point of a subsequence $\{\frac{y^\nu}{\|y^\nu\|}\}_{\nu\in K}$ or $\{y^\nu\}_{\nu\in K}$. We consider the following cases:

- Suppose that $z \in D \setminus \{0\}$. Without loss of generality, we consider $z = (q, 0)$ for some $q > 0$. Having assumed $\sup_{\nu\in\mathbb{N}} \beta^\nu < 1$, any sequence $z^\nu \to z$ eventually belongs to $\mathrm{H}^-_{\beta^\nu}$. Therefore, for ν sufficiently large, one has

$$y^\nu = \tfrac{1}{\lambda^\nu} R_{\beta^\nu}(z^\nu) = \frac{1}{\lambda^\nu(1-\beta^\nu)} \begin{pmatrix} 0 \\ z_2^\nu \end{pmatrix}.$$

  It follows that y belongs to $\{0\} \times \mathbb{R} = \hat{N}_D((q, 0)) \subseteq N_D^{\mathrm{C}}(z)$.
- Suppose that $z = 0$ and that $z^\nu \in \mathrm{H}^-_{\beta^\nu}$ holds for all $\nu \in K$. Then, for all $\nu \in K$ one again obtains the above relation. Therefore, y belongs to $\{0\} \times \mathbb{R} = N_D((0, 0)) \subseteq N_D^{\mathrm{C}}(z)$.
- Suppose that $z = 0$ and that $z^\nu \in \mathrm{O}_-$ holds for all $\nu \in K$. Then, for all $\nu \in K$ one has

$$y^\nu = \tfrac{1}{\lambda^\nu} R_{\beta^\nu}(z^k) = \frac{1}{\lambda^\nu(1-\beta^\nu)} \begin{pmatrix} z_1^\nu \\ z_2^\nu \end{pmatrix} \leq \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

  which implies that y belongs to $\mathbb{R}^2_- = \hat{N}_D((0, 0)) \subseteq N_D^{\mathrm{C}}(z)$.
- Suppose that $z = 0$ and that $z^\nu \in \mathrm{T}_{\beta^\nu}$ holds for all $\nu \in K$. Then, for all $\nu \in K$ one has

$$y^\nu = \tfrac{1}{\lambda^\nu} R_{\beta^\nu}(z^\nu) = \frac{1}{\lambda^\nu\beta^\nu(2-\beta^\nu)} \begin{pmatrix} z_2^\nu - (1-\beta^\nu) z_1^\nu \\ z_1^\nu - (1-\beta^\nu) z_2^\nu \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

  which implies that y belongs to $\mathbb{R}^2_+ \subseteq N_D^{\mathrm{C}}(z)$.

In this way, we conclude that $y \in N_D^{\mathrm{C}}(z)$, and that $y \in N_D(z)$ as long as $z^\nu \notin \operatorname{int} \mathrm{T}_{\beta^\nu}$ eventually always holds. This latter condition is guaranteed by the sufficient condition provided in the statement. □

The notation of well-behaved CCs in Definition 2.5 allows us to exclude the problematic region T_β from the convergence analysis. In particular, by using the sufficient conditions in Lemma 4.2, we establish the following result.

Corollary 4.3 *Suppose that F_i is a well-behaved CC for some $i \in \{1, \ldots, p\}$, and consider a sequence $\{x^\nu\} \subset \mathbb{R}^n$ converging to a point x such that $F_i(x) \in D$. Let $\{\lambda^\nu\} \subset \mathbb{R}_{++}$ and $\{\beta^\nu\} \subset (0, 1)$ be sequences such that $\lambda^\nu \searrow 0$ and $\limsup_{\nu\in\mathbb{N}} \beta^\nu < 1$, and let $y_i^\nu = \frac{1}{\lambda^\nu} R_{\beta^\nu}(F_i(x^\nu))$. Then, any cluster point of $\{\frac{y_i^\nu}{\|\boldsymbol{y}^\nu\|}\}$ or $\{y_i^\nu\}$ belongs to the cone $N_D(F_i(x))$, for such well-behaved F_i.*

Lemma 4.2 and Corollary 4.3 characterize the behavior of $\{\boldsymbol{y}^\nu\}$. We now use these results to analyze the boundedness of the multipliers in the sequence of gradients $\{\nabla(\sum_{i=1}^p \operatorname{env}_{\lambda^\nu,\beta^\nu\lambda^\nu} \delta_D(F_i(x^\nu)))\}$ based on BCCQ, as stated in the following lemma.

Lemma 4.4 *(Bounded multiplier) Let $\{\lambda^\nu\} \subset \mathbb{R}_{++}$ and $\{\beta^\nu\} \subset (0, 1)$ be sequences such that $\lambda^\nu \searrow 0$ and $\limsup_{\nu\in\mathbb{N}} \beta^\nu < 1$. Consider a sequence $\{x^\nu\} \subset \mathbb{R}^n$ such that $\operatorname{dist}(-\nabla s_{\lambda^\nu,\beta^\nu}(x^\nu), N_C(x^\nu)) \searrow 0$, and suppose that $\{x^\nu\} \to x$, where x satisfies $F_i(x) \in D$ for all $i \in \{1, \ldots, p\}$, and BCCQ holds at x. Let $\boldsymbol{y}^\nu = (y_1^\nu, \ldots, y_p^\nu)$ with $y_i^\nu = \frac{1}{\lambda^\nu} R_{\beta^\nu}(F_i(x^\nu))$. Then, the sequence $\{\boldsymbol{y}^\nu\}$ is bounded as $\{x^\nu\} \to x$.*

Proof To arrive at a contradiction, suppose that $\{\boldsymbol{y}^\nu\}$ is unbounded. Without loss of generality we can assume that $\{\frac{\boldsymbol{y}^\nu}{\|\boldsymbol{y}^\nu\|}\} \to \boldsymbol{y} = (y_1, \ldots, y_p) \neq 0$. It follows from the assumptions that there exists $r^\nu \to 0$ such that

$$r^\nu \in \nabla f(x^\nu) + \tfrac{1}{\lambda^\nu} \sum_{i=1}^{p} J F_i(x^\nu)^\top R_{\beta^\nu}(F_i(x^\nu)) + N_C(x^\nu).$$

Dividing the inclusion by $\|\boldsymbol{y}^\nu\|$ and using the fact that $N_C(x^\nu)$ is a cone, we have

$$\frac{r^\nu}{\|\boldsymbol{y}^\nu\|} \in \frac{\nabla f(x^\nu)}{\|\boldsymbol{y}^\nu\|} + \sum_{i=1}^{p} J F_i(x^\nu)^\top \frac{y_i^\nu}{\|\boldsymbol{y}^\nu\|} + N_C(x^\nu),$$

which, upon passing to the limit, yields that

$$0 \in \sum_{i=1}^{p} J F_i(x)^\top y_i + N_C(x).$$

We next consider the well-behaved and the remaining CCs separately.

- If $i \notin \mathcal{W}$, it follows from Lemma 4.2 that $y_i \in N_D^{\mathrm{C}}(F_i(x))$.
- If $i \in \mathcal{W}$, then Corollary 4.3 implies that $y_i \in N_D(F_i(x))$. Moreover, if $H_i \geq 0$ is a constant mapping for the well-behaved CC $F_i = (G_i, H_i)$, then the expression (3.5) of R_β implies that $(R_{\beta^\nu}(F_i(x^\nu)))_2 = 0$ for all ν large enough, implying that $y_i^H = 0$ in this case. The same reasoning shows that $y_i^G = 0$ if $G_i \geq 0$ is a constant mapping.

Altogether, this shows that $\boldsymbol{y}$ satisfies the conditions in BCCQ, and must thus be zero by the BCCQ assumption, yielding the contradiction. □

Remark 4.5 Since BCCQ is weaker than MPCC-MFCQ (see (2.3)), the statement of Lemma 4.4 also holds under MPCC-MFCQ, which is more commonly used in the MPCC literature. On the other hand, it can be seen from [20, 28] that MPCC-MFCQ is not always required; instead, MPCC-CRCQ [28, Definition 3(c)] or MPCC-CPLD [28, Definition 3(d)] may be employed. Establishing the implications between MPCC-CRCQ, MPCC-CPLD and BCCQ remains an interesting topic for future research.

Remark 4.6 For simplicity of exposition, unless explicitly mentioned otherwise we set $\mathcal{W} = \emptyset$ in the remaining proofs; all subsequent results have straightforward counterparts when $\mathcal{W} \neq \emptyset$ is considered, up to replacing the Clarke limiting cone with the limiting normal cone for the corresponding indices $i \in \mathcal{W}$.

We are now in a position to state whether the approximating problems (4.1), together with their optimality conditions (4.2), are well-justified surrogates of the original problem (1.1) and its optimality condition in Definition 2.1. To this end, we leverage the concept of *consistent approximation*.

Definition 4.7 *(Consistent approximation [43, Definition 2.2])* Given functions $s, s^\nu : \mathbb{R}^n \to \overline{\mathbb{R}}$ and set-valued mappings $S, S^\nu : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, $\nu \in \mathbb{N}$, the pairs $\{(s^\nu, S^\nu)\}$ form a *consistent approximation* of (s, S) if

$$s^\nu \xrightarrow{\text{e}} s \quad \text{and} \quad S^\nu \xrightarrow{\text{g}} S.$$

If the graphical convergence is relaxed to

$$\operatorname*{Lim\,sup}_{\nu\to\infty} \operatorname{gph} S^\nu \subseteq \operatorname{gph} S, \tag{4.6}$$

then the pairs $\{(s^\nu, S^\nu)\}$ form a *weakly consistent approximation* of (s, S).

Proposition 3.3(iv) implies that if the subproblem can be progressively solved to global optimality, then the global minimizer of the original problem can be retrieved. However, solving the subproblems to global optimality is typically intractable. As an alternative, the following theorem shows that using first-order conditions, the limit $\operatorname{dist}(-\nabla s_{\lambda^\nu,\beta^\nu}(x^\nu), N_C(x^\nu)) \searrow 0$ with $\lambda^\nu \searrow 0$ and $\limsup \beta^\nu < 1$, characterizes a C-stationary point of (1.1).

Theorem 4.8 (C-stationary point) *Let $\{\lambda^\nu\} \subset \mathbb{R}_{++}$ and $\{\beta^\nu\} \subset (0, 1)$ be sequences such that $\lambda^\nu \searrow 0$ and $\limsup_{\nu\in\mathbb{N}} \beta^\nu < 1$. Consider a sequence $\{x^\nu\} \subset \mathbb{R}^n$ such that $\operatorname{dist}(-\nabla s_{\lambda^\nu,\beta^\nu}(x^\nu), N_C(x^\nu)) \searrow 0$, and suppose that $\{x^\nu\} \to x$, where x satisfies $F_i(x) \in D$ for all $i \in \{1, \ldots, p\}$, and BCCQ holds at x. Let $\boldsymbol{y}^\nu = (y_1^\nu, \ldots, y_p^\nu)$ with $y_i^\nu = \frac{1}{\lambda^\nu} R_{\beta^\nu}(F_i(x^\nu))$. Then, we have $\{\boldsymbol{y}^\nu\} \to \boldsymbol{y}$, and every limit pair $(x, \boldsymbol{y})$ is a C-stationary primal-dual pair of (1.1), i.e.,*

$$0 \in \nabla f(x) + \sum_{i=1}^{p} JF_i(x)^\top y_i + N_C(x) \quad \text{with} \quad y_i \in N_D^{\text{C}}(F_i(x)) \quad \forall i \in \{1, \ldots, p\}.$$

Proof Lemma 4.4 implies that $\{\boldsymbol{y}^\nu\}$ is bounded under BCCQ, thus we have $\{\boldsymbol{y}^\nu\} \to \boldsymbol{y}$ as $\{x^\nu\} \to x$. On the other hand, to demonstrate that the approximating problems, $\operatorname{minimize}_{x\in C} s_{\lambda^\nu,\beta^\nu}(x)$, eventually become accurate relative to (1.1) in terms of stationary points, under the light of Definition 4.7 we need to prove that the surrogate function epi-converges to the original function, which is already established in (4.5). In addition, it is necessary to verify that the generalized equation corresponding to (4.2) at least satisfies the inclusion (4.6). To this end, let the limit

$$\begin{aligned} v \in & \operatorname*{Lim\,sup}_{x^\nu\to x, \lambda^\nu\searrow 0} \nabla s_{\lambda^\nu,\beta^\nu}(x^\nu) + \partial\delta_C(x^\nu) \\ = & \operatorname*{Lim\,sup}_{x^\nu\to x, \lambda^\nu\searrow 0} \nabla f(x^\nu) + \tfrac{1}{\lambda^\nu} \sum_{i=1}^{p} JF_i(x^\nu)^\top R_{\beta^\nu}(F_i(x^\nu)) + N_C(x^\nu) \end{aligned}$$

be given under a sequence $\{\beta^\nu\} \subset (0,1)$ such that $\limsup_{\nu\in\mathbb{N}} \beta^\nu < 1$. Then, the sequences $\{\lambda^\nu\}$, $\{\beta^\nu\}$, $\{x^\nu\}$ and $\{\boldsymbol{y}^\nu\}$ satisfy

$$S_{\lambda^\nu,\beta^\nu}(x^\nu) := \nabla f(x^\nu) + \sum_{i=1}^{p} JF_i(x^\nu)^\top y_i^\nu + N_C(x^\nu) \to v.$$

Lemma 4.2 establishes that every limit point of $\{y_i^\nu\}$ belongs to $N_D^{\mathrm{C}}(F_i(x))$, i.e., $y_i \in N_D^{\mathrm{C}}(F_i(x))$ for all $i \in \{1,\dots,p\}$. Thus, we have

$$S(x) := \nabla f(x) + \sum_{i=1}^{p} JF_i(x)^\top N_D^{\mathrm{C}}(F_i(x)) + N_C(x) \ni v,$$

which implies that $\operatorname{Lim\,sup}_{\nu\to\infty} \operatorname{gph} S_{\lambda^\nu,\beta^\nu} \subseteq \operatorname{gph} S$. Therefore, one has that the pairs $\{(s_{\lambda^\nu,\beta^\nu}, S_{\lambda^\nu,\beta^\nu})\}$ are a weakly consistent approximation of (s, S) based on Definition 4.7. Then, the condition $\operatorname{dist}(-\nabla s_{\lambda^\nu,\beta^\nu}(x^\nu), N_C(x^\nu)) \searrow 0$ results in all limit points of $\{x^\nu\}$ satisfying $S(x) \ni 0$ according to [43, Proposition 2.3]. In this way, we conclude that $(x, \boldsymbol{y})$ is a C-stationary primal-dual pair of (1.1). □

We now turn our attention to the M-stationarity, which hinges on the following important observation.

Lemma 4.9 *Fix $\beta \in (0,1)$ and let R_β be as in (3.5) and $\mathrm{T}_\beta \subset \mathbb{R}^2$ as in Fig. 2. For every $z \in \mathbb{R}^2 \setminus \operatorname{int} \mathrm{T}_\beta$ it holds that $R_\beta(z) \in \hat{N}_D(\bar{z})$, where $\bar{z} = \Pi_D(z)$.*

Proof Note that $\bar{z}$ as in the statement is well defined. Indeed, Π_D is single-valued everywhere except on the diagonal within the strictly positive orthant, which is entirely contained in $\operatorname{int} \mathrm{T}_\beta$ and thus not considered here. With O_- and $\mathrm{H}_\beta^\pm$ as in Fig. 2, the three possible scenarios are:

- $z \in \mathrm{O}_-$. In this case, $\bar{z} = (0,0)$ and $R_\beta(z) \le (0,0)$, as needed.
- $z \in \mathrm{H}_\beta^+ \setminus \mathrm{O}_-$. In this case, $\bar{z} = (0, z_2)$ and $(R_\beta(z))_2 = 0$, as needed.
- $z \in \mathrm{H}_\beta^- \setminus \mathrm{O}_-$. In this case, $\bar{z} = (z_1, 0)$ and $(R_\beta(z))_1 = 0$, as needed.

□

The region $\operatorname{int} \mathrm{T}_\beta$ turns out to be the problematic one, since all components of R_β are strictly positive there. Indeed, $\frac{1}{\beta(2-\beta)}(z_1+z_2) - \frac{1}{\beta} z_1 \le 0$ if and only if $z_2 \le (1-\beta) z_1$, and similarly $\frac{1}{\beta(2-\beta)}(z_1 + z_2) - \frac{1}{\beta} z_2 \le 0$ if and only if $z_2 \ge \frac{1}{1-\beta} z_1$. Therefore, in order to establish the M-stationarity at x, we need to assume that there is no constraint $F_i(x)$ that falls into the area $\operatorname{int} \mathrm{T}_\beta$; in other words, all constraints are well-behaved. In this case, the cone $N_D^{\mathrm{C}}(F_i(x))$ can be replaced by $N_D(F_i(x))$ in Definition 2.1, and BCQ$^+$, which is weaker than BCCQ (see (2.3)), suffices for guaranteeing the boundedness of $\{\boldsymbol{y}^\nu\}$.

Theorem 4.10 (AM- and M-stationary point) *Let $\{\lambda^\nu\} \subset \mathbb{R}_{++}$ and $\{\beta^\nu\} \subset (0,1)$ be sequences such that $\lambda^\nu \searrow 0$ and $\limsup_{\nu\in\mathbb{N}} \beta^\nu < 1$. Consider a sequence $\{x^\nu\} \subset \mathbb{R}^n$*

such that $\operatorname{dist}(-\nabla s_{\lambda^\nu,\beta^\nu}(x^\nu), N_C(x^\nu)) \searrow 0$, *and suppose that* $\{x^\nu\} \to x$, *where* x *satisfies* $F_i(x) \in D$, *and* F_i *is well-behaved for all* $i \in \{1,\dots,p\}$. *Let* $\boldsymbol{y}^\nu = (y_1^\nu,\dots,y_p^\nu)$ *with* $y_i^\nu = \frac{1}{\lambda^\nu} R_{\beta^\nu}(F_i(x^\nu))$. *Then, the point* x *is AM-stationary. Furthermore, any such limit point* x *at which* BCQ$^+$ *is satisfied is actually M-stationary.*

Proof By following the method in Lemma 4.4, we know that the sequence $\{y_i^\nu\}$ with $y_i^\nu = \frac{1}{\lambda^\nu} R_{\beta^\nu}(F_i(x^\nu))$ is bounded if BCQ$^+$ is satisfied at x. On the other hand, we have

$$\operatorname{dist}\left(-\nabla f(x^\nu) - \textstyle\sum_{i=1}^p JF_i(x^\nu)^\top y_i^\nu, N_C(x^\nu)\right) \searrow 0. \tag{4.7}$$

For any $i \in \{1,\dots,p\}$ we consider two cases:

- $F_i(x) \neq (0,0)$. For every ν let

$$z_i^\nu \in \Pi_D(F_i(x^\nu)).$$

  Since $\limsup \beta^\nu < 1$ one has that $z_i^\nu \notin \operatorname{int} T_{\beta^\nu}$ for all ν large enough, and Lemma 4.9 implies that

$$y_i^\nu \in N_D(z_i^\nu) \tag{4.8}$$

  for all such ν. Furthermore, since $F_i(x) \in D$, clearly one has

$$F_i(x^\nu) - z_i^\nu \to 0. \tag{4.9}$$

  All properties (4.7), (4.8), (4.9) and the feasibility of x combined show that all the conditions of Definition 2.2 are satisfied, hence that x is AM-stationary.
- $F_i(x) = (0,0)$. If $F_i(x^\nu) \notin \operatorname{int} T_{\beta^\nu}$ holds for all ν large enough, then the arguments appealing to Lemma 4.9 as in the previous case can be replicated and the same conclusion follows.

In case $\{\boldsymbol{y}^\nu\}$ is bounded, this being ensured by BCQ$^+$ at x as shown above, M-stationarity follows. □

4.1 Special Instance Assuming Feasible Initialization

Without further assumptions, limit points may be infeasible, even in convex composite optimization, such as the Gauss-Newton method for solving nonlinear systems of equations [7]. Therefore, it is reasonable to introduce a further restriction on how to generate $\{x^\nu\}$. Specifically, if it is possible to select a feasible starting point, then one can proceed along the lines of Algorithm 1.

Theorem 4.11 *Consider the iterates generated by Algorithm 1 with* $\epsilon^\nu \searrow 0$, $\lambda^\nu \searrow 0$ *and* $\sup_{\nu\in\mathbb{N}} \beta^\nu < 1$. *If* f *is bounded from below, then the following hold:*

(i) Any cluster point of $\{x^\nu\}$ *is feasible.*
(ii) Any cluster point of $\{x^\nu\}$ *at which BCCQ is satisfied is C-stationary.*
(iii) Any cluster point of $\{x^\nu\}$ *is AM-stationary as long as* F_i *is well-behaved for all* $i \in \{1,\dots,p\}$. *Any such cluster point at which* BCQ$^+$ *is satisfied is actually M-stationary.*

Algorithm 1 LL-based homotopy approach with feasible initialization.

Initialization: Let $\nu = 1$ and $x^0 \in C$ be such that $F_i(x^0) \in D$ for all $i \in \{1, \ldots, p\}$.

Repeat:

1: Choose $\epsilon^\nu, \lambda^\nu > 0$ and $\beta^\nu \in (0, 1)$, and set $s_{\lambda^\nu, \beta^\nu} := f + \sum_{i=1}^p \mathrm{env}_{\lambda^\nu, \beta^\nu \lambda^\nu} \delta_D \circ F_i$.

2: Find an approximate solution x^ν to the problem $\mathrm{minimize}_{x \in C}\, s_{\lambda^\nu, \beta^\nu}(x)$ such that

- $s_{\lambda^\nu, \beta^\nu}(x^\nu) \le \min\{s_{\lambda^\nu, \beta^\nu}(x^{\nu-1}), f(x^0)\}$ and
- $\mathrm{dist}(-\nabla s_{\lambda^\nu, \beta^\nu}(x^\nu), N_C(x^\nu)) \le \epsilon^\nu$.

3: Let $\nu \leftarrow \nu + 1$.

Proof According to the sandwich property Proposition 3.3(ii), we have

$$f(x^0) \ge s_{\lambda^\nu, \beta^\nu}(x^\nu) \ge f(x^\nu) + \tfrac{1}{2\lambda^\nu} \sum_{i=1}^p \mathrm{dist}\,(F_i(x^\nu), D)^2.$$

Suppose that x is a cluster point such that $F_i(x) \notin D$ for some $i \in \{1, \ldots, p\}$. Let $J \in \mathcal{N}_\infty^\sharp$ be such that $\{x^\nu\} \to_J x$, then $\frac{1}{2\lambda^\nu} \mathrm{dist}\,(F_i(x^\nu), D)^2 \to_J +\infty$ for such i, which gives a contradiction $f(x^\nu) \to_J -\infty$ (otherwise, the function f would be unbounded below). Thus, we have $F_i(x) \in D$ for all $i \in \{1, \ldots, p\}$. Claims (ii) and (iii) come from Theorems 4.8 and 4.10. □

5 Practical Algorithm and Complexity Analysis

The previous section analyzes the asymptotic behavior when $\lambda^\nu \searrow 0$ and $\epsilon^\nu \searrow 0$. However, a practical algorithm always requires termination in a finite number of iterations to meet a user-defined accuracy tolerance. In this section, we present a practical variant of Algorithm 1 and analyze its worst-case complexity to reach an approximate stationary point.

5.1 Assumptions and Auxiliary Lemmas

At the beginning of this paper, we assume that the solution set of (1.1) is nonempty and all functions are smooth. In this section, to establish the worst-case complexity, we additionally assume the compactness of the level set of the objective function f over the set C, as stated below.

Assumption 5.1 For $\alpha \in \mathbb{R}$, the level set $\mathrm{lev}_{\le\alpha}[f; C] := \{x \in C \mid f(x) \le \alpha\}$ is compact (possibly empty).

Assumption 5.1 implies that there exists $\underline{f} \in \mathbb{R}$ such that $f(x) \ge \underline{f}$ for all $x \in C$. In addition, we assume that the constraint function F_i is Lipschitz continuous. Together with the previously stated smoothness assumption, these conditions are formalized as follows:

Assumption 5.2 The constraint function $F_i : \mathbb{R}^n \to \mathbb{R}^2$ is K_i-Lipschitz continuous and its Jacobian $JF_i : \mathbb{R}^n \to \mathbb{R}^{2\times n}$ is L_i-Lipschitz continuous for all $i \in \{1, \ldots, p\}$. Furthermore, the objective gradient $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ is L_f-Lipschitz continuous.

As established in the previous section, the original problem (1.1) admits a smoothed approximation of the form: minimize$_{x\in C}\, s_{\lambda,\beta}(x)$. The following lemma establishes the property of $s_{\lambda,\beta}$.

Lemma 5.3 *Let Assumption 5.1 hold and $s_{\lambda,\beta}$ be as in (4.4), choose $\lambda > 0$, $\beta \in (0, 1)$ and $\tilde{x} \in C$, then the level set*

$$\mathrm{lev}_{\leq s_{\lambda,\beta}(\tilde{x})}[s_{\lambda,\beta}; C] = \{x \in C \mid s_{\lambda,\beta}(x) \leq s_{\lambda,\beta}(\tilde{x})\}$$

is compact.

Proof Let $x \in \mathrm{lev}_{\leq s_{\lambda,\beta}(\tilde{x})}[s_{\lambda,\beta}; C]$. Using the fact that r_β is nonnegative, as established in (3.6), we obtain

$$f(x) \leq f(x) + \tfrac{1}{\lambda}\sum_{i=1}^{p} r_\beta(F_i(x)) = s_{\lambda,\beta}(x) \leq s_{\lambda,\beta}(\tilde{x}),$$

implying $x \in \mathrm{lev}_{\leq s_{\lambda,\beta}(\tilde{x})}[f; C]$, and then $\mathrm{lev}_{\leq s_{\lambda,\beta}(\tilde{x})}[s_{\lambda,\beta}; C] \subseteq \mathrm{lev}_{\leq s_{\lambda,\beta}(\tilde{x})}[f; C]$. According to Assumption 5.1 and the continuity of f, F_i and r_β, we know that $\mathrm{lev}_{\leq s_{\lambda,\beta}(\tilde{x})}[s_{\lambda,\beta}; C]$ is compact. □

The following lemma establishes an important stopping criterion for our algorithm, ensuring that the violation of CCs remains bounded.

Lemma 5.4 *Let r_β be as in (3.4) and $F_i = (G_i, H_i)$. Suppose $\beta \in (0, 1)$ and $\epsilon > 0$. If x is such that*

$$\sum_{i=1}^{p} r_\beta(F_i(x)) \leq \frac{\epsilon^2}{2}, \tag{5.1}$$

then the complementarity constraints are satisfied approximately by

$$\left\|\min\{G(x), H(x)\}\right\| \leq \epsilon, \tag{5.2}$$

where the minimum is taken componentwise.

Proof For notational conciseness, we denote $\sum_{\mathrm{H}_\beta^\pm}$ as the sum over the indices $i = 1, \ldots, p$ for which $F_i(x) \in \mathrm{H}_\beta^\pm$, and similarly for the other two areas of Fig. 2. Building on this notation, we can reformulate (5.1) equivalently as

$$\sum_{\mathrm{int\,O}_-} \tfrac{1}{2(1-\beta)}\left(G_i(x)^2 + H_i(x)^2\right) + \sum_{\mathrm{H}_\beta^\pm} \tfrac{1}{2(1-\beta)}\min\{G_i(x), H_i(x)\}^2 + \sum_{\mathrm{int\,T}_\beta} r_\beta(F_i(x)) \leq \frac{\epsilon^2}{2}.$$

When $F_i(x) \in \operatorname{int} \mathrm{O}_-$, the value of $G_i(x)^2 + H_i(x)^2$ can be bounded from below by $\min\{G_i(x), H_i(x)\}^2$. Furthermore, given that $1/(1-\beta) > 1$ for $\beta \in (0, 1)$, it follows that

$$\sum_{\operatorname{int} \mathrm{O}_- \cup \mathrm{H}_\beta^{\pm}} \tfrac{1}{2} \min\{G_i(x), H_i(x)\}^2 + \sum_{\operatorname{int} \mathrm{T}_\beta} r_\beta(F_i(x)) \le \frac{\epsilon^2}{2}.$$

As for the case where $F_i(x) \in \operatorname{int} \mathrm{T}_\beta$, combining Proposition 3.3(ii) and the formulation of Moreau envelope (3.2), we have

$$\sum_{\operatorname{int} \mathrm{T}_\beta} \tfrac{1}{2} \min\{G_i(x), H_i(x)\}^2 = \sum_{\operatorname{int} \mathrm{T}_\beta} \lambda \operatorname{env}_\lambda \delta_D(F_i(x)) \le \sum_{\operatorname{int} \mathrm{T}_\beta} r_\beta(F_i(x)),$$

which leads to

$$\sum_{\operatorname{int} \mathrm{O}_- \cup \mathrm{H}_\beta^{\pm} \cup \operatorname{int} \mathrm{T}_\beta} \tfrac{1}{2} \min\{G_i(x), H_i(x)\}^2 = \sum_{i=1}^{p} \tfrac{1}{2} \min\{G_i(x), H_i(x)\}^2 \le \frac{\epsilon^2}{2}.$$

In this way, the claimed result is obtained. □

The following lemma establishes an upper bound of $1/\lambda$, which is determined by the violation of CCs.

Lemma 5.5 *Let $\lambda > 0$, $\beta \in (0, 1)$ and $\epsilon > 0$, and let r_β, $s_{\lambda,\beta}$ be as in (3.4) and (4.4). Choose $x, \tilde{x} \in C$ such that*

$$s_{\lambda,\beta}(x) \le s_{\lambda,\beta}(\tilde{x}) \quad \text{and} \quad \sum_{i=1}^{p} r_\beta(F_i(\tilde{x})) \le \frac{\epsilon^2}{4}.$$

If the condition (5.1) is not satisfied by x, i.e., $\sum_{i=1}^{p} r_\beta(F_i(x)) > \epsilon^2/2$, then the parameter $1/\lambda$ is upper bounded by

$$\frac{1}{\lambda} < \frac{4(f(\tilde{x}) - f(x))}{\epsilon^2}.$$

Proof From the definition of $s_{\lambda,\beta}$ in (4.4) and the condition $s_{\lambda,\beta}(x) \le s_{\lambda,\beta}(\tilde{x})$, it follows that

$$\begin{aligned} f(x) + \tfrac{1}{\lambda} \sum_{i=1}^{p} r_\beta(F_i(x)) = s_{\lambda,\beta}(x) &\le s_{\lambda,\beta}(\tilde{x}) \\ &= f(\tilde{x}) + \tfrac{1}{\lambda} \sum_{i=1}^{p} r_\beta(F_i(\tilde{x})) \le f(\tilde{x}) + \tfrac{\epsilon^2}{4\lambda}. \end{aligned}$$

Then, the violation of (5.1) implies

$$f(x) + \frac{\epsilon^2}{2\lambda} < f(x) + \tfrac{1}{\lambda} \sum_{i=1}^{p} r_\beta(F_i(x)) \leq f(\tilde{x}) + \tfrac{\epsilon^2}{4\lambda},$$

which leads to the result. □

5.2 Algorithm and Outer Complexity

Let us now consider Algorithm 2, which is a practical variant of Algorithm 1. In this version, the tolerance ϵ is held constant rather than decreasing to 0, and similarly, the parameter β is fixed. The starting point x^0 is near-feasible rather than strictly feasible. Furthermore, a specific stopping criterion and warm-start strategy are introduced. This subsection also analyzes the worst-case complexity of the outer iterations, i.e., the number of subproblems that need to be solved. The details of the inner solver will be discussed in the next subsection.

Algorithm 2 LL-based homotopy approach with near-feasible initialization.

Initialization: Let $\nu = 1$. Given $\epsilon > 0, \lambda^0 > 0, \beta \in (0, 1)$ and $x^0 \in C$ such that $\sum_{i=1}^{p} r_\beta(F_i(x^0)) \leq \epsilon^2/4$.

Repeat:

1: Choose λ^ν and set $s_{\lambda^\nu,\beta} := f + \sum_{i=1}^{p} \frac{1}{\lambda^\nu} r_\beta \circ F_i$.

2: Find an approximate solution x^ν to the problem minimize$_{x \in C}\, s_{\lambda^\nu,\beta}(x)$ such that

- $s_{\lambda^\nu,\beta}(x^\nu) \leq \min\{s_{\lambda^\nu,\beta}(x^{\nu-1}), s_{\lambda^\nu,\beta}(x^0)\}$ and
- $\mathrm{dist}(-\nabla s_{\lambda^\nu,\beta}(x^\nu), N_C(x^\nu)) \leq \epsilon$

by using an inner solver from the starting point $x^{\nu,0} := \arg\min\{s_{\lambda^\nu,\beta}(\hat{x}) \mid \hat{x} \in \{x^{\nu-1}, x^0\}\}$.

3: Let $\nu \leftarrow \nu + 1$.

Until: The stopping criterion $\sum_{i=1}^{p} r_\beta(F_i(x^\nu)) \leq \epsilon^2/2$ is satisfied.

Remark 5.6 Step 2 of Algorithm 2 is realistic for first-order methods. The descent condition (the first requirement) follows directly from standard convergence guarantees for descent methods; see, for instance, [5, Theorem 10.15]. The approximate stationarity condition (the second requirement) is widely employed as a practical stopping criterion. This criterion can be satisfied within at most $\mathcal{O}(\epsilon^{-2})$ iterations by standard first-order methods, such as a projected gradient method [31], or an accelerated gradient method [17]. In practice, quasi-Newton methods are well known to often outperform the aforementioned methods. Therefore, we adopt the off-the-shelf L-BFGS-B solver [50] in our numerical experiments presented in Section 6.

Before presenting the main result of this subsection, we introduce the following operator:

$$\Pi_D^{\mathrm{C},\beta}(z) := \begin{cases} \Pi_D(z) & \text{if } z \in \mathrm{O}_- \cup \mathrm{H}_\beta^{\pm} \\ (0, 0) & \text{if } z \in \operatorname{int} \mathrm{T}_\beta, \end{cases}$$

which projects the point z to the origin when $z \in \operatorname{int} \mathrm{T}_\beta$, rather than to one of the two branches of D. The following theorem provides an upper bound on the number of outer iterations in Algorithm 2, and shows that the returned point is an approximate stationary point of (1.1).

Theorem 5.7 (Outer complexity) *Let Assumption 5.1 hold and consider the iterates generated by Algorithm 2 with λ^ν being updated as $\lambda^\nu = \rho\lambda^{\nu-1}$ with $\rho \in (0,1)$. Then, Algorithm 2 terminates in at most*

$$\bar{\nu} < 1 + \log_{\frac{1}{\rho}} \left(\frac{4\lambda^0 (f(x^0) - \underline{f})}{\epsilon^2} \right) \tag{5.3}$$

iterations. Furthermore, the last iterate $x^{\bar{\nu}}$ satisfies the approximate C-stationary condition

$$\begin{aligned} &\operatorname{dist}\left(-\nabla f(x^{\bar{\nu}}) - \textstyle\sum_{i=1}^p JF_i(x^{\bar{\nu}})^\top y_i^{\bar{\nu}},\ N_C(x^{\bar{\nu}}) \right) \le \epsilon \\ &\left\| \min\{G(x^{\bar{\nu}}), H(x^{\bar{\nu}})\} \right\| \le \epsilon \\ &y_i^{\bar{\nu}} \in N_D^{\mathrm{C}}(z_i^{\bar{\nu}}) \end{aligned}$$

for all $i \in \{1, \ldots, p\}$, where $z_i^{\bar{\nu}} = \Pi_D^{\mathrm{C},\beta}(F_i(x^{\bar{\nu}}))$ and $y_i^{\bar{\nu}} = \frac{1}{\lambda^{\bar{\nu}}} R_\beta(F_i(x^{\bar{\nu}}))$.

Proof If Algorithm 2 is terminated by meeting $\sum_{i=1}^p r_\beta(F_i(x^{\bar{\nu}})) \le \epsilon^2/2$, the following inequality holds at the iteration $\bar{\nu} - 1$:

$$\sum_{i=1}^p r_\beta(F_i(x^{\bar{\nu}-1})) > \frac{\epsilon^2}{2}$$

(otherwise, the algorithm would have terminated at iteration $\bar{\nu} - 1$). Recalling that Assumption 5.1 implies the existence of $\underline{f}$, and applying Lemma 5.5 along with the condition $s_{\lambda^{\bar{\nu}-1},\beta}(x^{\bar{\nu}-1}) \le s_{\lambda^{\bar{\nu}-1},\beta}(x^0)$, we obtain

$$\frac{1}{\lambda^{\bar{\nu}-1}} < \frac{4(f(x^0) - f(x^{\bar{\nu}-1}))}{\epsilon^2} \le \frac{4(f(x^0) - \underline{f})}{\epsilon^2}.$$

On the other hand, the tuning strategy indicates $\lambda^{\bar{\nu}-1} = \rho^{\bar{\nu}-1}\lambda^0$, which leads to the following:

$$\frac{1}{\rho^{\bar{\nu}-1}} < \frac{4\lambda^0(f(x^0) - \underline{f})}{\epsilon^2}.$$

Since $\rho \in (0,1)$, applying $\log_{1/\rho}$ to both sides yields (5.3). Moreover, the above discussion implies that if

$$\bar{\nu} \ge 1 + \log_{\frac{1}{\rho}} \left(\frac{4\lambda^0(f(x^0) - \underline{f})}{\epsilon^2} \right),$$

then the stopping criterion (5.1) must be satisfied, i.e.,

$$\sum_{i=1}^{p} r_\beta(F_i(x^{\bar{\nu}-1})) \leq \frac{\epsilon^2}{2}.$$

Therefore, Algorithm 2 terminates finitely. In turn, this also shows that the complexity bound (5.3) must be true, for otherwise the algorithm would have terminated earlier.

Let us now consider the approximate C-stationarity. The first-order condition follows directly from Algorithm 2, while the approximate satisfaction of the CCs is ensured by Lemma 5.4. Regarding the inclusion $y_i^{\bar{\nu}} \in N_D^{\mathrm{C}}(z_i^{\bar{\nu}})$, Lemma 4.9 implies that $R_\beta(F_i(x^{\bar{\nu}})) \in \hat{N}_D(z_i^{\bar{\nu}})$ with $z_i^{\bar{\nu}} = \Pi_D(F_i(x^{\bar{\nu}}))$ when $F_i(x^{\bar{\nu}}) \notin \operatorname{int} \mathrm{T}_\beta$. In case $F_i(x^{\bar{\nu}}) \in \operatorname{int} \mathrm{T}_\beta$, we have $\Pi_D^{\mathrm{C},\beta}(F_i(x^{\bar{\nu}})) = (0,0)$. Hence, it follows that $R_\beta(F_i(x^{\bar{\nu}})) \in N_D^{\mathrm{C}}(z_i^{\bar{\nu}})$ with $z_i^{\bar{\nu}} = \Pi_D^{\mathrm{C},\beta}(F_i(x^{\bar{\nu}}))$. Since $\lambda^{\bar{\nu}} > 0$ (as Algorithm 2 terminates finitely), and due to the property of the cone, we conclude that $y_i^{\bar{\nu}} = \frac{1}{\lambda^{\bar{\nu}}} R_\beta(F_i(x^{\bar{\nu}})) \in N_D^{\mathrm{C}}(z_i^{\bar{\nu}})$ with $z_i^{\bar{\nu}} = \Pi_D^{\mathrm{C},\beta}(F_i(x^{\bar{\nu}}))$ for all $F_i(x^{\bar{\nu}}) \in \mathbb{R}^2$. □

Remark 5.8 If F_i is well-behaved for all $i \in \{1, \ldots, p\}$, then analogously to Theorem 4.10, we refer to $x^{\bar{\nu}}$ as an approximate M-stationary point. Furthermore, it is worth noting that the definitions of approximate C/M-stationarity do not require any CQs. Under BCCQ/ BCQ$^+$, in the limiting setting as $\epsilon \searrow 0$ and $\lambda \searrow 0$, they turn to be standard C/M-stationarity as in Definition 2.1.

5.3 Total Complexity

To determine the total complexity of Algorithm 2, it is necessary to estimate the computational cost of the inner solver for each subproblem. To this end, we first analyze a smoothness property of $s_{\lambda,\beta}$.

Lemma 5.9 *Let Assumptions 5.1 and 5.2 hold, and let $s_{\lambda,\beta}$ be as in (4.4) with $\lambda > 0$, $\beta \in (0,1)$ and $x^0 \in C$. Then, the gradient $\nabla s_{\lambda,\beta}$ is Lipschitz continuous on* $\operatorname{conv}(\operatorname{lev}_{\leq s_{\lambda,\beta}(x^0)}[s_{\lambda,\beta}; C])$ *with modulus*

$$L_\lambda(x^0) = L_f + \tfrac{1}{\lambda}\tilde{L}_\lambda(x^0), \tag{5.4}$$

where

$$\tilde{L}_\lambda(x^0) = \sum_{i=1}^{p}\left[L_i\left(L_\beta K_i D_\lambda^0 + \left\|R_\beta(F_i(x^0))\right\|\right) + L_\beta K_i\left(L_i D_\lambda^0 + \left\|R_\beta(F_i(x^0))\right\|\right)\right] \tag{5.5}$$

and the constant D_λ^0 is defined in (5.7).

Proof For all $x, \tilde{x} \in \operatorname{conv}(\operatorname{lev}_{\leq s_{\lambda,\beta}(x^0)}[s_{\lambda,\beta}; C])$, we have

$$\left\|\nabla s_{\lambda,\beta}(x) - \nabla s_{\lambda,\beta}(\tilde{x})\right\|$$

$$
\begin{aligned}
&= \left\| \nabla f(x) - \nabla f(\tilde{x}) + \tfrac{1}{\lambda} \textstyle\sum_{i=1}^{p} \left[JF_i(x)^\top R_\beta(F_i(x)) - JF_i(\tilde{x})^\top R_\beta(F_i(\tilde{x})) \right] \right\| \\
&\le L_f \left\| x - \tilde{x} \right\| + \tfrac{1}{\lambda} \sum_{i=1}^{p} \left\| JF_i(x)^\top R_\beta(F_i(x)) - JF_i(\tilde{x})^\top R_\beta(F_i(\tilde{x})) \right\| \\
&\le L_f \left\| x - \tilde{x} \right\| + \tfrac{1}{\lambda} \sum_{i=1}^{p} \left\| JF_i(x)^\top R_\beta(F_i(x)) - JF_i(\tilde{x})^\top R_\beta(F_i(x)) \right\| \\
&\qquad + \tfrac{1}{\lambda} \sum_{i=1}^{p} \left\| JF_i(\tilde{x})^\top R_\beta(F_i(x)) - JF_i(\tilde{x})^\top R_\beta(F_i(\tilde{x})) \right\| \\
&\le L_f \left\| x - \tilde{x} \right\| + \tfrac{1}{\lambda} \sum_{i=1}^{p} \left\| JF_i(x) - JF_i(\tilde{x}) \right\| \left\| R_\beta(F_i(x)) \right\| \\
&\qquad + \tfrac{1}{\lambda} \sum_{i=1}^{p} \left\| JF_i(\tilde{x}) \right\| \left\| R_\beta(F_i(x)) - R_\beta(F_i(\tilde{x})) \right\| \\
&\le L_f \left\| x - \tilde{x} \right\| + \tfrac{1}{\lambda} \sum_{i=1}^{p} L_i \left\| R_\beta(F_i(x)) \right\| \left\| x - \tilde{x} \right\| \\
&\qquad + \tfrac{1}{\lambda} \sum_{i=1}^{p} L_\beta \left\| JF_i(\tilde{x}) \right\| \left\| F_i(x) - F_i(\tilde{x}) \right\| \\
&\le L_f \left\| x - \tilde{x} \right\| + \tfrac{1}{\lambda} \sum_{i=1}^{p} L_i \left\| R_\beta(F_i(x)) \right\| \left\| x - \tilde{x} \right\| \\
&\qquad + \tfrac{1}{\lambda} \sum_{i=1}^{p} L_\beta K_i \left\| JF_i(\tilde{x}) \right\| \left\| x - \tilde{x} \right\|,
\end{aligned}
$$

where the Lipschitz modulus L_β is given in Proposition 3.6(i). We now aim to establish upper bounds for $\|R_\beta(F_i(x))\|$ and $\|JF_i(\tilde{x})\|$, which will be derived based on the point x^0. Specifically, we have

$$
\begin{aligned}
\left\| R_\beta(F_i(x)) \right\| &\le \left\| R_\beta(F_i(x)) - R_\beta(F_i(x^0)) \right\| + \left\| R_\beta(F_i(x^0)) \right\| \\
&\le L_\beta K_i \left\| x - x^0 \right\| + \left\| R_\beta(F_i(x^0)) \right\|
\end{aligned}
$$

and

$$
\begin{aligned}
\left\| JF_i(\tilde{x}) \right\| &\le \left\| JF_i(\tilde{x}) - JF_i(x^0) \right\| + \left\| JF_i(x^0) \right\| \\
&\le L_i \left\| \tilde{x} - x^0 \right\| + \left\| R_\beta(F_i(x^0)) \right\|.
\end{aligned}
$$

Lemma 5.3 establishes that the level set $\mathrm{lev}_{\le s_{\lambda,\beta}(x^0)}[s_{\lambda,\beta}; C]$ is compact, then so is $\mathrm{conv}(\mathrm{lev}_{\le s_{\lambda,\beta}(x^0)}[s_{\lambda,\beta}; C])$, which implies that the constant

$$
D_\lambda^0 := \sup_{x \in \mathrm{conv}(\mathrm{lev}_{\le s_{\lambda,\beta}(x^0)}[s_{\lambda,\beta}; C])} \left\| x - x^0 \right\| \tag{5.7}
$$

is finite. By summarizing the above results, we conclude that

$$\begin{aligned}&\left\|\nabla s_{\lambda,\beta}(x)-\nabla s_{\lambda,\beta}(\tilde{x})\right\|\\ &\le L_f\left\|x-\tilde{x}\right\|+\tfrac{1}{\lambda}\sum_{i=1}^{p}L_i\left(L_\beta K_i D_\lambda^0+\left\|R_\beta(F_i(x^0))\right\|\right)\left\|x-\tilde{x}\right\|\\ &\quad+\tfrac{1}{\lambda}\sum_{i=1}^{p}L_\beta K_i\left(L_i D_\lambda^0+\left\|R_\beta(F_i(x^0))\right\|\right)\left\|x-\tilde{x}\right\|.\end{aligned}$$

Therefore, the gradient $\nabla s_{\lambda,\beta}$ is Lipschitz continuous on $\mathrm{conv}(\mathrm{lev}_{\le s_{\lambda,\beta}(x^0)}[s_{\lambda,\beta};C])$ with the claimed modulus. □

In view of Lemma 5.9, a first-order inner solver can be used to solve the subproblem in Algorithm 2. In particular, we consider first-order solvers $\mathcal{S}_1$ that achieve a complexity bound of $\mathcal{O}(\epsilon^{-2})$ for finding an approximate stationary point of the subproblem $\mathrm{minimize}_{x\in C}\, s_{\lambda^\nu,\beta}(x)$. By combining Theorem 5.7 and Lemma 5.9, we can now establish the total complexity of Algorithm 2 for generating an ϵ-approximate C-stationary point of (1.1).

Theorem 5.10 (Total complexity) *Let Assumptions 5.1 and 5.2 hold, and suppose that a first-order solver $\mathcal{S}_1$ is used to solve the subproblem approximately. Consider the iterates generated by Algorithm 2 with λ^ν being updated as $\lambda^\nu=\rho\lambda^{\nu-1}$ with $\rho\in(0,1)$. Then, the last iterate $x^{\bar{\nu}}$ is an ϵ-approximate C-stationary point of (1.1) and at most*

$$\mathcal{O}\left(\epsilon^{-4}\log_{\frac{1}{\rho}}\left[\tfrac{4\lambda^0(f(x^0)-\underline{f})}{\epsilon^2}\right]\right)$$

gradient evaluations are needed.

Proof By following [18], to obtain an ϵ-stationary point x^ν of each subproblem, i.e.,

$$\mathrm{dist}\left(-\nabla f(x^\nu)-\tfrac{1}{\lambda^\nu}\textstyle\sum_{i=1}^{p}JF_i(x^\nu)^\top R_\beta(F_i(x^\nu)),\ N_C(x^\nu)\right)\le\epsilon,$$

the worst-case complexity bound of the inner solver $\mathcal{S}_1$ is given by

$$\frac{c_1 L_{\lambda^\nu}(x^0)\left(s_{\lambda^\nu,\beta}(x^{\nu,0})-\underline{s}_{\lambda^\nu,\beta}\right)}{\epsilon^2},\tag{5.9}$$

where $L_{\lambda^\nu}(x^0)$ is defined in Lemma 5.9, $x^{\nu,0}$ is the starting point of the subproblem and $c_1>0$ is a constant that depends on the first-order method. Note that x^0 is used in $L_{\lambda^\nu}(x^0)$ as specified in the step 2 of Algorithm 2, i.e., $s_{\lambda^\nu,\beta}(x^\nu)\le s_{\lambda^\nu,\beta}(x^0)$. Furthermore, $\underline{s}_{\lambda^\nu,\beta}$ is a lower bound to $s_{\lambda^\nu,\beta}$. It follows from (3.6), (4.4) and Assumption 5.1 that

$$s_{\lambda,\beta}(x)=f(x)+\tfrac{1}{\lambda}\sum_{i=1}^{p}r_\beta(F_i(x))\ge f(x)\ge\underline{f}$$

for all $x\in\mathbb{R}^n$. Therefore, we can let $\underline{s}_{\lambda^\nu,\beta}=\underline{f}$. We now aim to establish an upper bound for $s_{\lambda^\nu,\beta}(x^{\nu,0})$. To this end, let us start by establishing an upper bound of $1/\lambda^\nu$.

Suppose that $x^{\bar{\nu}}$ is the point returned by meeting $\sum_{i=1}^{p} r_\beta(F_i(x^{\bar{\nu}})) \leq \epsilon^2/2$, and let $\nu \in \{1, \ldots, \bar{\nu}\}$. Then, according to Theorem 5.7, we have

$$\frac{1}{\lambda^\nu} = \frac{1}{\rho^\nu} \cdot \frac{1}{\lambda^0} \leq \frac{1}{\rho^{\bar{\nu}}} \cdot \frac{1}{\lambda^0} = \frac{1}{\rho} \cdot \frac{1}{\rho^{\bar{\nu}-1}} \cdot \frac{1}{\lambda^0} < \frac{1}{\rho} \cdot \frac{4(f(x^0) - \underline{f})}{\epsilon^2}. \tag{5.10}$$

Furthermore, using the sandwich property Proposition 3.3(ii), the initial condition $\sum_{i=1}^{p} r_\beta(F_i(x^0)) \leq \epsilon^2/4$ implies that

$$\sum_{i=1}^{p} \lambda^\nu \operatorname{env}_{\lambda^\nu} \delta_D(F_i(x^0)) \leq \sum_{i=1}^{p} \lambda^\nu \operatorname{env}_{\lambda^\nu, \beta\lambda^\nu} \delta_D(F_i(x^0)) \leq \frac{\epsilon^2}{4}.$$

Then, by applying the formulation of the Moreau envelope in (3.2), we obtain

$$\sum_{i=1}^{p} \operatorname{dist}\left(F_i(x^0), D\right)^2 \leq \frac{\epsilon^2}{2}. \tag{5.11}$$

Using again the sandwich property Proposition 3.3(ii), we obtain an upper bound for $s_{\lambda^\nu,\beta}(x^{\nu,0})$:

$$\begin{aligned}
s_{\lambda^\nu,\beta}(x^{\nu,0}) \leq s_{\lambda^\nu,\beta}(x^0) &= f(x^0) + \sum_{i=1}^{p} \tfrac{1}{\lambda^\nu} r_\beta(F_i(x^0)) \\
&\leq f(x^0) + \sum_{i=1}^{p} \operatorname{env}_{\lambda^\nu(1-\beta)} \delta_D(F_i(x^0)) \\
\text{by (3.2)} \quad &= f(x^0) + \frac{1}{2\lambda^\nu(1-\beta)} \sum_{i=1}^{p} \operatorname{dist}\left(F_i(x^0), D\right)^2 \\
\text{by (5.11)} \quad &\leq f(x^0) + \frac{\epsilon^2}{4\lambda^\nu(1-\beta)} \\
\text{by (5.10)} \quad &< \underbrace{f(x^0) + \frac{f(x^0) - \underline{f}}{\rho(1-\beta)}}_{c_2 :=}.
\end{aligned} \tag{5.12}$$

Recalling that D_λ^0 is finite for all $\lambda > 0$, as established in Lemma 5.9, it follows that the constant

$$L(x^0) := \max_{\nu \in \{1,2,\ldots,\bar{\nu}\}} \tilde{L}_{\lambda^\nu}(x^0) \tag{5.13}$$

is also finite, where $\tilde{L}_\lambda(x^0)$ is defined in (5.5). In this way, by (5.9), the total number of gradient evaluations in Algorithm 2 is bounded by

$$\sum_{\nu=1}^{\bar{\nu}} c_1 L_{\lambda^\nu}(x^0) \left(s_{\lambda^\nu,\beta}(x^{\nu,0}) - \underline{f}\right) \epsilon^{-2} \leq \sum_{\nu=1}^{\bar{\nu}} c_1 L_{\lambda^\nu}(x^0) \left(s_{\lambda^\nu,\beta}(x^0) - \underline{f}\right) \epsilon^{-2}$$

$$\text{by (5.4) and (5.12)} < \sum_{\nu=1}^{\bar{\nu}} c_1(c_2 - \underline{f}) \left(L_f + \tfrac{1}{\lambda^\nu} \tilde{L}_{\lambda^\nu}(x^0)\right) \epsilon^{-2}$$

$$\text{by (5.13)} \leq \sum_{\nu=1}^{\bar{\nu}} c_1(c_2 - \underline{f}) \left(L_f + \tfrac{1}{\lambda^\nu} L(x^0)\right) \epsilon^{-2}.$$

Finally, with the aid of Theorem 5.7, the above result leads to

$$\begin{aligned}
&\sum_{\nu=1}^{\bar{\nu}} c_1(c_2 - \underline{f}) \left(L_f + \tfrac{1}{\lambda^\nu} L(x^0)\right) \epsilon^{-2} \\
\text{by (5.10)} &\leq c_1(c_2 - \underline{f}) \bar{\nu} \left(L_f + \tfrac{4(f(x^0) - \underline{f})}{\rho \epsilon^2} L(x^0)\right) \epsilon^{-2} \\
\text{by (5.3)} &< c_1(c_2 - \underline{f}) \left(1 + \log_{\frac{1}{\rho}} \left(\tfrac{4\lambda^0 (f(x^0) - \underline{f})}{\epsilon^2}\right)\right) \left(L_f + \tfrac{4(f(x^0) - \underline{f})}{\rho \epsilon^2} L(x^0)\right) \epsilon^{-2}.
\end{aligned}$$

Therefore, we conclude that the total complexity of Algorithm 2 to obtain an ϵ-approximate C-stationary point is $\mathcal{O}(\epsilon^{-4} \log_{\frac{1}{\rho}} [4\lambda^0 (f(x^0) - \underline{f}) \epsilon^{-2}])$. □

5.4 Two-Stage Approach

The complexity bound in the previous subsection is derived under the assumption that the starting point of the homotopy method is near-feasible, namely such that $\sum_{i=1}^{p} r_\beta(F_i(x^0)) \leq \epsilon^2/4$. To obtain such a point, we solve the following problem starting from an (infeasible) initial point $\tilde{x}^0$:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ \mathcal{G}(x) = \sum_{i=1}^{p} r_\beta(F_i(x)) + \delta_C(x), \tag{5.14}$$

where δ_C is the indicator function of the set C. The solution x^0 of (5.14) is then used to initialize Algorithm 2. This yields a two-stage approach for solving (1.1). Proposition 3.3(iii) implies that the global minimum of $\sum_{i=1}^{p} r_\beta \circ F_i$ is 0, which occurs when x is feasible for all F_i. Moreover, since we assume $\{x \in C \mid F_i(x) \in D, i = 1, \ldots, p\}$ is nonempty, the global minimum of (5.14) is 0. Therefore, the problem (5.14) can be interpreted as a feasibility problem for (1.1).

The above observation motivates the introduction of the *proximal-PL* property for (5.14), as it ensures that all first-order stationary points attain the global minimum; see [29, Section 4] for details. Although Proposition 3.6(ii) establishes that r_β has the PL property, this does not necessarily imply that $\mathcal{G}$ has the proximal-PL property. However, as shown in [29, Appendix G], the proximal-PL property is equivalent to the *Kurdyka-Łojasiewicz (KL)* property over $\operatorname{dom} \mathcal{G}$. Therefore, in the special case of $p = 1$, we provide the following result to characterize conditions under which $r_\beta \circ F_i + \delta_C$ has the KL property over $\operatorname{dom}(r_\beta \circ F_i + \delta_C)$. To maintain notational consistency, we retain the subscript i of F_i even if $p = 1$.

Proposition 5.11 (KL property of $r_\beta \circ F_i + \delta_C$) *Fix $\beta \in (0,1)$, let r_β, R_β be as in (3.4) and (3.5), and $F_i(x) = (G_i(x), H_i(x))$. Suppose that*

either (i) $\nabla G_i(x)$ and $\nabla H_i(x)$ are linearly independent for all $x \in \mathbb{R}^n$,
or (ii) $\nabla G_i(x) \neq 0$ and $H_i(x) = 0$ for all $x \in \mathbb{R}^n$.

Then, any point x such that $\nabla(r_\beta \circ F_i)(x) = 0$ satisfies $F_i(x) \in D$.
If the conditions can be strengthened to

either (i') $\inf_{x\in\mathbb{R}^n} \lambda_{\min}(JF_i(x)JF_i(x)^\top) > 0$, where $\lambda_{\min}$ denotes the minimal eigenvalue,
or (ii') $\inf_{x\in\mathbb{R}^n} \|\nabla G_i(x)\| > 0$ and $H_i(x) = 0$,

and the following inclusion

$$\mathrm{B}\left(-\nabla(r_\beta \circ F_i)(x), \left\|\nabla(r_\beta \circ F_i)(x)\right\|\right) \cap N_C(x) = \{0\} \tag{5.15}$$

holds, then the composite function $r_\beta \circ F_i + \delta_C$ has the KL property over $\mathrm{dom}(r_\beta \circ F_i + \delta_C)$.

Proof Let us first consider condition (i). We have

$$\begin{aligned}\left\|\nabla(r_\beta \circ F_i)(x)\right\|^2 &= \left\|JF_i(x)^\top R_\beta(F_i(x))\right\|^2 \\ &= R_\beta(F_i(x))^\top \left(JF_i(x)JF_i(x)^\top\right) R_\beta(F_i(x)).\end{aligned}$$

Linear independence of $\nabla G_i(x)$ and $\nabla H_i(x)$ implies that the minimal eigenvalue of $JF_i(x)JF_i(x)^\top$, denoted by $q_i(x)$, satisfies $q_i(x) > 0$. Then, by Proposition 3.6(ii) , we obtain

$$\begin{aligned}\tfrac{1}{2}\left\|\nabla(r_\beta \circ F_i)(x)\right\|^2 &\geq \tfrac{q_i(x)}{2}\left\|R_\beta(F_i(x))\right\|^2 \\ &\geq q_i(x) l_\beta(r_\beta \circ F_i)(x).\end{aligned} \tag{5.16}$$

Since $q_i(x) > 0$, if $\nabla(r_\beta \circ F_i)(x) = 0$ one must have $r_\beta(F_i(x)) = 0$, which by Proposition 3.3(iii) implies that $F_i(x) \in \arg\min \delta_D = D$.

For condition (ii), the transpose of the Jacobian becomes $JF_i(x)^\top = (\nabla G_i(x), 0)$. Let $z = (z_1, z_2) = (z_1, 0)$. Then, according to (3.4) and (3.5), we have

$$r_\beta(z) = \begin{cases} \frac{1}{2(1-\beta)} z_1^2 \\ 0 \end{cases} \quad \text{and} \quad R_\beta(z) = \begin{cases} \frac{1}{1-\beta}\binom{z_1}{0} & \text{if } z_1 \leq 0 \\ \binom{0}{0} & \text{if } z_1 > 0. \end{cases}$$

With $z_1 := G_i(x) \leq 0$, it holds that

$$\begin{aligned}\tfrac{1}{2}\left\|\nabla(r_\beta \circ F_i)(x)\right\|^2 &= \tfrac{1}{2}(R_\beta(F_i(x)))_1^2 \|\nabla G_i(x)\|^2 \\ &= \tfrac{1}{2}\left(\tfrac{1}{1-\beta} z_1\right)^2 \|\nabla G_i(x)\|^2 \\ &= \tfrac{1}{1-\beta}\|\nabla G_i(x)\|^2 (r_\beta \circ F_i)(x).\end{aligned} \tag{5.17}$$

According to the assumption, we have $\|\nabla G_i(x)\| > 0$, and again we conclude that $r_\beta(F_i(x)) = 0$. Finally, we have $r_\beta(F_i(x)) = 0$ trivially when $z_1 := G_i(x) > 0$.

Let us now consider the KL property under condition (5.15), along with either condition (i') or (ii'). Since $\inf_{x\in\mathbb{R}^n} q_i(x) > 0$ under condition (i') and $\inf_{x\in\mathbb{R}^n} \|\nabla G_i(x)\| > 0$ under condition (ii'), in both cases it follows from (5.16) and (5.17) that $r_\beta \circ F_i$ satisfies the PL inequality with some constant $l_{\beta,i} > 0$. According to [29, Appendix G], the KL property of $r_\beta \circ F_i + \delta_C$ is defined by the existence of a constant $\tilde{l}_\beta > 0$ such that

$$\min_{s\in\partial(r_\beta\circ F_i+\delta_C)(x)} \|s\|^2 \;\geq\; 2\tilde{l}_\beta \left(r_\beta \circ F_i + \delta_C\right)(x)$$

for all $x \in \mathrm{dom}(r_\beta \circ F_i + \delta_C)$. In our case where $r_\beta \circ F_i$ is differentiable and δ_C is convex, we have

$$\partial(r_\beta \circ F_i + \delta_C)(x) = \left\{\nabla(r_\beta \circ F_i)(x) + \xi \;\middle|\; \xi \in N_C(x)\right\}.$$

Based on this, let us consider the inequality:

$$\left\|\xi\right\|^2 + 2\left\langle\nabla(r_\beta \circ F_i)(x), \xi\right\rangle \leq 0 \quad\Leftrightarrow\quad \left\|\xi + \nabla(r_\beta \circ F_i)(x)\right\| \leq \left\|\nabla(r_\beta \circ F_i)(x)\right\|.$$

Therefore, the condition (5.15) implies that

$$\inf_{\xi\in N_C(x)} \left(\left\|\xi\right\|^2 + 2\left\langle\nabla(r_\beta \circ F_i)(x), \xi\right\rangle\right) \geq 0. \tag{5.18}$$

Let $\xi \in N_C(x)$, then we have

$$\begin{aligned}
&\left\|\nabla(r_\beta \circ F_i)(x) + \xi\right\|^2 \\
=\;& \left\|\nabla(r_\beta \circ F_i)(x)\right\|^2 + \left\|\xi\right\|^2 + 2\left\langle\nabla(r_\beta \circ F_i)(x), \xi\right\rangle \\
\geq\;& \left\|\nabla(r_\beta \circ F_i)(x)\right\|^2 + \inf_{\tilde{\xi}\in N_C(x)} \left(\left\|\tilde{\xi}\right\|^2 + 2\left\langle\nabla(r_\beta \circ F_i)(x), \tilde{\xi}\right\rangle\right) \\
\text{by (5.18)}\;\geq\;& \left\|\nabla(r_\beta \circ F_i)(x)\right\|^2 \\
\text{by (i') or (ii')}\;\geq\;& 2l_{\beta,i}(r_\beta \circ F_i)(x).
\end{aligned}$$

Since the condition (5.15) implies $N_C(x) \neq \emptyset$, thus $x \in C$ and we have

$$\left\|\nabla(r_\beta \circ F_i)(x) + \xi\right\|^2 \geq 2l_{\beta,i}\left(r_\beta \circ F_i + \delta_C\right)(x).$$

In this way, we conclude that $r_\beta \circ F_i + \delta_C$ has the KL property over $\mathrm{dom}(r_\beta \circ F_i + \delta_C)$. □

Remark 5.12 Note that the null gradient in conditions (ii) and (ii') of Proposition 5.11 correspond to the setting in Example 2.6. Furthermore, although conditions (i), (ii), (i') and (ii') may be difficult to verify for all $x \in \mathbb{R}^n$, an important and commonly encountered special case arises when the constraints are linear. Specifically, consider

$F_i(x) = (a_i^\top x + b_i, c_i^\top x + d_i)$, where $a_i, c_i \in \mathbb{R}^n$ and $b_i, d_i \in \mathbb{R}$. Then, all conditions in Proposition 5.11 simply reduce to: either (i) a_i and b_i being linearly independent, or (ii) $a_i \neq 0$ and $c_i = 0, d_i = 0$.

Assuming that all conditions in Proposition 5.11 are satisfied, the result in [29, Appendix G] implies that $r_\beta \circ F_i + \delta_C$ has the proximal-PL property. Then, according to [29, Theorem 5], the projected gradient method with the step-size $1/L$,

$$\tilde{x}^{\nu+1} = \Pi_C\left(\tilde{x}^\nu - \tfrac{1}{L}\nabla(r_\beta \circ F_i)(\tilde{x}^\nu)\right)$$

converges linearly to the global minimum 0 from the starting point $\tilde{x}^0$, where L is the Lipschitz modulus of $r_\beta \circ F_i$. Therefore, the complexity to meet $(r_\beta \circ F_i)(x^0) \leq \epsilon^2/4$ is $\mathcal{O}(\log \frac{1}{\epsilon})$. When the Lipschitz modulus L is unknown, backtracking is commonly used to determine the step-size. In practice, however, it is well known that backtracking often yields faster practical convergence than the fixed step-size, even when L is known. Excluding function evaluations needed in the linesearch, the complexity remains $\mathcal{O}(\log \frac{1}{\epsilon})$ calls to the gradient in this case.

The condition (5.15) in Proposition 5.11 imposes an additional geometric structure between F_i and C that ensures the KL property. However, this condition is nontrivial to verify. When $p > 1$, the situation becomes even more complex. The condition (5.15) now becomes

$$\mathrm{B}\left(-\textstyle\sum_{i=1}^p \nabla(r_\beta \circ F_i)(x), \left\|\sum_{i=1}^p \nabla(r_\beta \circ F_i)(x)\right\|\right) \cap N_C(x) = \{0\}$$

and, in addition, the following inequality is required:

$$\left\|\textstyle\sum_{i=1}^p \nabla(r_\beta \circ F_i)(x)\right\|^2 \geq \kappa \displaystyle\sum_{i=1}^p \left\|\nabla(r_\beta \circ F_i)(x)\right\|^2 \quad \text{for some } \kappa > 0. \tag{5.19}$$

Eq. (5.19) can be interpreted as a *gradient alignment* condition, which requires that the gradients do not cancel each other out excessively, or, equivalently, exhibit some degree of positive correlation. The problem structures under which (5.15) and (5.19) naturally arise will be explored in future work. Nevertheless, even when $\mathcal{G}$ lacks the proximal-PL property, we can still solve (5.14) for a starting point of Algorithm 2, albeit heuristically, as the method then only guarantees convergence to a stationary point of (5.14). The complete algorithm is presented in Algorithm 3, and its effectiveness is demonstrated through a range of numerical experiments in the next section.

6 Numerical Results

6.1 Implementation

In this section, we test Algorithm 3 on several benchmark problems. Following Example 2.6, all equality and inequality constraints are reformulated as CCs, which leads to

Algorithm 3 LL-based homotopy (two-stage) approach

Initialization: Let $\nu = 1$. Given $\epsilon > 0$, $\lambda^0 > 0$, $\beta \in (0, 1)$ and $\tilde{x}^0 \in \mathbb{R}^n$.

Stage 1:

Find an approximate solution x^0 to the problem $\underset{x \in C}{\text{minimize}} \sum_{i=1}^p r_\beta(F_i(x))$ such that

either $\sum_{i=1}^p r_\beta(F_i(x^0)) \leq \epsilon^2/4$ ▷ ideal: near-feasible point

or $\text{dist}(-\nabla(\sum_{i=1}^p r_\beta \circ F_i)(x^0), N_C(x^0)) \leq \epsilon$ ▷ heuristic: ϵ-stationary point

by using an inner solver from the starting point $\tilde{x}^0$.

Stage 2:

Repeat:

Choose λ^ν and set $s_{\lambda^\nu,\beta} := f + \sum_{i=1}^p \frac{1}{\lambda^\nu} r_\beta \circ F_i$.

Find an approximate solution x^ν to the problem $\text{minimize}_{x \in C}\, s_{\lambda^\nu,\beta}(x)$ such that

- $s_{\lambda^\nu,\beta}(x^\nu) \leq \min\{s_{\lambda^\nu,\beta}(x^{\nu-1}), s_{\lambda^\nu,\beta}(x^0)\}$ and
- $\text{dist}(-\nabla s_{\lambda^\nu,\beta}(x^\nu), N_C(x^\nu)) \leq \epsilon$

by using an inner solver with starting point $x^{\nu,0} := \arg\min\{s_{\lambda^\nu,\beta}(\hat{x}) \mid \hat{x} \in \{x^{\nu-1}, x^0\}\}$.

Let $\nu \leftarrow \nu + 1$.

Until: The stopping criterion $\sum_{i=1}^p r_\beta(F_i(x^\nu)) \leq \epsilon^2/2$ is satisfied.

the form of (1.1). Algorithm 3 with a first-order inner solver is referred to as *LL2(1st)* (two-stage LL-based homotopy approach with a first-order inner solver). The method without trying to find a feasible point at the beginning is called *LL1* (one-stage LL-based homotopy approach). Furthermore, we observe that the second-order solver [9] may perform well in stage 1, and result in a better final solution. Therefore, it is also included in the comparison and referred to as *LL2(2nd)*. It is well known that the Scholtes method [45] exhibits strong numerical performance compared to other regularization methods in practice [20]. It is implemented as follows:

$$\begin{aligned} \underset{x \in C}{\text{minimize}} \;\; & f(x) \\ \text{subject to} \;\; & G(x) \geq 0,\; H(x) \geq 0,\; G_i(x)H_i(x) \leq \tau,\;\; i \in \{1, \ldots, p\} \\ & c_i(x) = 0,\; i \in \mathcal{E} \\ & c_i(x) \leq 0,\; i \in \mathcal{I}, \end{aligned}$$

where $\tau \in \mathbb{R}_+$ is the relaxation parameter. We also implemented the LL method without relaxing the general equality and inequality constraints:

$$\begin{aligned} \underset{x \in C}{\text{minimize}} \;\; & f(x) + \sum_{i=1}^p \text{env}_{\lambda,\beta\lambda}\, \delta_D(F_i(x)) \\ \text{subject to} \;\; & c_i(x) = 0,\; i \in \mathcal{E} \\ & c_i(x) \leq 0,\; i \in \mathcal{I}, \end{aligned}$$

which is referred to as *LL penalty*. For the comparison of open source and commercial solvers, we used ALGENCAN [1], IPOPT [47] and MINOS [37]. These NLP solvers

have been shown to be efficient for MPCCs to some extent, as demonstrated in [23, 30, 34, 40]. Following the methods in [23, 30], we report the performance of NLP solvers based on the following formulation:

$$G(x) \geq 0, \;\; H(x) \geq 0 \quad \text{and} \quad \langle G(x), H(x) \rangle \leq 0.$$

We set the tolerance as $\epsilon = 10^{-8}$ for LL-based, Scholtes and NLP methods. The parameters for LL-based methods are: $\beta = 0.999$, $\rho = 0.8$ and $\lambda^0 = 1$. The parameters in [45] are adopted for Scholtes method, i.e., $\tau^\nu = 0.1\tau^{\nu-1}$ and $\tau^0 = 1$. The maximum number of outer iterations for the LL-based and Scholtes methods is 200. Eq. (5.1) and (5.2) serve as the stopping criteria for the LL-based and Scholtes methods, respectively. The above methods are implemented in MATLAB (R2024b). L-BFGS-B[1] is used as the inner solver for LL1, LL2(1st) and stage 2 of LL2(2nd), with the memory parameter set to 5. The problems are modeled in AMPL (A Mathematical Programming Language) [16]. The NLP solvers are invoked via AMPL MATLAB API.

6.2 MacMPEC Benchmark

In order to obtain quantitative results, we first benchmark Algorithm 3 on the MacMPEC collection.[2] As a preliminary result, we use 77 problems from the collection with fewer than 100 variables. For each test problem, we perform 20 runs from randomly generated starting points, where each element is a uniform random value in the interval $[-50, 50]$. Table 1 reports the number of cases in which the algorithm converges to an optimal point, a suboptimal but feasible point, an infeasible point, or encounters a solver failure. The returned objective value is regarded as optimal if its relative suboptimality with respect to the best known value, i.e., $(|\text{obj}-\text{obj}^*|/|\text{obj}^*|) \times 100\%$, is within 5%. We use IPOPT as the inner solver for both the Scholtes and LL penalty methods, as it outperforms other NLP solvers on MacMPEC benchmark. Table 1 shows that the LL penalty performs best in finding the optimal solution, and it is followed by the Scholtes method. However, the Scholtes method returns more infeasible points compared to LL-based methods. LL2(1st) and LL2(2nd) are particularly effective at avoiding infeasible points, demonstrating the effectiveness of the two-stage approach. Since LL2(2nd) initializes Algorithm 2 using a second-order stationary point of problem (5.14), it has a higher likelihood of converging to the global minimizer. In contrast, LL2(1st), which uses only a first-order stationary point, may be more prone to converging to local minimizers. While LL1 returns the highest number of infeasible points among the LL-based methods, it still performs better in this regard than the Scholtes method and other NLP solvers. Overall, we conclude that Algorithm 3 exhibits superior robustness compared to the other methods. It consistently translates infeasible points into at least suboptimal points, indicating strong potential for generating high-quality solutions in practice.

[1] https://github.com/stephenbeckr/L-BFGS-B-C

[2] https://wiki.mcs.anl.gov/leyffer/index.php/MacMPEC

Table 1 Problem number counting of optimal, suboptimal and infeasible points, and solver failures

	LL1	LL penalty	LL2 (1st)	LL2 (2nd)	Scholtes	ALGENCAN	IPOPT	MINOS
optimal	1262	1372	1273	1286	1339	1240	1293	1076
suboptimal	243	148	251	238	151	148	171	340
infeasible	35	20	16	16	50	49	76	105
solver failure	0	0	0	0	0	103	0	19

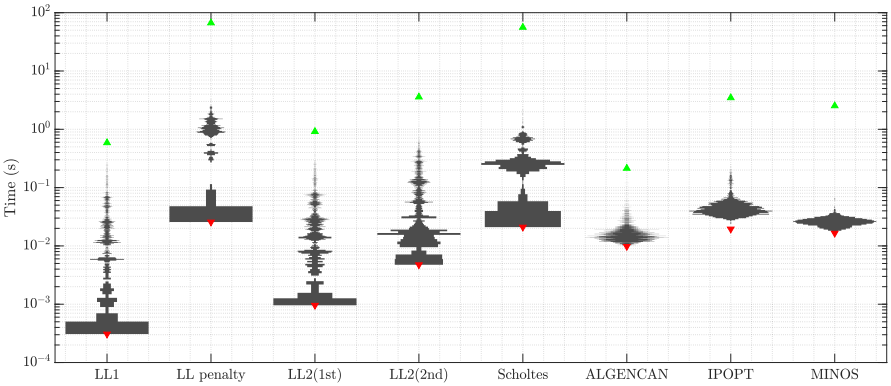


Fig. 5 Computation time (time value distribution and minimal/maximal values)

The computation time of each method is presented in Fig. 5, excluding cases of solver failure. LL1 and LL2(1st) present higher computation efficiency compared to the LL penalty, LL2(2nd) and Scholtes method, which is not surprising since it benefits from the fast first-order solver for smooth subproblems. Although the LL penalty and Scholtes methods achieve better performance in terms of solution quality, see Table 1, they incur significantly higher computational costs due to their reliance on IPOPT as the inner solver. On the other hand, LL2(2nd) is slower than LL2(1st) since the second-order solver is used in stage 1. ALGENCAN, IPOPT and MINOS exhibit more consistent and less volatile computation times, which come from their efficient and mature implementation. Even though the current implementation of Algorithm 3 is a prototype in MATLAB and not optimized, it is still competitive in terms of the computation efficiency to the NLP solvers on the MacMPEC benchmark.

6.3 Bound-Constrained Quadratic MPCCs

In this part, we consider the larger size MPCCs with quadratic objective and bound-constrained variables:

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & \tfrac{1}{2}x^\top Qx + g^\top x \\ \text{subject to} \quad & l_0 \le x_0 \le u_0 \\ & 0 \le x_1 \perp x_2 \ge 0, \end{aligned} \tag{6.1}$$

where $x := (x_0, x_1, x_2) \in \mathbb{R}^n$, $x_0 \in \mathbb{R}^{n_0}$, $x_1, x_2 \in \mathbb{R}^p$, $Q \in \mathbb{R}^{n\times n}$ is a symmetric matrix, $g \in \mathbb{R}^n$ is a vector, and the lower and upper bounds satisfy $(l_0)_i < (u_0)_i$ for $i \in \{1, \ldots, n_0\}$. The problem of the form (6.1) is studied in [30], as it arises as a subproblem within an augmented Lagrangian framework for solving more general MPCCs. Since it also constitutes a special case of (1.1), it provides a meaningful benchmark for evaluating our algorithm. A batch of test problems are randomly generated by using the code[3] in [30] and the problem information is: Q is sparse with the density $n^2/4$, whose elements are normally distributed, the elements in g are uniform random values in the interval $[-10, 10]$, the bounds l_0, u_0 are uniform random values in the interval $[-10, 10]$ and $[0, 20]$. In order to increase the difficulty of solving, the matrix Q is generated to be indefinite. We consider two problem sizes: $(n_0 = 100, p = 200)$ and $(n_0 = 200, p = 400)$, where the total number of variables is 500 and 1000. We randomly generate 10 problems for each size. The comparison with MINOS is excluded due to instability issues observed in the previous subsection. ALGENCAN is used as the inner solver of the Scholtes method due to its good performance on (6.1), as shown in Table 2. Since larger problem sizes are considered, we focus primarily on the first-order methods among the LL-based methods from this point onward. Furthermore, it is trivial to find a feasible point of (6.1), so we initialize all methods with $x^0 = (l_0, 0, 0)$ without loss of generality. As a result, stage 1 of the LL-based method is not required. Noting that only bound constraints are involved in (6.1), LL penalty is equivalent to LL1.

In this subsection, we include a comparison with the projected gradient method (PGM), taking advantage of the fact that the constraint set in (6.1) admits an efficient projection. Specifically, we define the projection operator and the objective function as

$$\Pi_{[l_0,u_0]\times D}(x) := (\Pi_{[l_0,u_0]}(x_0), \Pi_D((x_1, x_2))) \quad \text{and} \quad f(x) := \tfrac{1}{2}x^\top Qx + g^\top x.$$

The PGM with Armijo linesearch for solving problem (6.1) begins by initializing x^0 and setting parameters $\epsilon > 0$, $\eta_{\max} > 0$, $\gamma \in (0, 1)$ and $c \in (0, 1)$. At each iteration, the algorithm sets $\eta \leftarrow \eta_{\max}$, computes the projected trial point $x_{\text{trial}} \leftarrow \Pi_{[l_0,u_0]\times D}(x^\nu - \eta\nabla f(x^\nu))$, and backtracks via $\eta \leftarrow \gamma\eta$ until the sufficient decrease condition

$$f(x_{\text{trial}}) \le f(x^\nu) - c\langle \nabla f(x^\nu), x^\nu - x_{\text{trial}}\rangle$$

is satisfied. The next iterate is then set as $x^{\nu+1} \leftarrow x_{\text{trial}}$, and the process terminates when $\|(x^{\nu+1} - x^\nu)/\eta\| \le \epsilon$. This method follows the standard PGM framework [6, Section 2.3], with commonly used parameter values: $\eta_{\max} = 1$, $\gamma = 0.5$ and $c = 10^{-4}$. The tolerance in PGM is consistent with the other methods, i.e., $\epsilon = 10^{-8}$.

The objective value, the maximal violation of CCs, i.e., $\|\min\{x_1, x_2\}\|_\infty$, and the computation time of each method are presented in Table 2. The minimal objective value in each problem is highlighted in bold and the minimal computation time is in bold with an underline. The suboptimal objective value is reported in terms of relative suboptimality. Computation times exceeding 50 seconds are rounded to the nearest

[3] https://wiki.mcs.anl.gov/leyffer/index.php/BndMPCC

Table 2 Objective, constraint violation and computation time on bounded quadratic MPCCs

n_0–p–num.	LL1			Scholtes			ALGENCAN			IPOPT			PGM	
	obj.	vio.	time(s)	obj.	vio.	time(s)	obj.	vio.	time(s)	obj.	vio.	time(s)	obj.	time(s)
100–200–0	0.90%	1.1e-11	1.82	**-58355.23**	2.9e-09	6.37	2.78%	0	**1.41**	4.29%	1.4e-05	58	1.24%	4.64
100–200–1	0.46%	1.2e-11	2.10	**-78425.76**	2.2e-09	11.12	3.79%	0	0.79	4.41%	7.3e-06	70	1.11%	**0.11**
100–200–2	3.61%	1.9e-10	1.82	**-46340.69**	2.8e-09	10.07	13.89%	0	0.68	14.45%	2.0e-05	66	7.29%	**0.19**
100–200–3	**-64054.32**	4.4e-11	1.53	0.21%	2.5e-09	7.46	2.46%	0	0.63	3.90%	1.5e-05	82	0.21%	**0.16**
100–200–4	**-55115.72**	1.1e-10	1.69	0.53%	2.6e-09	11.52	11.88%	0	1.15	15.87%	1.4e-05	94	0.01%	**0.09**
100–200–5	**-67241.94**	1.0e-10	1.86	0.83%	2.8e-09	6.99	5.57%	0	0.89	16.00%	2.5e-05	75	0.47%	**0.19**
100–200–6	0.70%	1.6e-11	1.88	**-71063.23**	3.6e-09	10.59	0.79%	0	0.72	2.94%	2.8e-05	62	0.64%	**0.09**
100–200–7	**-61504.67**	4.9e-11	2.04	4.21%	0	9.08	8.17%	0	0.93	20.39%	5.3e-06	82	9.58%	**0.06**
100–200–8	**-54981.98**	1.6e-11	2.30	0.34%	2.4e-09	19.86	22.47%	0	0.89	14.95%	2.4e-05	138	8.01%	**0.10**
100–200–9	**-54423.15**	1.6e-10	1.66	0.15%	3.0e-09	7.52	1.89%	0	0.53	4.35%	1.9e-05	77	0.19%	**0.28**
200–400–0	1.09%	3.9e-12	11.73	**-210017.27**	9.4e-09	93	7.02%	0	8.65	4.57%	6.0e-05	883	1.52%	**0.11**
200–400–1	4.50%	2.0e-11	10.89	**-190225.77**	0	126	6.41%	0	8.33	8.76%	1.6e-05	1296	0.38%	**0.12**
200–400–2	4.25%	3.0e-11	8.63	**-195462.75**	0	106	8.53%	0	7.01	6.45%	1.1e-04	1551	5.26%	**0.15**
200–400–3	1.51%	5.8e-12	10.80	1.00%	3.5e-09	104	4.52%	0	6.49	7.99%	5.8e-05	811	**-176247.85**	**0.16**
200–400–4	1.54%	6.5e-11	8.59	2.08%	7.3e-09	196	8.19%	0	5.16	6.37%	1.3e-05	743	**-197038.66**	**3.07**
200–400–5	6.29%	8.1e-12	8.51	1.93%	4.6e-09	93	5.37%	0	8.15	9.71%	4.4e-05	907	**-175486.10**	**0.23**
200–400–6	**-172420.88**	6.9e-12	9.80	5.81%	3.6e-09	89	6.36%	0	**6.54**	3.66%	3.5e-05	1619	4.83%	9.26
200–400–7	**-192654.03**	1.0e-10	7.76	4.41%	7.9e-10	116	7.41%	0	6.89	7.97%	1.8e-05	943	4.27%	**0.14**
200–400–8	**-147134.93**	3.4e-12	9.31	2.62%	0	151	7.27%	0	**5.31**	3.59%	7.1e-05	1049	3.80%	9.65
200–400–9	**-166871.82**	8.5e-11	7.98	0.67%	0	78	2.87%	0	8.99	10.34%	2.3e-05	754	6.24%	**0.16**

integer. The column for constraint violation of the PGM is omitted, since the projection operator $\Pi_{[l_0,u_0]\times D}$ naturally ensures feasibility for problem (6.1). Notably, ALGENCAN also ensures feasibility, as reported by its zero violation in good performance on Table 2. However, if a small violation of CCs (on the level of 1e-10) is acceptable, then LL1 is preferable, as it yields improved objective values in many instances. The Scholtes method delivers comparable solution quality to LL1 but is significantly slower. On the other hand, it is expected that the PGM is the fastest method overall, as it employs a single-loop structure (requiring no inner solver) and operates over a simple constraint set. However, it typically does not outperform LL1 in terms of solution quality. Finally, IPOPT reports a struggling performance on (6.1) with indefinite Hessian Q. Overall, we find that the LL-based method is efficient at finding high-quality solutions for large-scale bound-constrained quadratic MPCCs.

6.4 Linearly Constrained Quadratic MPCCs

In this subsection, we consider some larger size MPCCs with quadratic objective and linearly constrained variables:

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \;& \tfrac{1}{2}x^\top Qx + g^\top x \\ \text{subject to} \;& Ax_0 + a \le 0 \\ & 0 \le x_1 \perp Nx_0 + Mx_1 + q \ge 0, \end{aligned} \tag{6.2}$$

where $x := (x_0, x_1) \in \mathbb{R}^n$, $x_0 \in \mathbb{R}^{n_0}$ and $x_1 \in \mathbb{R}^p$. Test problems are randomly generated by using the code[4] in [25], which also provides the global minima. We use the same number of variables as the previous subsection, i.e., $n = 500, 1000$. Specifically, we choose ($n_0 = 100$, $p = 400$) and ($n_0 = 200$, $p = 800$), noting that $n = n_0 + p$ for (6.2). For the problem data, we set the following: the matrix Q is positive definite with a condition number 10^4, the number of rows of A is $1/2 \times n_0$, and the matrix M is nonsymmetric and nonmonotone ($M + M^\top$ is not positive semidefinite) with a condition number 2×10^4. The optimality conditions of (6.2) follow from Definition 2.1, in which the primal-dual feasibility and complementary slackness are

$$\begin{aligned} & (Ax_0 + a)_i \le 0,\ \alpha_i \ge 0 \text{ and } \alpha_i (Ax_0 + a)_i = 0,\ i \in \mathcal{I} \\ & x_1 \ge 0,\ Nx_0 + Mx_1 + q \ge 0,\ y_i^G x_1 = 0 \text{ and } y_i^H (Nx_0 + Mx_1 + q) = 0,\ i \in \mathcal{P}. \end{aligned}$$

By following [25, Definition 1], an index $i \in \mathcal{I}$ is called *first-level degeneracy* if $\alpha_i = (Ax_0 + a)_i = 0$, an index $i \in \mathcal{P}$ is called *second-level degeneracy* if $i \in I_{00}$ and an index $i \in I_{00}$ is called *mixed degeneracy* if either $y_i^G = 0$ or $y_i^H = 0$, in which the set I_{00} is defined in (2.2), and $\mathcal{I}, \mathcal{P}$ are index sets of inequality and complementarity constraints, respectively. We first test problems with *zero degeneracy*, i.e., the cardinalities of the first-level, second-level, and mixed degeneracy index sets are set to 0. Subsequently, to increase the problem-solving difficulty, we generate

[4] https://www3.eng.cam.ac.uk/~dr241/Papers/qpecgen.m

problems with *high degeneracy*. Specifically, we choose the index set cardinality of the first-level degeneracy as $3/5$ of the number of rows of A, the second-level degeneracy as $3/4 \times p$ and the mixed degeneracy as $2/3$ of the second-level degeneracy. Furthermore, for each problem size, 10 random problems are generated and tested, with the starting point of each instance randomly selected from the interval $[-50, 50]$.

The results are presented in Table 3. We omit the Scholtes method from comparison, as it requires even more computational time than ALGENCAN. In most instances, LL-based methods find the best objective values, which are consistent with the global minima provided by the problem generator, and they are also faster than NLP solvers. Therefore, it is expected that LL penalty offers limited improvement in objective values, while incurring higher computational costs due to the use of an NLP solver as the inner solver. For these reasons, LL penalty is also omitted in Table 3. Although ALGENCAN also returns high-quality solutions, it performs significantly slower, especially compared to the fast computation observed in the previous subsection. This indicates inefficiency in handling complex CCs and infeasible starting points. While IPOPT is faster than ALGENCAN, it returns lower-quality solutions, often yielding highly suboptimal solutions ($>100\%$ deviation). On the other hand, the computation times of ALGENCAN and IPOPT increase significantly when the degeneracy level is raised from *zero* to *high*. In contrast, LL-based methods demonstrate robust performance regardless of degeneracy. It is also worth noting that LL-based methods violate the CCs minimally compared to ALGENCAN and IPOPT. Therefore, we conclude that LL-based methods are efficient for solving large-scale, linearly constrained quadratic MPCCs, particularly in the presence of infeasible starting points and high degeneracy, and are well suited for problems that demand high satisfaction of the CCs.

7 Conclusion

We have introduced a new homotopy-based approach for solving MPCCs. The indicator function associated with the CC is nonsmooth and nonconvex, and is approximated using the Lasry–Lions double envelope to obtain a smooth reformulation. The original problem is then approximated by solving a sequence of smoothed subproblems with increasing penalization. When the proposed approach, Algorithm 1, is initialized from a feasible point, we show that it converges to a C/M-stationary point in the limiting setting. To obtain an approximate stationary point within a finite number of steps, we further propose a practical and easily implementable variant, Algorithm 3. It is shown that an ϵ-approximate C/M-stationary point is generated within $\mathcal{O}(\log(\epsilon^{-2}))$ outer iterations, and $\mathcal{O}(\epsilon^{-4}\log(\epsilon^{-2}))$ total iterations when a first-order method is used as the inner solver. In Algorithm 3, a smooth subproblem with simple constraints is used to approximate the general constrained MPCCs, and thus fast optimization algorithms can be readily deployed. The effectiveness of the proposed approach is validated through a comprehensive set of benchmark problems against the classical regularization method and state-of-the-art NLP solvers. As for future work, the following three directions will be pursued: (i) To make BCCQ usable in practice, verifiable conditions for BCCQ will be studied. (ii) As mentioned in Remark 4.5, the implications among MPCC-CRCQ, MPCC-CPLD, and BCCQ will be established. (iii) As observed in

Table 3 Objective, constraint violation and computation time on linearly constrained quadratic MPCCs

n_0–p–degen.–num.	LL1			LL2(1st)			ALGENCAN			IPOPT		
	obj.	vio.	time(s)	obj.	vio.	time(s)	obj.	vio.	time(s)	obj.	vio.	time(s)
100–400–zero–0	**-221887.01**	2.5e-10	97	**0%**	3.0e-10	**93**	<0.1%	2.4e-04	712	>100%	3.1e-03	525
100–400–zero–1	**-100443.01**	5.5e-11	**9.21**	**0%**	6.7e-11	10.65	<0.1%	1.1e-04	631	1.39%	2.3e-05	457
100–400–zero–2	**-127464.67**	7.7e-11	**30.42**	**0%**	6.3e-11	30.49	<0.1%	9.3e-05	776	>100%	1.1e-05	232
100–400–zero–3	**-78290.04**	4.5e-11	**2.25**	**0%**	5.2e-11	3.61	57.99%	8.3e-09	672	2.12%	2.2e-04	295
100–400–zero–4	**-111968.19**	1.1e-10	35.18	**0%**	9.0e-11	**34.50**	<0.1%	2.3e-03	340	>100%	1.7e-04	170
100–400–zero–5	**-75648.18**	1.8e-10	**28.96**	**0%**	3.4e-11	29.17	<0.1%	3.6e-04	322	>100%	1.4e-03	130
100–400–zero–6	**-152564.60**	5.6e-11	**13.58**	**0%**	1.8e-10	16.67	1.05%	8.5e-10	1077	32.86%	1.1e-05	69
100–400–zero–7	**-38928.21**	5.8e-11	**35.00**	**0%**	5.8e-11	39.22	<0.1%	9.5e-05	667	0.15%	2.7e-03	324
100–400–zero–8	**-24295.04**	1.5e-09	**5.10**	**0%**	1.1e-08	6.49	<0.1%	1.8e-04	420	>100%	2.6e-05	584
100–400–zero–9	**-153419.83**	7.0e-10	175	<0.1%	7.4e-10	198	<0.1%	4.6e-04	930	>100%	8.3e-03	**163**
100–400–high–0	**-68683.11**	3.3e-10	196	**0%**	2.9e-10	**192**	<0.1%	2.1e-03	4073	>100%	9.8e-04	265
100–400–high–1	**-55672.63**	1.1e-09	**17.19**	**0%**	4.9e-11	19.14	**0%**	6.8e-06	2545	0.38%	9.9e-04	533
100–400–high–2	**-53030.90**	6.1e-11	**78**	**0%**	7.5e-11	93	<0.1%	5.0e-05	4092	0.24%	3.2e-05	681
100–400–high–3	**-32820.07**	1.5e-09	**4.01**	**0%**	1.2e-09	5.53	**0%**	9.2e-05	2454	42.55%	8.2e-06	1047
100–400–high–4	**-42400.77**	1.1e-10	**52**	**0%**	6.2e-11	57	<0.1%	1.0e-02	3376	>100%	1.9e-05	293
100–400–high–5	**-25807.63**	1.1e-10	118	**0%**	1.0e-10	**98**	<0.1%	2.8e-03	3982	>100%	5.3e-04	102
100–400–high–6	3.08%	2.5e-03	292	0.41%	1.0e-03	**285**	**-2729.96**	5.9e-03	3938	>100%	5.4e-05	443

Table 3 continued

n_0–p–degen.–num.	LL1			LL2(1st)			ALGENCAN			IPOPT		
	obj.	vio.	time(s)	obj.	vio.	time(s)	obj.	vio.	time(s)	obj.	vio.	time(s)
100–400–high–7	**-59925.08**	1.1e-10	**99**	**0%**	2.8e-10	136	<0.1%	3.2e-03	3739	0.39%	1.1e-03	447
100–400–high–8	**-45550.65**	3.8e-07	172	**0%**	5.2e-08	**156**	<0.1%	1.2e-03	3674	0.15%	1.0e-03	418
100–400–high–9	**-52139.38**	6.3e-07	184	**0%**	6.0e-07	**175**	<0.1%	7.1e-03	2945	>100%	6.2e-03	505
200–800–zero–0	**-25562.79**	5.3e-11	**655**	**0%**	6.0e-11	691	<0.1%	1.7e-03	12803	>100%	1.2e-03	3328
200–800–zero–1	**-232452.69**	4.8e-11	**151**	**0%**	3.6e-11	162	<0.1%	3.5e-03	2764	>100%	2.5e-04	2007
200–800–zero–2	**-323667.78**	3.3e-11	**11.82**	**0%**	7.4e-11	18.80	**0%**	3.6e-04	1984	20.01%	3.7e-04	1890
200–800–zero–3	**-244977.07**	1.2e-09	**682**	<0.1%	6.3e-10	753	<0.1%	3.4e-03	2606	>100%	1.1e-04	2580
200–800–zero–4	**-128268.60**	3.4e-11	**13.72**	**0%**	5.1e-11	21.06	**0%**	7.1e-05	2486	11.97%	1.0e-04	2184
200–800–zero–5	**-117778.88**	3.5e-11	**38.50**	**0%**	9.6e-12	45.72	**0%**	2.4e-03	4054	57.22%	5.1e-05	4305
200–800–zero–6	**-234265.58**	1.6e-10	**224**	<0.1%	2.2e-10	230	<0.1%	3.1e-03	2247	>100%	7.3e-05	1935
200–800–zero–7	**-28086.92**	7.4e-12	**23.52**	**0%**	3.3e-11	30.75	<0.1%	3.0e-04	1708	14.89%	1.7e-04	7822
200–800–zero–8	**-292242.58**	1.0e-11	**72**	**0%**	1.8e-11	79	<0.1%	2.1e-03	6099	>100%	1.6e-04	6265
200–800–zero–9	**-668174.30**	1.0e-10	**62**	**0%**	5.2e-11	70	<0.1%	1.3e-03	2149	>100%	5.3e-05	5044
200–800–high–0	**-11090.41**	6.7e-11	382	**0%**	8.4e-11	**358**	<0.1%	1.2e-02	18047	0.36%	6.6e-05	8932
200–800–high–1	**-65477.02**	8.2e-07	**858**	**0%**	2.1e-07	907	<0.1%	6.0e-03	16212	0.16%	2.0e-03	10552
200–800–high–2	**-122853.43**	2.2e-10	**25.50**	**0%**	3.5e-09	29.68	**0%**	1.0e-04	10190	38.39%	7.8e-05	8718
200–800–high–3	**-81872.19**	1.6e-07	**952**	**0%**	1.2e-06	961	<0.1%	1.8e-02	18445	>100%	1.4e-02	9795
200–800–high–4	**-59376.49**	2.3e-09	**32.30**	**0%**	2.7e-09	37.18	**0%**	1.5e-05	12224	79.89%	3.8e-05	12542
200–800–high–5	**-46166.97**	2.7e-11	39.90	**0%**	9.4e-10	**39.42**	<0.1%	1.4e-03	22028	>100%	1.0e-04	7088
200–800–high–6	**-76357.26**	8.8e-11	686	**0%**	7.0e-11	**592**	<0.1%	8.3e-03	21230	>100%	8.4e-05	2966
200–800–high–7	<0.1%	8.2e-07	**1253**	<0.1%	2.9e-06	1262	**-14294.96**	7.1e-03	16925	0.30%	8.0e-04	6985
200–800–high–8	**-102912.14**	7.3e-10	**75**	**0%**	4.1e-11	88	<0.1%	3.1e-03	18732	>100%	6.3e-05	4821
200–800–high–9	**-241410.59**	5.7e-11	**123**	**0%**	3.5e-11	124	<0.1%	1.9e-03	15742	>100%	4.5e-05	5331

Tables 2 and 3, the computational time increases with the problem dimension, the number of CCs and the level of degeneracy. Therefore, more efficient algorithms, such as those incorporating preconditioning in the inner solver, are worth developing.

Acknowledgements This work was supported by: Fond for Scientific Research Vlaanderen (FWO) projects G033822N, G081222N, G0A0920N; KU Leuven internal funding C14/24/103; Japan Society for the Promotion of Science KAKENHI grants JP21K17710 and JP24K20737.

A Proof of Theorem 3.5

The complementarity set D, defined in (1.2), can be equivalently written as

$$D = D_1 \cup D_2 \quad \text{with} \quad D_1 := \{0\} \times \mathbb{R}_+ \quad \text{and} \quad D_2 := \mathbb{R}_+ \times \{0\}, \tag{A.1}$$

which corresponds to the so-called *disjunctive constraint*; see, for instance, [24, Section 5.1]. Then, the Moreau envelope in (3.2) admits the following representation:

$$\mathrm{env}_\lambda\, \delta_D(z) = \tfrac{1}{2\lambda}\, \mathrm{dist}\,(z, D)^2 = \tfrac{1}{2\lambda} \min\{\mathrm{dist}\,(z, D_1)^2, \mathrm{dist}\,(z, D_2)^2\}. \tag{A.2}$$

Based on the above notation, we derive the Lasry–Lions double envelope following Definition 3.2. Specifically, we have

$$\begin{aligned}
&\mathrm{env}_{\lambda,\mu}\, \delta_D(z) \\
&\overset{(3.1)}{=} \sup_{z' \in \mathbb{R}^2} \{\mathrm{env}_\lambda\, \delta_D(z') - \tfrac{1}{2\mu}\|z' - z\|^2\} \\
&\overset{(A.2)}{=} \tfrac{1}{\lambda} \sup_{z' \in \mathbb{R}^2} \left\{ \tfrac{1}{2} \min\{\mathrm{dist}\,(z', D_1)^2, \mathrm{dist}\,(z', D_2)^2\} - \tfrac{1}{2\beta}\|z' - z\|^2 \right\} \\
&= -\tfrac{1}{\lambda} \inf_{z' \in \mathbb{R}^2} \left\{ \tfrac{1}{2\beta}\|z' - z\|^2 - \tfrac{1}{2} \min\{\mathrm{dist}\,(z', D_1)^2, \mathrm{dist}\,(z', D_2)^2\} \right\} \\
&= -\tfrac{1}{\lambda} \inf_{z' \in \mathbb{R}^2} \left\{ \tfrac{1}{2\beta}\|z' - z\|^2 + \max\{-\tfrac{1}{2}\, \mathrm{dist}\,(z', D_1)^2, -\tfrac{1}{2}\, \mathrm{dist}\,(z', D_2)^2\} \right\} \\
&= -\tfrac{1}{\lambda} \inf_{z' \in \mathbb{R}^2} \left\{ \max\{\tfrac{1}{2\beta}\|z' - z\|^2 - \tfrac{1}{2}\, \mathrm{dist}\,(z', D_1)^2, \tfrac{1}{2\beta}\|z' - z\|^2 - \tfrac{1}{2}\, \mathrm{dist}\,(z', D_2)^2\} \right\}
\end{aligned} \tag{A.3a}$$

$$\begin{aligned}
= -\tfrac{1}{\lambda} \inf_{z' \in \mathbb{R}^2} \Bigg\{ \max_{y \in \Delta_2} \Big[& y_1 \left(\tfrac{1}{2\beta}\|z' - z\|^2 - \tfrac{1}{2}\, \mathrm{dist}\,(z', D_1)^2 \right) \\
& + y_2 \left(\tfrac{1}{2\beta}\|z' - z\|^2 - \tfrac{1}{2}\, \mathrm{dist}\,(z', D_2)^2 \right) \Big] \Bigg\},
\end{aligned} \tag{A.3b}$$

where $y := (y_1, y_2) \in \mathbb{R}^2$ and $\Delta_2 := \{y \in \{0, 1\}^2 \mid y_1 + y_2 = 1\}$. Hence, the either-or structure of the maximization in (A.3a) is equivalently reformulated as a maximization over the discrete set Δ_2 in (A.3b). Note that the inner maximization problem in (A.3b) corresponds to the *support function* of the set Δ_2; see [42, 8(10)]. Thus, in light of [4,

Example 11.2], it follows that (A.3b) admits an equivalent formulation:

$$\operatorname{env}_{\lambda,\mu}\delta_D(z) = -\tfrac{1}{\lambda}\inf_{z'\in\mathbb{R}^2}\left\{\max_{y\in\operatorname{cl}\operatorname{conv}\Delta_2}\begin{bmatrix} y_1\left(\frac{1}{2\beta}\|z'-z\|^2 - \frac{1}{2}\operatorname{dist}(z', D_1)^2\right) \\ +y_2\left(\frac{1}{2\beta}\|z'-z\|^2 - \frac{1}{2}\operatorname{dist}(z', D_2)^2\right)\end{bmatrix}\right\}, \tag{A.4}$$

where $\operatorname{cl}\operatorname{conv}\Delta_2 = \{y\in\mathbb{R}^2_+ \mid y_1+y_2=1\}\subset\mathbb{R}^2$. Note that the distance functions appearing in (A.4) can be expressed explicitly. In particular, we recall that the sets D_1 and D_2 are defined in (A.1), the associated projections $\Pi_{D_1}(z)$ and $\Pi_{D_2}(z)$ admit

$$\Pi_{D_1}(z) = \begin{pmatrix} 0 \\ [z_2]_+\end{pmatrix} \quad\text{and}\quad \Pi_{D_2}(z) = \begin{pmatrix} [z_1]_+ \\ 0\end{pmatrix}.$$

Therefore, the distance function associated with D_1 can be written as

$$\begin{aligned}\tfrac{1}{2}\operatorname{dist}(z', D_1)^2 = \tfrac{1}{2}\|z'-\Pi_{D_1}(z')\|^2 &= \tfrac{1}{2}\|z'\|^2 - \langle z', \Pi_{D_1}(z')\rangle + \tfrac{1}{2}\|\Pi_{D_1}(z')\|^2 \\ &= \tfrac{1}{2}\|z'\|^2 - z_2'[z_2']_+ + \tfrac{1}{2}[z_2']_+^2 \\ &= \tfrac{1}{2}\|z'\|^2 - [z_2']_+^2 + \tfrac{1}{2}[z_2']_+^2 \\ &= \tfrac{1}{2}\|z'\|^2 - \tfrac{1}{2}[z_2']_+^2\end{aligned}$$

and, analogously, the distance function associated with D_2 is given by $\frac{1}{2}\|z'\|^2 - \frac{1}{2}[z_1']_+^2$. Substituting these expressions into (A.4) yields

$$\begin{aligned}&\operatorname{env}_{\lambda,\mu}\delta_D(z) \\ &= -\tfrac{1}{\lambda}\inf_{z'\in\mathbb{R}^2}\left\{\max_{y\in\operatorname{cl}\operatorname{conv}\Delta_2}\begin{bmatrix} y_1\left(\frac{1}{2\beta}\|z'-z\|^2 - \frac{1}{2}\|z'\|^2 + \frac{1}{2}[z_2']_+^2\right) \\ +y_2\left(\frac{1}{2\beta}\|z'-z\|^2 - \frac{1}{2}\|z'\|^2 + \frac{1}{2}[z_1']_+^2\right)\end{bmatrix}\right\} \\ &= -\tfrac{1}{\lambda}\max_{y\in\operatorname{cl}\operatorname{conv}\Delta_2}\left\{\inf_{z'\in\mathbb{R}^2}\begin{bmatrix} y_1\left(\frac{1}{2\beta}\|z'-z\|^2 - \frac{1}{2}\|z'\|^2 + \frac{1}{2}[z_2']_+^2\right) \\ +y_2\left(\frac{1}{2\beta}\|z'-z\|^2 - \frac{1}{2}\|z'\|^2 + \frac{1}{2}[z_1']_+^2\right)\end{bmatrix}\right\} \\ &= -\tfrac{1}{\lambda}\max_{y\in\operatorname{cl}\operatorname{conv}\Delta_2}\left\{\inf_{z'\in\mathbb{R}^2}\begin{pmatrix} \frac{1}{2\beta}\|z'-z\|^2 - \frac{1}{2}\|z'\|^2 \\ +\frac{1}{2}[z_1']_+^2 + \frac{1}{2}[z_2']_+^2 - \frac{1}{2}y_1[z_1']_+^2 - \frac{1}{2}y_2[z_2']_+^2\end{pmatrix}\right\} \\ &= -\tfrac{1}{\lambda}\max_{y\in\operatorname{cl}\operatorname{conv}\Delta_2}\left\{\inf_{z'\in\mathbb{R}^2}\begin{pmatrix} \frac{1}{2\beta}\sum_{i=1}^2(z_i'-z_i)^2 - \frac{1}{2}\sum_{i=1}^2(z_i')^2 \\ +\frac{1}{2}\sum_{i=1}^2[z_i']_+^2 - \frac{1}{2}\sum_{i=1}^2 y_i[z_i']_+^2\end{pmatrix}\right\} \\ &= -\tfrac{1}{\lambda}\max_{y\in\operatorname{cl}\operatorname{conv}\Delta_2}\left\{\sum_{i=1}^2\inf_{z_i'\in\mathbb{R}}\left(\tfrac{1}{2\beta}(z_i'-z_i)^2 - \tfrac{1}{2}(z_i')^2 + \tfrac{1}{2}[z_i']_+^2 - \tfrac{1}{2}y_i[z_i']_+^2\right)\right\},\end{aligned} \tag{A.5}$$

in which the second equality follows from the *minimax theorem*, since the inner function is linear in y and convex in z', owing to the fact that $\beta\in(0,1)$. The third equality follows from the constraint $y_1+y_2=1$. Consequently, we obtain two one-dimensional

inner convex minimization problems in (A.5), each of which admits the minimizer, together with the corresponding function value, given by

$$z_i'^{\star} = \begin{cases} \frac{1}{1-\beta} z_i, & \text{if } z_i \le 0 \\ \frac{1}{1-\beta y_i} z_i, & \text{if } z_i \ge 0 \end{cases} \tag{A.6a}$$

and

$$\tfrac{1}{2\beta}(z_i'^{\star} - z_i)^2 - \tfrac{1}{2}(z_i'^{\star})^2 + \tfrac{1}{2}[z_i'^{\star}]_+^2 - \tfrac{1}{2} y_i [z_i'^{\star}]_+^2 = \begin{cases} -\frac{1}{2(1-\beta)} z_i^2, & \text{if } z_i \le 0 \\ -\frac{y_i}{2(1-\beta y_i)} z_i^2, & \text{if } z_i \ge 0. \end{cases} \tag{A.6b}$$

Depending on the positions of z_1 and z_2, we consider the following four cases.

(i) $z_1 \le 0$ and $z_2 \le 0$. This case corresponds to the region O_- shown in Fig. 2, and (A.5) becomes

$$\begin{aligned} \mathrm{env}_{\lambda,\mu}\, \delta_D(z) &\overset{(A.6)}{=} -\tfrac{1}{\lambda} \max_{y \in \mathrm{cl\,conv}\, \Delta_2} \left\{ -\tfrac{1}{2(1-\beta)} z_1^2 - \tfrac{1}{2(1-\beta)} z_2^2 \right\} \\ &= \tfrac{1}{\lambda} \tfrac{1}{2(1-\beta)} \|z\|^2 = \tfrac{1}{\lambda} r_\beta(z), \end{aligned}$$

as needed in (3.3) and (3.4).

(ii) $z_1 \le 0$ and $z_2 \ge 0$. This case corresponds to the left rectangular region of H_β^+ as labeled in Fig. 2, and (A.5) can be formulated as

$$\begin{aligned} \mathrm{env}_{\lambda,\mu}\, \delta_D(z) &\overset{(A.6)}{=} -\tfrac{1}{\lambda} \max_{y \in \mathrm{cl\,conv}\, \Delta_2} \left\{ -\tfrac{1}{2(1-\beta)} z_1^2 - \tfrac{y_2}{2(1-\beta y_2)} z_2^2 \right\} \\ &= \tfrac{1}{\lambda} \min_{y \in \mathrm{cl\,conv}\, \Delta_2} \left\{ \tfrac{1}{2(1-\beta)} z_1^2 + \tfrac{y_2}{2(1-\beta y_2)} z_2^2 \right\} \\ &= \tfrac{1}{\lambda} \tfrac{1}{2(1-\beta)} z_1^2 + \tfrac{1}{\lambda} \min_{y \in \mathrm{cl\,conv}\, \Delta_2} \left\{ \tfrac{y_2}{2(1-\beta y_2)} z_2^2 \right\}. \end{aligned} \tag{A.7}$$

Recalling that $\mathrm{cl\,conv}\, \Delta_2 = \{y \in \mathbb{R}_+^2 \mid y_1 + y_2 = 1\}$, the minimization problem in (A.7) is equivalent to

$$\underset{y \in \mathbb{R}_+^2}{\text{minimize}} \ \tfrac{y_2}{2(1-\beta y_2)} z_2^2 \quad \text{subject to } \ y_1 + y_2 = 1.$$

Let $\tau \in \mathbb{R}$ be the Lagrange multiplier associated with the constraint $y_1 + y_2 = 1$. Then, the associated Lagrangian is given by

$$\mathcal{L}(y, \tau) = \tfrac{y_2}{2(1-\beta y_2)} z_2^2 + \tau \left(1 - y_1 - y_2\right),$$

and the dual function is obtained as

$$\psi(\tau) := \inf_{y \in \mathbb{R}_+^2} \mathcal{L}(y, \tau) = \tau + \inf_{y_1 \in \mathbb{R}_+} \{-\tau y_1\} + \inf_{y_2 \in \mathbb{R}_+} \left\{ \tfrac{y_2}{2(1-\beta y_2)} z_2^2 - \tau y_2 \right\}. \tag{A.8}$$

The first minimization problem in (A.8) implies that $\tau \leq 0$; otherwise $\psi(\tau) = -\infty$. Thus, its minimizer is $y_1(\tau) = 0$. Considering the second minimization problem in (A.8), the gradient with respect to y_2 is $\frac{z_2^2}{2(1-\beta y_2)^2} - \tau$, which is nonnegative, owing to the facts that $z_2 \geq 0$ and $\tau \leq 0$. Hence, the minimum is attained at $y_2(\tau) = 0$. Consequently, the minimizer in the dual function (A.8) is $y(\tau) = 0$, and the dual problem admits

$$\max_{\tau \in \mathbb{R}_-} \psi(\tau) = \max_{\tau \in \mathbb{R}_-} \tau = 0.$$

It is straightforward to verify that the minimization problem in (A.7) is convex, and thus *strong duality* holds. The Lasry–Lions double envelope is then given by

$$\mathrm{env}_{\lambda,\mu}\, \delta_D(z) \overset{(A.7)}{=} \tfrac{1}{\lambda} \tfrac{1}{2(1-\beta)} z_1^2 = \tfrac{1}{\lambda} r_\beta(z),$$

which yields the expressions in (3.3) and (3.4).

(iii) $z_1 \geq 0$ and $z_2 \leq 0$. This case corresponds to the lower rectangular region of H_β^- shown in Fig. 2. Similar to the previous case, we have

$$\mathrm{env}_{\lambda,\mu}\, \delta_D(z) = \tfrac{1}{\lambda} \tfrac{1}{2(1-\beta)} z_2^2 = \tfrac{1}{\lambda} r_\beta(z).$$

(iv) $z_1 > 0$ and $z_2 > 0$. This case corresponds to the region T_β, the right triangular region of H_β^+ and the upper triangular region of H_β^-, as labeled in Fig. 2. Proceeding as in the case ii, we have

$$\mathrm{env}_{\lambda,\mu}\, \delta_D(z) \overset{(A.5),(A.6)}{=} \tfrac{1}{\lambda} \min_{y \in \mathrm{cl\,conv}\, \Delta_2} \left\{ \tfrac{y_1}{2(1-\beta y_1)} z_1^2 + \tfrac{y_2}{2(1-\beta y_2)} z_2^2 \right\}. \tag{A.9}$$

Let $\tau \in \mathbb{R}$ be the Lagrange multiplier, the associated Lagrangian and the dual function are

$$\begin{aligned} \mathcal{L}(y, \tau) &= \tau + \sum_{i=1}^{2} \left(\tfrac{y_i z_i^2}{2(1-\beta y_i)} - \tau y_i \right) \quad \text{and} \\ \psi(\tau) &= \tau + \sum_{i=1}^{2} \inf_{y_i \in \mathbb{R}_+} \left\{ \tfrac{y_i z_i^2}{2(1-\beta y_i)} - \tau y_i \right\}. \end{aligned} \tag{A.10}$$

The minimizers of the inner minimization problems in $\psi(\tau)$ are given by

$$y_i(\tau) = \begin{cases} \frac{1}{\beta} - \frac{z_i}{\beta\sqrt{2\tau}}, & \text{if } \tau \geq \frac{z_i^2}{2} \\ 0, & \text{if } \tau \leq \frac{z_i^2}{2}. \end{cases}$$

In this way, we obtain the function $\psi_i(\tau)$ and its gradient with respect to τ as follows:

$$\psi_i(\tau) := \inf_{y_i \in \mathbb{R}_+} \left\{ \tfrac{y_i}{2(1-\beta y_i)} z_i^2 - \tau y_i \right\} = \begin{cases} -\frac{(\sqrt{2\tau}-z_i)^2}{2\beta}, & \text{if } \tau \geq \frac{z_i^2}{2} \\ 0, & \text{if } \tau \leq \frac{z_i^2}{2} \end{cases} \tag{A.11a}$$

and

$$\nabla\psi_i(\tau) = \begin{cases} -\frac{\sqrt{2\tau}-z_i}{\beta\sqrt{2\tau}}, & \text{if } \tau \geq \frac{z_i^2}{2} \\ 0, & \text{if } \tau \leq \frac{z_i^2}{2}. \end{cases} \tag{A.11b}$$

To solve the dual problem maximize$_\tau$ $\psi(\tau)$, we consider the following three cases, depending on the positions of z_1 and z_2.

iv.1 $\tau \geq \frac{z_1^2}{2}$ and $\tau \leq \frac{z_2^2}{2}$. In this case, the gradient of the dual function admits

$$\nabla\psi(\tau) = 1 + \nabla\psi_1(\tau) + \nabla\psi_2(\tau) = 1 - \tfrac{\sqrt{2\tau}-z_1}{\beta\sqrt{2\tau}},$$

and setting $\nabla\psi(\tau) = 0$ yields $\tau^\star = \frac{1}{2}\left(\frac{z_1}{1-\beta}\right)^2$. This implies that $z_1 > 0$ and $\frac{z_1}{1-\beta} \leq z_2$, which characterizes the right triangular region of H_β^+. By invoking strong duality once again, the Lasry–Lions double envelope is given by

$$\mathrm{env}_{\lambda,\mu}\, \delta_D(z) \overset{(A.9)}{=} \tfrac{1}{\lambda} \max_\tau \psi(\tau) \overset{(A.10),(A.11)}{=} \tfrac{1}{\lambda}\left(\tau^\star - \tfrac{(\sqrt{2\tau^\star}-z_1)^2}{2\beta}\right) = \tfrac{1}{\lambda}\tfrac{1}{2(1-\beta)} z_1^2 = \tfrac{1}{\lambda} r_\beta(z),$$

as required in (3.3) and (3.4).

iv.2 $\tau \leq \frac{z_1^2}{2}$ and $\tau \geq \frac{z_2^2}{2}$. Similar to the previous case, we obtain $z_2 > 0$ and $\frac{z_2}{1-\beta} \leq z_1$, which characterizes the upper triangular region of H_β^-. Moreover, the Lasry–Lions double envelope is expressed as

$$\mathrm{env}_{\lambda,\mu}\, \delta_D(z) = \tfrac{1}{\lambda}\tfrac{1}{2(1-\beta)} z_2^2 = \tfrac{1}{\lambda} r_\beta(z).$$

iv.3 $\tau \geq \frac{z_1^2}{2}$ and $\tau \geq \frac{z_2^2}{2}$. Finally, the gradient of the dual function in this case reads

$$\nabla\psi(\tau) = 1 + \nabla\psi_1(\tau) + \nabla\psi_2(\tau) = 1 - \tfrac{\sqrt{2\tau}-z_1}{\beta\sqrt{2\tau}} - \tfrac{\sqrt{2\tau}-z_2}{\beta\sqrt{2\tau}}.$$

Setting $\nabla\psi(\tau) = 0$ yields $\tau^\star = \frac{1}{2}\left(\frac{z_1+z_2}{2-\beta}\right)^2$, which implies that $(1-\beta)z_1 \leq z_2 \leq \frac{z_1}{1-\beta}$, thus characterizing the region T_β. The Lasry–Lions double envelope then becomes

$$\mathrm{env}_{\lambda,\mu}\, \delta_D(z) \overset{(A.9)}{=} \tfrac{1}{\lambda} \max_\tau \psi(\tau)$$

$$\text{by (A.10) and (A.11)} = \frac{1}{\lambda}\left(\tau^\star - \frac{\left(\sqrt{2\tau^\star}-z_1\right)^2}{2\beta} - \frac{\left(\sqrt{2\tau^\star}-z_2\right)^2}{2\beta}\right)$$
$$= \frac{1}{\lambda}\left(\frac{1}{2\beta(2-\beta)}(z_1+z_2)^2 - \frac{1}{2\beta}\|z\|^2\right) = \frac{1}{\lambda}r_\beta(z),$$

as in (3.3) and (3.4).

In this way, the formulation of the Lasry–Lions double envelope is verified.

References

1. Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: On augmented Lagrangian methods with general lower-level constraints. SIAM J. Optim. **18**(4), 1286–1309 (2008)
2. Anitescu, M., Tseng, P., Wright, S.J.: Elastic-mode algorithms for mathematical programs with equilibrium constraints: global convergence and stationarity properties. Math. Program. **110**, 337–371 (2007)
3. Attouch, H., Aze, D.: Approximation and regularization of arbitrary functions in Hilbert spaces by the Lasry-Lions method. Annales de l'Institut Henri Poincaré C, Analyse non linéaire **10**(3), 289–312 (1993)
4. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, New York, New York, NY (2011)
5. Beck, A.: First-Order Methods in Optimization. SIAM, Philadelphia (2017)
6. Bertsekas, D.P.: Nonlinear Programming, 2nd edn. Athena Scientific, Belmont, MA (1999)
7. Burke, J.V., Hoheisel, T.: Epi-convergent smoothing with applications to convex composite functions. SIAM J. Optim. **23**(3), 1457–1479 (2013)
8. Chen, Y., Florian, M.: The nonlinear bilevel programming problem: Formulations, regularity and optimality conditions. Optimization **32**, 193–209 (1995)
9. Coleman, T.F., Li, Y.: An interior trust region approach for nonlinear minimization subject to bounds. SIAM J. Optim. **6**(2), 418–445 (1996)
10. Ferris, M.C., Pang, J.S.: Engineering and economic applications of complementarity problems. SIAM Rev. **39**(4), 669–713 (1997)
11. Flegel, M.L.: Constraint Qualifications and Stationarity Concepts for Mathematical Programs with Equilibrium Constraints. PhD thesis, Institute of Applied Mathematics and Statistics, Universität Würzburg (2005)
12. Flegel, M.L., Kanzow, C.: A Fritz John approach to first order optimality conditions for mathematical programs with equilibrium constraints. Optimization **52**(3), 277–286 (2003)
13. Flegel, M.L., Kanzow, C.: Abadie-type constraint qualification for mathematical programs with equilibrium constraints. J. Optim. Theory Appl. **124**(3), 595–614 (2005)
14. Flegel, M.L., Kanzow, C.: On the Guignard constraint qualification for mathematical programs with equilibrium constraints. Optimization **54**(6), 517–534 (2005)
15. Fletcher, R., Leyffer, S., Ralph, D., Scholtes, S.: Local convergence of SQP methods for mathematical programs with equilibrium constraints. SIAM J. Optim. **17**(1), 259–286 (2006)
16. Fourer, R., Gay, D.M., Kernighan, B.W.: AMPL: A Modeling Language for Mathematical Programming, 2nd edn. Thomson, Pacific Grove, CA (2003)
17. Ghadimi, S., Lan, G., Zhang, H.: Generalized uniformly optimal methods for nonlinear programming. J. Sci. Comput. **79**, 1854–1881 (2019)
18. Grapiglia, G.N.: Worst-case evaluation complexity of a quadratic penalty method for nonconvex optimization. Optimization Methods and Software **38**(4), 781–803 (2023)
19. Hermans, B., Pipeleers, G., Patrinos, P.: A penalty method for nonlinear programs with set exclusion constraints. Automatica **127**(109500), 1–8 (2021)
20. Hoheisel, T., Kanzow, C., Schwartz, A.: Theoretical and numerical comparison of relaxation methods for mathematical programs with complementarity constraints. Math. Program. **137**, 257–288 (2013)
21. Hu, X.M., Ralph, D.: Convergence of a penalty method for mathematical programming with complementarity constraints. J. Optim. Theory Appl. **123**, 365–390 (2004)

22. Izmailov, A.F., Solodov, M.V.: An active-set Newton method for mathematical programs with complementarity constraints. SIAM J. Optim. **19**(3), 1003–1027 (2008)
23. Izmailov, A.F., Solodov, M.V., Uskov, E.I.: Global convergence of augmented Lagrangian methods applied to optimization problems with degenerate constraints, including problems with complementarity constraints. SIAM J. Optim. **22**(4), 1579–1606 (2012)
24. Jia, X., Kanzow, C., Mehlitz, P., Wachsmuth, G.: An augmented Lagrangian method for optimization problems with structured geometric constraints. Math. Program. **199**, 1365–1415 (2023)
25. Jiang, H., Ralph, D.: QPECgen, a MATLAB generator for mathematical programs with quadratic objectives and affine variational inequality constraints. Comput. Optim. Appl. **13**, 25–59 (1999)
26. Kadrani, A., Dussault, J.P., Benchakroun, A.: A new regularization scheme for mathematical programs with complementarity constraints. SIAM J. Optim. **20**(1), 78–103 (2009)
27. Kanzow, C., Schwartz, A.: A new regularization method for mathematical programs with complementarity constraints with strong convergence properties. SIAM J. Optim. **23**(2), 770–798 (2013)
28. Kanzow, C., Schwartz, A.: The price of inexactness: convergence properties of relaxation methods for mathematical programs with complementarity constraints revisited. Math. Oper. Res. **40**(2), 253–275 (2015)
29. Karimi, H., Nutini, J., Schmidt, M.: Linear convergence of gradient and proximal-gradient methods under the Polyak-Lojasiewicz condition. In: P. Frasconi, N. Landwehr, G. Manco, J. Vreeken (eds.) Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2016, Riva del Garda, Italy, September 19-23, 2016, Proceedings, Part I, *Lecture Notes in Computer Science*, vol. 9851, pp. 795–811. Springer (2016)
30. Kirches, C., Larson, J., Leyffer, S., Manns, P.: Sequential linearization method for bound-constrained mathematical programs with complementarity constraints. SIAM J. Optim. **32**(1), 75–99 (2022)
31. Lan, G., Li, T., Xu, Y.: Projected gradient methods for nonconvex and stochastic optimization: new complexities and auto-conditioned stepsizes, (2024). arXiv:2412.14291 arXiv preprint
32. Lasry, J.M., Lions, P.L.: A remark on regularization in Hilbert spaces. Israel J. Math. **55**, 257–266 (1986)
33. Leyffer, S.: MacMPEC: AMPL Collection of MPECs, Argonne National Laboratory (2000)
34. Leyffer, S., López-Calva, G., Nocedal, J.: Interior methods for mathematical programs with complementarity constraints. SIAM J. Optim. **17**(1), 52–77 (2006)
35. Lin, G.H., Fukushima, M.: A modified relaxation scheme for mathematical programs with complementarity constraints. Ann. Oper. Res. **133**, 63–84 (2005)
36. Murphy, F., Pierru, A., Smeers, Y.: A tutorial on building policy models as mixed-complementarity problems. Interfaces **46**(6), 465–481 (2016)
37. Murtagh, B.A., Saunders, M.A.: MINOS 5.0 User's Guide. Technical report SOL 83.20, Stanford University, Stanford, CA (1983)
38. Outrata, J.V.: Optimality conditions for a class of mathematical programs with equilibrium constraints. Math. Oper. Res. **24**, 627–644 (1999)
39. Pang, J.S., Trinkle, J.C.: Complementarity formulations and existence of solutions of dynamic multi-rigid-body contact problems with Coulomb friction. Math. Program. **73**, 199–226 (1996)
40. Raghunathan, A.U., Biegler, L.T.: An interior point method for mathematical programs with complementarity constraints (MPCCs). SIAM J. Optim. **15**(3), 720–750 (2005)
41. Ralph, D., Wright, S.J.: Some properties of regularization and penalization schemes for MPECs. Optimization Methods and Software **19**(5), 527–556 (2004)
42. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis. Springer-Verlag, Berlin, Germany (1998)
43. Royset, J.O.: Consistent approximations in composite optimization. Math. Program. **201**, 339–372 (2023)
44. Scheel, H., Scholtes, S.: Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. Math. Oper. Res. **25**(1), 1–22 (2000)
45. Scholtes, S.: Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. SIAM J. Optim. **11**(4), 918–936 (2001)
46. Simões, M., Themelis, A., Patrinos, P.: Lasry-Lions envelopes and nonconvex optimization: A homotopy approach. In: 2021 29th European Signal Processing Conference (EUSIPCO), pp. 2089–2093. IEEE, Dublin, Ireland (2021)
47. Wächter, A., Biegler, L.T.: On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Math. Program. **106**(1), 25–57 (2006)

48. Ye, J.J.: Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints. J. Math. Anal. Appl. **307**(1), 350–369 (2005)
49. Ye, J.J., Zhu, D.L., Zhu, Q.J.: Exact penalization and necessary optimality conditions for generalized bilevel programming problems. SIAM J. Optim. **7**(2), 481–507 (1997)
50. Zhu, C., Byrd, R.H., Lu, P., Nocedal, J.: Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. ACM Transactions on Mathematical Software **23**(4), 550–560 (1997)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.